# TESTS OF THE GENERALITY OF THE CUE-OVERLOAD PRINCIPLE

Margaret Mary Gardiner

Department of Social Science and Humanities
The City University

Thesis submitted for the degree of
Doctor of Philosophy at The City University

December 1983

# CONTAINS PULLOUTS

CONTENTS

Page

Page

Page

LIST OF TABLES

Page

## LIST OF FIGURES

Page

Page

ACKNOWLEDGEMENTS

I am especially grateful to my supervisor, Dr. James A. Hampton, for his encouragement and constructive advice at every stage of the research, and for his thorough criticism of earlier versions of this thesis.

I would also like to acknowledge the contribution of [redacted] to the experimental work reported in the first chapter of the thesis, and to thank him for many hours of patient and sympathetic listening and for always being there when needed.

Finally, my thanks to 520 anonymous subjects, who so generously donated some of their free time for little or no pay.

## DECLARATION TO THE UNIVERSITY LIBRARIAN

I grant powers to the University Librarian to allow this thesis to be copied in whole or in parts without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

ABSTRACT

The primary objective of the research reported here was to test the generality of a Cue-Overload interpretation of memory phenomena. This states that recall is mediated by cues and that these cues get overloaded, and become less effective for recall, as they come to subsume more and more items. A secondary aim of the thesis was to investigate the effects of typicality on recall.

The experimental work reported falls into three broad categories. Experiments 1 and 2 establish the effects of cue-overload in a release from retroactive inhibition paradigm, and support a common interpretation of release from retroactive and proactive inhibition in terms of Cue-Overload. Experiments 3, 5, 6 and 7 demonstrate that the Cue-Overload Principle breaks down when it is required to predict the effects on recall of systematic changes in the strength of association among to-be-recalled items and between these and the cues used for recall, as determined by the items' rated typicality. These experiments also provide evidence of the effects of typicality on recall, and test the appropriateness of some of the available theories concerning the nature of typicality effects in general. Experiment 4 describes the collection of a comprehensive set of normative data -- rated typicality, rated familiarity and associative frequency -- for a 531-word corpus, and describes a correlational analysis of the interdependence of the three measures. Such an analysis is not possible using the commonly used, American, sets of norms, and it extends our understanding of the nature of typicality.

The final chapter of the thesis summarizes the principal experimental findings and describes a proposed Similarity/Accessibility Hypothesis, which is shown to complement the Cue-Overload formulation and to greatly improve its ability to account for the data reported. The implications of the Hypothesis for future research are outlined.

# CHAPTER ONE : INTRODUCTION

## 1.1 Introduction

This chapter can be divided into four major sections. The first section (1.2) defines the Cue-Overload Principle, outlines its scope and evaluates its theoretical import. The principal aims of the thesis are then presented (Section 1.3). The next section (1.4) reviews the major theories associated with typicality (the main independent variable used in the research) and describes a subsidiary aim of the research. Finally, the experimental method chosen for most of the experiments reported is discussed (Section 1.5).

## 1.2 The Cue-Overload Principle

The Cue-Overload Principle states that recall is mediated by cues and that these cues are subject to overload -- and hence they lose some of their effectiveness -- as they come to subsume more and more events. The Principle was first described in a series of studies by Watkins and Watkins (1975; 1976; see also Watkins, 1979). Like Craik and Lockhart's (1972) levels-of-processing approach, Cue-Overload provides a general framework for memory research. It can cut across paradigms and explain a variety of memory phenomena in terms of the same set of basic principles. However, unlike the levels-of-processing approach, Cue-Overload is concerned with the way items are retrieved from (rather than encoded into) memory.

The Principle does not describe the mechanisms leading to overload, nor does it specify the exact relationship between the effectiveness of a given cue and the number of events it subsumes. This relationship is simply assumed to be inverse and monotonic, the "load" being brought about by the number of events, e.g. word presentations, in a given memory task.

Though it was formulated quite some time ago, the Cue-Overload Principle has not been extensively tested. Of the handful of studies which have investigated the Principle, most were carried out by Watkins and Watkins to demonstrate its scope and explanatory power (for instance, Watkins, 1975; 1979; Watkins & Watkins, 1975; 1976). Of the remaining studies, only Parkin (1980) tested the Principle's generality, using a levels-of-processing manipulation. By and large, researchers have tended to take the Principle as given, and used it as a means of interpreting their data, rather than tested its appropriateness as a predictive tool. For instance, Glenberg (in press) used Cue-Overload to interpret the modality effect on recall; Todres and Watkins (1981) quote it as one possible interpretation of the part-set cueing effect they obtained with recognition tests; Roediger and Neely (1982) mention the Principle as one way to explain retrieval blocks in memory.

In addition, Cue-Overload can account for the effects on recall of factors such as list-length, categorization (subjective and experimenter-imposed), part-list cueing, paired-associate learning, and build-up of, and release from, proactive and retroactive inhibition (see, for instance, Watkins & Watkins, 1975; Watkins, 1979). In the following section, each of these factors will be considered in turn, to illustrate the manner in which Cue-Overload is said to operate. The section does not attempt to provide a comprehensive theoretical review of each of these paradigms. Some theories are dealt with in more detail at appropriate points later in the thesis.

### 1.2.1 Cue-Overload explanations of memory phenomena

List-length effects   It is usually found that, as list length increases, so the probability of recalling individual list items monotonically decreases (for instance, Murdock, 1960). To explain this effect, the Cue-Overload Principle assumes that recall of a given list is mediated by cues like "list", "last list", or "last few words". Increasing list length has the effect of increasing the load on the cue used, leading to cue overload and decreasing the effectiveness of the cue for retrieval of the list items.

Recall of categorized lists   Long lists of words can, however, be easily recalled, if they are structured. For example, if lists are made up of items from several discrete categories, recall tends to be higher than if they are simply random arrays of words (for instance, Tulving & Pearlstone, 1966; Lewis, 1972). According to Cue-Overload, the better recall for categorized lists comes about because the category names are used as retrieval cues. Since these subsume fewer items than a general "list" cue, they are more effective for retrieval. Presumably, a general "list" cue is used to retrieve the category names themselves, prior to recall of the list items. This can also be overloaded, as demonstrated by the findings of experiments by Tulving and Pearlstone (1966), Hudson and Austin (1970) and Slamecka (1972). These studies showed that post-recall presentation of either the category names, or of items from the categories used, could lead to more words being recalled, from categories which the subjects had not initially generated.

Effects of subjective organization    If list items are randomly selected from a uniform population, recall can be considerably improved if subjects are instructed to group items together, or to sort them into categories (e.g., Tulving, 1962; 1964; Mandler & Pearlstone, 1966; Mandler, 1967; Mayhew, 1967). Cue-Overload explains this finding by assuming that the subjective category names serve as recall cues, and improve recall because they subsume fewer words than a general "list" cue used for unsorted lists. Alternatively, even if the subjects fail to name their subjective categories, individual inter-item associations established at the learning stage (for instance, "B seems to go with A, C seems to go with B and A", etc.) may be used to retrieve items sorted together, so that a list word, rather than a category name, may serve as the recall cue for a particular cluster.

Extralist cueing effects    Poor recall due to overloaded recall cues can be dramatically improved if the subjects are provided, prior to recall, with cues which specify certain items in the list (for instance, Bahrick, 1969; Bilodeau, 1967; see also Gardiner, Craik & Birtwistle, 1972). Because the extralist cues are not overloaded, they are highly effective for recall. The effects of extralist cueing also demonstrate that poor recall is not due solely to items being badly encoded. In other words, overload on a cue does not reflect progressively poorer encoding of the items it subsumes. If it did, then provision of additional cues would not have led to retrieval of extra items. Instead, the cue originally set up at the time of learning becomes overloaded and not effective for retrieval, though all items are equally well encoded. Experiment 2 in the present thesis followed the rationale behind Gardiner, Craik and Birtwistle's (1972) study. The effects of extralist cueing are discussed in more detail in the introduction to that experiment.

Part-set cueing effects     Research on part-set, or part-list cueing stemmed from work by Slamecka (1968; 1969). He found that if, prior to recall, subjects were shown again portions of the list they had just studied, their recall for the remaining list items tended to be inhibited, compared with recall by subjects who were simply asked to free-recall the list. This effect has been extensively replicated, using a variety of list materials (Roediger, 1973; Rundus, 1973; Mueller & Watkins, 1977) and can also be obtained with tasks tapping semantic, rather than episodic, memory (Brown, 1968). It occurs, too, when the words presented prior to recall are members of the categories represented in the list, but were not themselves included in it (Watkins, 1975), and when recognition, rather than recall, is used to test performance (Todres & Watkins, 1981). The fact that inhibition, and not facilitation, is obtained when extralist cues are provided in this paradigm may seem puzzling, when considered in conjunction with the extralist cueing effects described in the previous section. This discrepancy has been explained in terms of the information provided by the cues, in each paradigm (Roediger, 1974). If the additional cues enable the subject to gain access to higher-order units which he would otherwise have failed to recall (e.g., extra categories or clusters), then extralist cues will facilitate recall. If, however, the cues provide more information than is necessary to gain access to higher-order units, recall will be inhibited. Models of memory which interpret encoding and storage as the formation of intralist associations (e.g., Anderson, 1972; Anderson & Bower, 1972; 1973; 1974; Estes, 1972) have difficulty accommodating the inhibitory effects of part-list cueing, since provision at recall of part of the list should reinstate at least some of these intralist associations and hence improve, not impair, recall. Cue-Overload can, however, explain the inhibitory effect by assuming that re-presenting a list item at recall is functionally equivalent to adding another item to the list. Thus the part-set cueing effect can be interpreted in the same manner as list-length effects were: In the cued condition, retrieval cues are seen to subsume all the items in the study list, plus the re-presented items. In the free-recall condition, they subsume only the list items. So recall should be higher in this latter condition, because overload is not as great. By simple extension, if the cues presented at recall are extralist words consistent with the category set, then recall with these extralist cues should also be impaired -- and to the same extent -- compared with the free-recall situation. This is indeed the case (Watkins, 1975; Mueller & Watkins, 1977). The part-set cueing paradigm was used in Experiment 6 of the present thesis. The findings of experiments using this paradigm are discussed in more detail in Chapter Four, together with other theories which have been put forward to account for the effect.

Effects of paired-associate learning If the paired-associate refers to more than one term (for instance, in the A-B, A-C paradigm, where the paired-associate A refers to words B and C) recall is lower than if the associated word refers to one single item (Postman, Stark & Fraser, 1968). This finding follows directly from the basic Cue-Overload formulation, that a recall cue loses part of its effectiveness as it comes to refer to more and more events.

In this case, the recall cue is the paired-associate A. If it refers to two, or more items (B, C, etc.) recall should be lower than if it refers to one single term. In the former case it is more overloaded, and hence less efficient.

Build-up of proactive and retroactive inhibition The usual finding of build-up of proactive inhibition within the Brown-Peterson paradigm (e.g., Wickens, 1970; 1972; Wickens, Born & Allen, 1963) is readily interpretable in terms of Cue-Overload. As more and more triads are presented, all sharing the same conceptual attribute, so the retrieval cue based on that attribute will be overloaded, and recall impaired. Shifts in the nature of the triads with respect to the encoded attribute lead to release, or improved recall, because a new, non-overloaded cue is set up at the time of study, and this cue is highly efficient for retrieval. Note, however, that this assumes that cues relating to the temporal separation of the triads prior to a shift are not easily utilized by the subjects, to discriminate between lists. If they were, then recall would be high, and no inhibition would occur. The same interpretation can be applied to explain the build-up of and release from retroactive inhibition (see, e.g., Watkins & Watkins, 1976). In this case, too, cues relating to the attributes encoded are more loaded in the inhibitory condition, because they subsume both target and distractor items. Shifting the attribute shared by the list words, in the release condition, improves recall because in this condition only target words share the retrieval cue; the distractors are subsumed by another cue.

Cue-Overload can also explain the difference in the levels of release observed, for the same attributes, between the proactive and retroactive inhibition paradigms (cf. Underwood, 1948; see also Watkins & Watkins, 1976).

The magnitude of the release effect tends to be greater for the proactive inhibition paradigm, for the same attribute, compared with that obtained under retroactive inhibition. The difference need not reflect intrinsic differences in the processes involved in each paradigm. In the Brown-Peterson procedure, recall is required after each triad is presented. In the Watkins and Watkins (1976) retroactive inhibition paradigm, recall is tested once only, after all words have been presented. Thus in the Brown-Peterson paradigm, the attribute cue (e.g., a category name) is loaded by all the words presented in the inhibitory triads (normally 12 words, or 4 triads), plus the words recalled by the subject after each triad and, in the inhibitory condition, also the additional 3 words of the 5th triad. In the Watkins and Watkins procedure, cue loading corresponds only to the 6 words presented for recall plus, in the inhibitory condition, a further 6 words. The load is hence considerably smaller in this latter case. Release from proactive inhibition may, therefore, be a more sensitive procedure, involving greater cue loading, than the retroactive inhibition procedure.

### 1.2.2 Evaluation of the Cue-Overload Principle

The description of the scope of the Cue-Overload Principle presented in the previous section also serves to highlight one of its shortcomings. The fact that the Principle does not attempt to specify which cues will be utilized in particular retrieval attempts means that certain cues which in some paradigms are assumed to be operative and efficient are, for unspecified reasons, not used or not efficient in other paradigms. As a consequence, the Principle can only determine whether a cue was used by post hoc observation. If recall is poor, overload occurred; if it is not, then the cue or cues provided, or assumed to be present, were not overloaded, or not used. There is therefore an inherent danger of circularity in the Principle's formulation. For example, Cue-Overload can explain the decline in recall with increased list length by assuming that it reflects increased overload on a "last list" cue. In the proactive inhibition paradigm, because there is decline in recall across lists, the "last list" cue is presumably not used. Similarly, in a study by Gardiner, Klee, Redman and Ball (1976), within the release from proactive inhibition paradigm, cues referring to the ink colour used to write stimulus items were assumed to be effective retrieval cues when used to retrieve meaningless trigrams (release was obtained with shifts in ink colour), but not when used to retrieve words (release was not obtained with similar colour shifts).

Careful control of the experimental situation is therefore highly desirable, if the results of experiments intended to test Cue-Overload are to be unambiguously interpreted. The generality of the Principle's formulation is such that it can explain virtually any pattern of results by simply changing the emphasis placed on different cues, in different situations.

It has been argued (Watkins, 1979) that one of the advantages of a Cue-Overload interpretation of recall performance is the way in which it brings together a large number of seemingly disparate findings and explains them all in terms of a very simple set of general principles. This ability to cut across paradigms, it is argued, may provide a better insight into the way memory works than any collection of separate theories which concentrate on only one or two experimental paradigms at a time. One further advantage of the Principle, according to Watkins, is that, unlike these separate theories, it does not assume a spatial metaphor, where items are put into, stored in, and taken out of memory. Perhaps it is their attempts to describe what happens between items being put in memory and taken out of it that eventually renders spatial models inconsistent with the experimental data (see Roediger, 1980, for more discussion). It was this state of affairs which prompted the development of Cue-Overload theory for, according to Watkins, "... It is sobering to reflect, amidst the impressive array of paradigms and paraphernalia we now have before us, that still we can only observe what happens to our subject at time t1 and how this affects his or her behavior at time t2." (Watkins, 1979, p. 370). Because it does not make assumptions about what happens in the intervening time, Cue-Overload may provide a better way to interpret memory phenomena.

## 1.3 Summary of the principal objectives of the thesis

Research involving Cue-Overload is scarce, and has tended to use the Principle more as a "loose framework for research" (Watkins & Watkins, 1976, p. 289) than as a formal theory of memory. Formulation of such a theory requires that a lot more data be collected concerning, for example, the mechanisms determining the choice of cues, and cue effectiveness across a variety of types of cue and experimental conditions. The collection of such data was one of the main aims of the research reported in this thesis. The first two studies reported concentrate on testing the Principle's predictive power at a general level, by manipulating directly the degree of loading on the cues used for retrieval. Subsequent research then tested the Principle's generality at a more specific level, which will now be described.

Watkins and Watkins (1976) provide some information about which types of attributes are used as retrieval cues, in a retroactive inhibition paradigm, in much the same way that studies by Wickens (1970; 1972) mapped out the attributes used for encoding in a proactive inhibition paradigm. Watkins and Watkins (1976) demonstrated that shifts in category membership led to significant release from retroactive inhibition, indicating that category-membership cues are effective functional retrieval cues. However, not all items are equally representative of the categories in which they have membership (see Rosch, 1975a; 1978). Some items are considered more "typical" of a category than are other category members, and this typicality is seen to reflect the strength of association between a given item and other category members. Does the degree to which an item is considered a good, representative example of a category also determine the extent to which it overloads the category cue? Or is overload really brought about simply by the total number of to-be-recalled items to which the category name applies, irrespective of typicality? Similarly, does the prior strength of association between a cue and the items it subsumes determine the degree to which the cue is overloaded, and can an overloaded cue still be used effectively for recall, so long as it is used to retrieve items with which it is strongly associated? It is to these questions that the bulk of the research reported in this thesis is addressed.

## 1.4 Measures of associative strength

As stated in Section 1.3, the bulk of the research reported in this thesis investigated the degree to which the prior strength of association among items and between an item and its recall cue affected cue overload. Strength of association, of these two types, was defined for this purpose as the rated typicality of items within a category. This section gives a brief review of the main theories associated with the study of typicality, to justify this assumption.

Research by Rosch and her associates (e.g., Rosch, 1973; 1974; 1975a, b; 1978; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976; Rosch, Simpson & Miller, 1976) first introduced the notion of typicality. It reflects the fact that subjects find it a meaningful task to rate the degree of membership of an item in a given category, and that the ratings thus obtained are highly consistent among the subjects tested.

Rated typicality has been shown to be a good predictor of categorization time (Smith, Shoben & Rips, 1974; Hampton, 1979; McCloskey & Glucksberg, 1979) and of associative frequency (the measure of the probability of an item being generated to the category label; Mervis, Catlin & Rosch, 1976; Hampton, 1979). It also correlates well with measures of "semantic distance" (Rips, Shoben & Smith, 1973) and of the featural overlap between an item and either its fellow category members (Rosch & Mervis, 1975) or the category concept itself (Hampton, 1979; 1981).

Typicality has been the object of much research (mainly using categorization-speed tasks) particularly since it correlates well with associative frequency and the two measures have been seen to reflect the proposed category structure of two rival sets of theories of memory organization. Network-search models of semantic memory (e.g., Collins & Loftus, 1975) suggest that associative frequency is the most direct measure of category structure, and claim that the typicality effects described above are determined by the strength and search order of pathways linking the category node with its subordinate item-nodes. On the other hand, rated typicality is the more direct measure of category structure for featural and prototype models (e.g., Rosch, 1978). Rosch & Mervis (1975) argued that what lies behind the ratings of typicality is an estimation of the "family resemblance" between an item and its fellow items. This "family resemblance" was conceptualized in terms of the number of features shared by members of the category. Typical items, the "good examples" of the category, share many more features with other category members than do atypical category instances. Typicality effects are assumed to reflect this featural similarity among different category members.

So far, despite repeated efforts, there has been no unambiguous way to discredit either theory, and it is likely that both factors, and possibly others as well, may play some role in determining performance in tasks involving categorized materials (see, e.g., Collins & Loftus, 1975). Because of this, the approach adopted in this thesis is somewhat atheoretical with respect to the debate outlined above. Instead, typicality was used as a convenient independent variable, to see whether inter-item, and item-cue, strength of association in any way affected degree of cue overload.

However, the design of some of the experiments reported tried to take into account some factors of theoretical importance for the study of typicality, to the extent that these factors, if left unchecked, might interact with the testing of a Cue-Overload interpretation of the experimental results. Two uninteresting explanations of typicality were considered in this context:

(a) that "atypical" category members are really more typical items of categories other than that being tested, so that any effects of typicality are due in part to mispriming or confusion (Loftus, 1975); and

(b) that familiarity with the category members used in typicality experiments may at least in part be responsible for the patterns of results obtained. Atypical members may be worse recalled, or take longer to categorize, simply because they are not very familiar to the subjects (McCloskey, 1980; Malt & Smith, 1982).

Where possible, appropriate controls were adopted to try to eliminate the effects of these two alternative interpretations of typicality.

Finally, research on typicality has concentrated to a large extent on paradigms involving categorization speed and judgements of category membership. The effects of typicality on recall are less well researched, though it is usually assumed that typical items are better recalled than atypical items. Two studies which did manipulate typicality in the context of recall tests have particular relevance to this thesis. Keller & Kellas (1978) demonstrated release from proactive inhibition with shifts in typicality, as well as a typical-word recall advantage; Greenberg &

Bjorklund (1981) reported a similar recall advantage in the absence of associative frequency differences between typical and atypical words. However, neither study controlled word familiarity. These studies are considered in detail in Chapter Four (pp. 127 and 148 ). It is a subsidiary aim of the thesis to try to provide additional evidence of the effects of typicality on recall performance, under conditions where alternative, uninteresting explanations of such effects can be discounted.

## 1.5 Choice of experimental method

The Watkins and Watkins (1976) study described in Section 1.3 also introduced a modified version of the traditional method of retroaction, which is particularly well-suited for investigating the Cue-Overload Principle. This experimental procedure, dubbed the Method of Interpolated Attributes, involves presenting subjects with a short list of words (typically six words) which the subjects are instructed to memorize. They are then shown a second short list (usually same length) which they are requested to copy down onto appropriate booklets. Recall for the memory list is tested after the copying task. Within each list, all words are homogeneous with respect to a given attribute or dimension. The factor of interest is the relationship between the memory and copy lists. This can either be congruent -- both lists sharing the same attribute -- or incongruent -- each list sharing a different attribute. The congruent, or SAME condition is said to be the inhibition condition. The degree of inhibition is measured by comparison with the incongruent, or shift, or DIFFERENT condition. The amount of "release" (*)

---

(*) It may be argued that "release" is not an appropriate name for the effect, since with this experimental design inhibition is not "removed", as is the case with the Brown-Peterson procedure. Rather, two different levels of overload are compared, in the release from retroactive inhibition paradigm, one produced by two similar lists, and the other produced by one single list, composed of target words only. For simplicity's sake, however, "release" from retroactive inhibition will be used to name this effect.

obtained is expressed proportionalised on the percentage recall for the SAME condition, using the formula (D-S)/S, where S and D are the final recall percentages for SAME and DIFFERENT conditions, respectively.

The Method of Interpolated Attributes has the advantage that it guarantees encoding equivalence for memory list words across the two experimental conditions, since provided that condition order is unpredictable subjects cannot know which condition they are being tested under until after the presentation of the memory list. This means that the effects of inhibition and release can be confidently attributed to the retrieval stage. In addition, it means that subjects can be tested under both conditions (within-subjects designs), and that more observations can be obtained from each subject, increasing the reliability of the measurements obtained, and decreasing the size of the total sample of subjects needed. For these reasons, the Method of Interpolated Attributes was used for most of the experiments reported in the following chapters.

# CHAPTER TWO : CUE-LOADING MANIPULATIONS

## 2.1 Introduction

The three studies reported in this chapter fall into two broad categories. Experiments 1 and 2 test the generality of a Cue-Overload account of the release from retroactive inhibition phenomenon under conditions where cue loading is directly manipulated. Experiment 3 varied instead the prior strength of association among list items, and between these and the cues used for retrieval, and tested Cue-Overload's ability to predict or explain the results obtained.

## 2.2 Experiment 1 - Copy-list length and release from retroactive inhibition

### 2.2.1 Introduction

Experiment 1 served two very different purposes. First, because Watkins and Watkins (1976) give only a very brief description of the experimental procedure involved in the Method of Interpolated Attributes, replication of one of their experiments was desirable, to make sure that the procedure, as used here, was correctly understood. Their taxonomic category shift experiment was chosen for replication, both because the magnitude of the release effect with shifts in this attribute was the second highest in the series of experiments reported by Watkins and Watkins, and because later research in the thesis involved categorized materials.

The second purpose of the experiment had a more marked theoretical orientation, since it involved a direct test of the predictive power of the Cue-Overload Principle. According to the Principle, the release from retroactive inhibition effect obtained with shifts in taxonomic category (using the Method of Interpolated Attributes) is due to the different degrees of cue overload associated with each of the two conditions of the Method. In the SAME condition, the load on the category recall cue corresponds with the total number of word presentations contained in both the memory and the copy list, since they both belong in the same category. In the DIFFERENT condition, that load is halved, because copy-list words belong in a different category from the memory list. Thus recall is poorer in the SAME than in the DIFFERENT condition.

It follows from the above reasoning that, should the length of the copy list increase, recall for memory-list words in the SAME condition should proportionately decrease, while recall for memory-list items in the DIFFERENT condition should remain largely unaffected. This is because the recall cue in the DIFFERENT condition is always loaded by the total number of word presentations in the memory list alone, so that the length of the copy list should to a large extent (within limits) be immaterial. Thus Cue-Overload would predict that the actual amount of release should be greater with a longer, than with a shorter, copy list. Experiment 1 tested this hypothesis.

### 2.2.2 Method

Design and Materials Each subject was tested under all four conditions obtained by a factorial combination of condition type (SAME or DIFFERENT) and copy-list length (six or twelve words). Twelve categories were selected from the 56 listed in Battig and Montague (1969). They were: trees, flowers, fruit, animals, food flavourings, insects, birds, vegetables, sports, occupations, countries and musical instruments. The last four categories were used exclusively as copy material for the DIFFERENT condition. A total of 18 words were selected from each of the first eight categories, and 12 words from each of the remaining four. This sampling procedure excluded the top three or four words with the highest associative frequency scores, for each category. The words selected had a mean Battig and Montague production frequency rating of 72.2 (S.D.=68.9).

The words from each of the first eight categories were randomly divided into three sets of six items, and each set was allocated to a different list as follows: The first set was labelled "memory list", the second set "copy (6) list", and this second set was randomly mixed with the third set, to form the "copy (12) list". The same procedure was followed for the last four categories, with the obvious omission of the memory list. Thus each of the first eight categories yielded 3 separate lists, and each of the last four produced 2. The composition of the lists, and the order of words within each list, was kept constant throughout the experiment. Copies of the experimental materials can be found in Appendix One.

Memory-list category membership was yoked across condition type and copy-list length, so that each category was used equally often in each of the four experimental conditions. Each subject was tested using all 8 memory-list categories, going twice over the full set of four conditions. Condition presentation order was separately randomized for each subject, with thw only constraint that the same condition type (SAME or DIFFERENT) did not occur more than twice in succession. The four categories used for the DIFFERENT condition were also fully balanced across the 8 memory-list categories.

Subjects The subjects were 16 undergraduate and graduate students at The City University, who were tested individually and were paid for their assistance. They were all native English speakers.

Procedure The lists were printed in stencilled capitals, columnwise, on 5" x 8" flashcards and manually presented by the experimenter. The memory list was in view for a total of 6 seconds, and the copy lists were shown for as long as the subjects needed to complete the copying task. The subjects were provided with prepared booklets containing alternate pages labelled "Copy" and "Recall" and appropriately lined (12 and 6 lines, respectively). Interpolated between the pairs of testing sheets were pages containing a distractor task: a series of 6 sums (3 three-digit numbers per sum), which the subjects had to complete in their own time after each recall test. Together with the randomization of condition order per subject, the distractor task was meant to avoid confounding possible effects of proactive inhibition with condition, and to emphasize the separate identity of each pair of memory and copy lists. Recall was written and unconstrained, and typically lasted about 45 seconds.

The subjects were told that they were going to be shown a series of pairs of lists. Within each pair, they would have to memorize the first list, copy down the second list, and then try to recall as many words as they could from the first list. They were asked to give equal attention to both learning and copying tasks, and to write carefully and clearly while copying the second list. Subjects were then shown the response booklets and instructed as to their use. The distractor task was explained and the session started. The subjects were offered a break after the first four trials. Each testing session lasted approximately 30 minutes.

### 2.2.3 Results

The results are shown collapsed across category for each of the four experimental conditions -- a total of 192 observations per condition (2 categories x 16 subjects x 6 target words). Table 2.1 gives the mean recall percentages and standard deviations for each condition, and the total percent release for each list length, using the Watkins and Watkins (1976) formula (see Section 1.5, p. 34 ). Figure 2.1 (overleaf) shows the plot of mean recall percentage for each list length, as a function of condition type (left-hand panel), and the final percent release for each list length (right-hand panel).

Table 2.1 - Experiment 1: Mean recall and release percentages as a function of copy-list length

| | 6 words | | 12 words | |
|---|---|---|---|---|
| | SAME | DIFFERENT | SAME | DIFFERENT |
| Mean recall | 61.5% | 67.7% | 49.5% | 65.1% |
| Standard deviation | 17.6 | 17.4 | 14.8 | 13.2 |
| % Release | 10.1% | | 31.5% | |

Increasing the length of the copy list from 6 to 12 words does seem to significantly affect the total number of words recalled in the SAME condition (decreasing recall from 61.5% to 49.5%), but leaves recall in the DIFFERENT condition largely unaffected (67.7% and 65.1%, respectively). This pattern of results is also reflected by the very

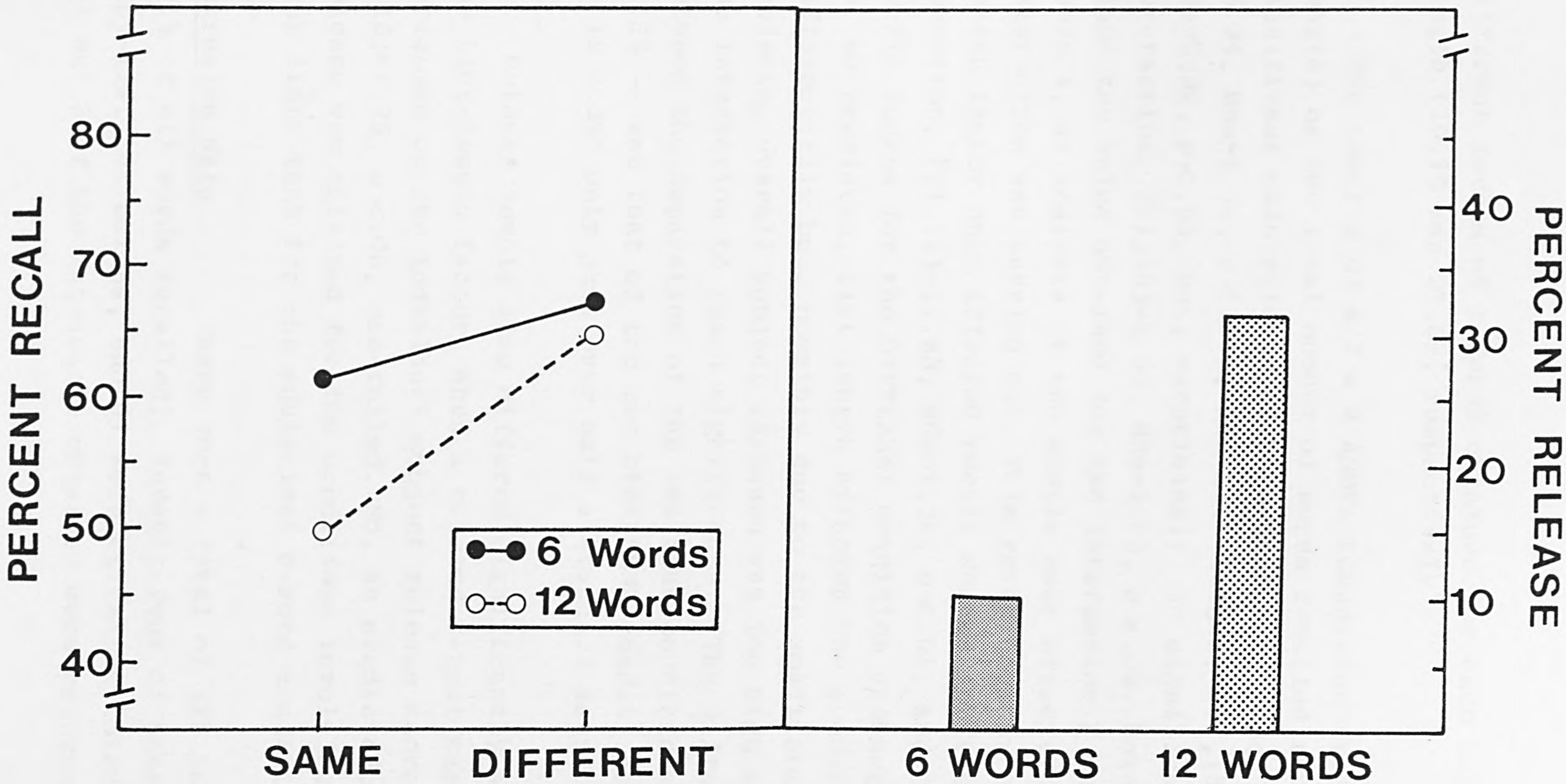



Figure 2.1 - Experiment 1: Percent recall and release from retroactive inhibition as a function of experimental condition and list length.

different levels of release obtained for each copy-list length (10.1% and 31.5%, respectively).

The results of a 2 x 2 ANOVA (Condition type x List length) on the total number of words recalled revealed significant main effects of both Condition type, $F(1,15)=25.94$, $MSe=1.06$, $p<.001$, and List length, $F(1,15)=5.53$, $MSe=2.22$, $p<.05$, but, surprisingly, no significant interaction, $F(1,15)=2.50$, $MSe=2.03$, $p>.05$. However, since the value obtained for the interaction F was well above 1, an analysis of the simple main effects of the interaction was carried out. This revealed that the List-length factor only affected recall scores for the SAME condition, $F(1,15)=11.83$, $MSe=1.39$, $p<.01$, and left recall scores for the DIFFERENT condition unchanged, $F<1$. So, as predicted, list length affected the conditions differentially but, possibly due to the small number of subjects, overall subject variance was too high to allow the interaction to reach significance. (The difference between the separation of the two SAME conditions -- 11.9% -- and that of the two DIFFERENT conditions -- 2.6% -- is 9.3%: only just over half a standard deviation.)

Release levels also differed significantly across the List-length factor, when a related t-test was performed on the individual subject release scores, $t(15)=1.79$, $p<.05$, one-tailed. So, as predicted, greater release was obtained for the conditions involving 12-word copy lists than for the equivalent 6-word conditions.

Intrusion data There were a total of 125 intrusions (21% of all words recalled). Seventy-four of these were copy-list intrusions, and 51 were extralist intrusions. All but 9 of the intrusions obtained were category-appropriate. Table 2.2 gives the breakdown of intrusion data for each experimental condition.

Table 2.2 - Experiment 1: Breakdown of intrusion data.

| Source | Condition | | | |
|---|---|---|---|---|
| | SAME | | DIFFERENT | |
| | 6 words | 12 words | 6 words | 12 words |
| Copy-list intrusions | 26 | 39 | 7 | 2 |
| Extralist intrusions | 8 | 5 | 16 | 22 |
| Total | 34 | 44 | 23 | 24 |

As the table shows, the SAME condition appears to have produced many more copy- than extralist intrusions, while the reverse was true for the DIFFERENT condition (65 vs. 13 in the SAME condition; 9 vs. 38 in the DIFFERENT condition), a finding which is not surprising, considering that copy-list intrusions in the DIFFERENT condition are intrusions from a category other than the memory list. The 9 copy-list intrusions that did occur in this condition were produced instead of words from the memory-list category, i.e., subjects mistakenly recalled the copy-list category. Overall intrusion rate appears to be slightly higher in the SAME (12) than in the SAME (6) condition (44 vs. 34 words), mainly because of the large difference in the total number of copy-list intrusions produced in the former condition (39 vs. 26). In the DIFFERENT condition, however, overall intrusion rate appears to be the same across copy-list length, even though slightly more extralist intrusions were produced with long copy lists (22 vs. 16).

Separate ANOVAS (List length x Intrusion type) were performed on the individual subject intrusion data for each condition (SAME and DIFFERENT). The results of these analyses support well the above interpretation. Considering first the SAME condition, the ANOVA revealed no main effect of copy-list length, $F(1,15)=1.60$, $MSe=0.97$, $p > .05$; but a main effect of Intrusion type, $F(1,15)=27.61$, $MSe=1.53$, $p < .001$, indicating that more copy- than extralist intrusions had been produced in this condition. The interaction was also significant, $F(1,15)=15.32$, $MSe=0.31$, $p < .001$, showing that increasing copy-list length also produced more intrusions from the copy list (39 compared with 26).

The ANOVA for the DIFFERENT condition again showed no main effect of list length, $F < 1$, but a main effect of intrusion type, $F(1,15)=7.14$, $MSe=1.84$, $p < .025$, showing that more extra- than copy-list intrusions were produced in this condition. The interaction was <u>not</u> significant, $F(1,15)=4.20$, $MSe=0.45$, $p > .05$, indicating that for the DIFFERENT condition increasing copy-list length had no effect on intrusion rate, either from the copy list or from words not used in the experiment.

The significant interaction between intrusion type and copy-list length in the SAME condition would seem to indicate that increasing list length made it more difficult for subjects to discriminate between targets and distractors -- the more the distractors, the more difficult the subjects found it to decide whether a given word had indeed been a memory list member. However, because of the generally low number of intrusions, it is unlikely that this can <u>fully</u> account for the effects on recall of increasing copy-list length.

### 2.2.4 Discussion

The results of this experiment support the hypothesis, derived from Cue-Overload theory, that increasing copy-list length should affect recall differentially in the SAME and DIFFERENT conditions of the Method of Interpolated Attributes. Recall in the SAME condition was lower when subjects had to copy down a 12-word copy list (total cue load = 18 words) than when they copied a 6-word list (total cue load = 12 words). Increasing list length did not affect recall in the DIFFERENT condition, where cue load was equivalent across list length (6 words, all of them targets). This pattern of results was reflected too in the significantly different levels of release obtained for each list length.

The results also replicate the Watkins and Watkins' (1976) finding of release from retroactive inhibition with a taxonomic category shift -- the other aim of the experiment.

Nevertheless, the intrusion data indicate that factors other than just simple overload may contribute to the outcome of experiments using the Method of Interpolated Attributes. These data showed that a subject's ability to discriminate between the memory and copy lists may also affect final recall level. Where discrimination is difficult, leading to high numbers of intrusions (e.g., the SAME (12) condition), recall may be prematurely stopped when the subject's written recall matches the known length of the memory list. In addition, the intrusions themselves overload the recall cue, and further contribute to poor recall (cf. the explanation offered in Section 1.2.1, p.24 , for the difference between release levels in the release from proactive and retroactive inhibition paradigms, and the interpretation Cue-Overload offers for the inhibitory effects of part-set cueing, in the same section, p.21 ).

## 2.3 Experiment 2 - Retroactive inhibition and the pre-recall cueing of subcategories.

### 2.3.1 Introduction

Gardiner, Craik and Birtwistle (1972) showed that it is possible to manipulate release from proactive inhibition, for the same set of materials, by providing some subjects with additional pre-recall cues specifying a subset of the general category from which list items had been selected. They used two semantic categories (sports and flowers), and two of their logical subdivisions (indoor and outdoor sports; wild and garden flowers). A typical trial consisted of the presentation of 3 Brown-Peterson-type word triads, all drawn from the same category subset, followed by a fourth triad, drawn from the alternative category subset. Recall was tested after each triad. The manipulation of interest concerned the information the subjects were given regarding the materials used for the shift triad. The subjects were divided into three groups: Group I received the general category name as a retrieval cue; Group II received the subset category title immediately before presentation of the 4th triad, and Group III received the same title after trial 4, but before recall. With appropriate counterbalancing of materials, all subjects were shown the same lists, so that encoding was equivalent for all groups right up to the 4th trial; and continued to be equivalent for Groups I and III up until the final recall test. The results of the experiment showed that, for the same materials, release from proactive inhibition only obtained for Groups II and III. Group I, who were given only the general category name, showed no release when words on trial 4 were shifted from one subcategory to the other. Groups II and III showed equivalent amounts of release.

The pattern of results obtained showed that interpretations of proactive inhibition build-up exclusively in terms of progressively poorer encoding of items sharing the same attribute could be discounted. Encoding was equivalent for subjects in Groups I and III, yet Group III showed release when given additional cues. Instead, Gardiner et al. opted for a retrieval interpretation of release from proactive inhibition. Two possible retrieval explanations were proposed, one based on increased accessibility of items cued with the subcategory title, the other based on increased discriminability of the freshly cued items. In a footnote to the paper, E. Tulving was said to favour this latter explanation, pointing out that, by the time subjects had to retrieve the last batch of words, the general category cue had to a large extent lost its effectiveness. This meant that current words became very difficult to distinguish from previous, already tested items, and so tended to show poor recall. Thus the provision of a fresh, highly efficient selection device, in the shape of the subcategory cue, improved recall by enabling the subjects to discriminate better between target words and already-tested items.

A Cue-Overload interpretation of the Gardiner et al. results would attribute the release effect to the increased accessibility of items cued with the subset title. Because the subset cue is less overloaded than the general category cue, it will be a more efficient retrieval prompt. In this form, the Principle can readily account for the results of Groups I and II, where the subset cue was made explicit before list presentation. The results for Group III (who received the subset cue only after having encoded the items in the 4th triad as members of the general category) cannot be explained simply in terms of accessibility, though.

Because items were encoded as members of the general category, it is likely that they were retrieved via the same cue (cf. Tulving & Thomson , 1973, Encoding Specificity Principle). The subset cue thus improved recall, in Group III, by providing the subjects with a means of discriminating between target words (the only ones to which the subset cue applied) and previously-tested items. Alternatively, even if subjects used the subset cue directly at the retrieval stage -- for instance, used it to generate subset category members -- so that the overloaded general category cue was not used, they would still need to decide whether the words they generated had, in fact, been in the 4th triad, or were extralist intrusions. A discrimination process must still have been involved (cf. Experiment 1, in this thesis, where it also seemed likely that "overload" reflected, in part, difficulties in discrimination between targets and intrusions).

Experiment 2 was based on the design used by Gardiner et al. (1972). The rationale behind the experiment was that, if Cue-Overload does indeed offer a comprehensive account of build-up of, and release from inhibition, then the results obtained by Gardiner et al. ought also to occur in the release from <u>retroactive</u> inhibition paradigm. The experiment was carried out using the Method of Interpolated Attributes. Both memory and copy lists were drawn from the same general category, but each list was made up of words from a different logical subset of the general category. SAME and DIFFERENT conditions were obtained by either providing the subset category name prior to recall (DIFFERENT, since different cues apply to each list) or by simply repeating the general category name prior to recall (SAME, since both lists share the same cue). Thus the experiment tried to replicate the results obtained for Groups I and III of the Gardiner et al. experiment.

### 2.3.2 Method

Materials and Design     A total of 6 categories -- flowers, fish, sports, vegetables, birds and trees -- were selected from the Battig and Montague (1969) norms, with the only constraint that it should be possible to subdivide each into two subsets which could be clearly defined. The division was carried out on a purely intuitive basis by three judges, who also agreed on the titles to be given to each subcategory: garden and wild flowers, indoor and outdoor sports (as used by Gardiner et al., 1972), freshwater and saltwater fish, green and root vegetables, edible and predatory birds, and evergreen and deciduous trees. The same judges then selected six words for each subcategory, again using the Battig and Montague norms. The only criteria used in this choice were that the words should be good examples of the subset in question and not applicable as examples for the complementary set. A full list of materials is enclosed in Appendix Two. The lists were stencilled in capitals on 6" x 4" flashcards, one subset per card, in one single column.

A within-subjects design was used, with each subject doing a total of 6 trials, half in the SAME (free recall) and half in the DIFFERENT (cued) condition. Materials were fully yoked across subjects, such that each subset list was used equally often as a memory and as a copy list, in the SAME and in the DIFFERENT condition. Condition order was randomized separately for each pair of subjects with the sole constraint that the same condition should not occur in more than two consecutive trials. Within each pair of subjects, one subject received one random condition order, and the second subject its complementary series. Category order was kept constant across all subjects.

Subjects     Twenty-eight undergraduate and graduate students at The City University served as paid volunteers. All subjects were native English speakers.

Procedure     The instructions given to the subjects followed closely those for Experiment 1, but subjects were also informed that, to help them recall the memory list better, before each pair of lists was presented the experimenter would tell them the name of the category from which the words had been drawn. After list presentation, the subjects were instructed to wait before proceeding to the recall test, as the experimenter would either refresh their memory as regards the category name, or give them "more information" about the words in the memory list, to aid their recall.

Lists were presented manually by the experimenter. Each memory list remained on display for six seconds before being replaced by the copy list. The copying task was again self-paced. Prior to recall, the experimenter picked up the recall booklet, turned over and tucked under the page containing the subject-written copy list and, before returning the booklet to the subject, said aloud the recall cue (e.g., "All the words in the first list were garden/wild flowers" -- DIFFERENT condition --, or "All the words in the first list were names of flowers" -- SAME condition). Recall was written and unconstrained, and lasted approximately 45 seconds. One experimental session (6 trials) lasted about 20 minutes.

### 2.3.3 Results

Table 2.3 shows the mean recall percentages, and respective standard deviations, for the two experimental conditions. Results are given both broken down by category, and collapsed across category, as the added scores for the three categories tested under each condition, for each subject. The overall result is therefore based on a total of 504 observations per condition (28 subjects x 3 categories x 6 targets).

Table 2.3 - Experiment 2: Mean recall percentages as a function of cueing condition

| Category | | Condition | |
|---|---|---|---|
| | | SAME (not cued) | DIFFERENT (cued) |
| Birds | Mean | 54.8 | 60.7 |
| | S.D. | 19.3 | 25.7 |
| Fish | Mean | 54.8 | 53.6 |
| | S.D. | 26.3 | 18.0 |
| Flowers | Mean | 50.0 | 45.2 |
| | S.D. | 19.9 | 23.1 |
| Sports | Mean | 58.3 | 61.9 |
| | S.D. | 18.6 | 17.2 |
| Trees | Mean | 58.3 | 65.5 |
| | S.D. | 20.7 | 21.3 |
| Vegetables | Mean | 54.8 | 72.6 |
| | S.D. | 24.7 | 18.4 |
| Overall | Mean | 55.4 | 59.9 |
| | S.D. | 17.5 | 16.1 |

The Watkins and Watkins (1976) formula was used to calculate the amount of release obtained within and across categories. Table 2.4 lists these results.

Table 2.4 - Experiment 2: Percent release within and across category.

| Category | % Release |
|---|---|
| Birds | 10.8% |
| Fish | - 2.2% |
| Flowers | -10.0% |
| Sports | 6.2% |
| Trees | 12.3% |
| Vegetables | 32.5% |
| Overall | 8.2% |

Collapsed across category, the results appear to go in the predicted direction. Recall was higher in the DIFFERENT than in the SAME condition (59.9% compared with 55.4%) and release was obtained when subcategories were cued (8.2%). Individual category data are less clear, though, since two categories (flowers and fish) actually showed better recall when not cued, leading to negative release. One possible interpretation is that, for some reason, for these categories subjects either guessed the subset cue (and hence eliminated the difference between the two conditions) or had created their own subset cues, which were different from those used by the experimenter in the DIFFERENT condition. This would have promoted a certain amount of confusion in this condition, when another subset cue was provided, and could explain the lower recall, and negative release, obtained.

The results were analysed collapsed across the pairs of subjects used for yoking categories with condition and condition presentation-order (see Section 2.3.2). A 2 x 6 ANOVA (Condition x Category) showed a significant main effect of Condition, $F(1,13)=4.76$, $MSe=0.72$, $p < .05$; a significant main effect of Category, $F(5,65)=2.78$, $MSe=1.26$, $p < .05$; and no interaction, $F(5,65)=1.13$, $MSe=1.36$, $p > .05$. Thus, even though there were wide variations in the level of category recall, the overall effect of cueing was significant, as predicted.

Intrusion data There were a total of 75 intrusions (11% of all words produced). 41 were copy- and 34 were extralist intrusions (all category-appropriate). Table 2.5 shows the distribution of copy- and extralist intrusions across the two experimental conditions.

Table 2.5 - Experiment 2: Breakdown of intrusion data.

| Source | Condition | | Total |
|---|---|---|---|
| | DIFFERENT | SAME | |
| Copy-list intrusions | 13 | 28 | 41 |
| Extralist intrusions | 21 | 13 | 34 |
| Total | 34 | 41 | 75 |

Of the 21 extralist intrusions produced in the DIFFERENT condition, 17 were subcategory-appropriate and 4 were not. The equivalent figures for the SAME condition were 8 and 5.

A 2 x 2 ANOVA (Condition x Intrusion type), again collapsed across the pairs of subjects used for yoking category and condition order, revealed a significant interaction, $F(1,13)=4.85$, $MSe=1.95$, $p < .05$, and no significant main effects of either factor ($F < 1$ in both cases). Thus extralist intrusions were more frequent when subjects received a subcategory cue than when they did not, and when using the general category label for recall subjects produced more copy- than extralist intrusions. This pattern of results extends to this paradigm the interpretation offered in the introduction section for the results of Group III in the Gardiner et al. (1972) experiment: Subjects appear to have used a subcategory cue to discriminate between memory and copy list words (fewer copy-list intrusions when the subcategory cue was provided), and in some cases may also have simply generated words to the subcategory cue and then used other cues to distinguish between memory and extralist words (most extralist intrusions were category <u>and</u> subcategory-appropriate). Recall in the SAME condition appears to have been mediated by the general category cues, as evidenced by the greater number of copy-list intrusions in this condition, and the fact that those extralist intrusions which did occur were about equally distributed between subcategory-appropriate instances, and subcategory-inappropriate ones.

### 2.3.4 Discussion

The experiment replicated, in the release from retroactive inhibition paradigm, the pattern of results obtained by Gardiner et al. (1972, Groups I and III). It showed that it is possible to obtain release from retroactive inhibition (using the same set of materials for both conditions) when post-presentation cues are provided which specify a subset of the general category from which items were obtained.

However, though the experiment provided further, converging evidence for a common explanation of release from proactive and retroactive inhibition in terms of Cue-Overload, it also demonstrated that at least part of the effects attributed to cue overload in both paradigms may be due to problems of discrimination between target words and distractors. In this experiment, the intrusion data suggest that provision of subcategory cues made the subjects adopt a strategy of first generating words for recall (either to the general category cue or to the subcategory title) and then using either the subcategory cue, or additional task-specific cues, to discriminate between memory and copy lists, or between memory and extralist items. So even though the load on the general category cue was the same for both experimental conditions, and that cue was overloaded (recall was only about 50% of a subspan memory list), it was possible to improve recall somewhat by using other, non-overloaded cues to generate items or to differentiate between targets and intrusions.

The intrusion data also show that some of the subcategories used may have been unfamiliar to the subjects, so that in some cases the cues given prior to recall may have hindered, rather than facilitated, performance. For instance, telling a subject that memory list items were all names of wild flowers, when the subject has no very clear idea of which flowers are wild and which are cultivated, does not provide any additional information to help the subject retrieve the memory list. In addition, if the subject then goes on to try to use the subcategory name (wild flowers) to generate items for recall, his performance may actually be poorer than if no cue had been supplied, since he would not be able to produce many wild flowers (not knowing much about the subcategory) and any intrusions produced will only contribute to overload the general category cue even more. It is the case that the two categories for which slightly negative release was obtained were also the two categories which contributed most subcategory-inappropriate (copy-list) intrusions in the DIFFERENT condition: Flowers, 6 and fish, 3 copy-list intrusions. Of the remaining categories, trees contributed 2 copy-list intrusions, and sports and birds one each. Thus some of the subcategories used may have assumed too much knowledge on the part of the subjects.

## 2.4 Experiment 3 - Typicality and release from retroactive inhibition

### 2.4.1 Introduction

Experiment 1 looked at the effects on recall of increasing degree of overload on a category cue by increasing the length of the copy list, leaving unchanged the length of the memory list. Experiment 3 involved a similar procedure but, rather than manipulating cue loading by varying the number of items in the copy list, it varied instead the degree to which list items were associated with the category recall cue and with one another. Copy-list words were either strongly associated with the category cue and with one another (all very typical instances of the category), or weakly associated with it and with other list words (all very atypical instances of the category). In addition, a third condition used copy-list words unrelated with the memory list category -- i.e., it involved a taxonomic category shift.

As formulated at present (see Section 1.2, p.17 ), the Cue-Overload Principle would predict equal interference by both typical and atypical words in this experiment, so long as both types of distractor are encoded as members of the same category as the memory list. If it is simply the total number of items subsumed by the retrieval cue that matters, then the category-cue loading should be the same regardless of the typicality of the category members chosen for the copy list, and should lead to similar levels of recall and release in each case.

However, as mentioned in Section 1.4, it has been argued that some of the typicality effects reported in the literature may reflect differential encoding for typical and atypical words with regard to the experimenter-chosen categories (e.g., Loftus, 1975). Atypical words may simply be more typical members of other categories, and hence may be encoded as members of those, rather than the experimenter-chosen categories. Should this be the case, in the present experiment, it would be difficult to interpret the results obtained, since subjects would be using different cues for retrieval of atypical items. It is thus essential to make sure that the cues the subjects use to encode the memory list will apply equally well to both typical and atypical copy-list words subsequently presented. In order to achieve this, Experiment 3 used a modified version of the Method of Interpolated Attributes, which will now be described.

The original Watkins and Watkins (1976) design involved presenting the subjects with a homogeneous memory list (all items sharing the same attribute to the same extent), followed by another, homogeneous copy list. So, in an experiment like this one, which looked at the effects of prior strength of association among items and between these and their retrieval cue, both memory and copy list composition should be homogeneous with respect to typicality. However, Loftus would argue that a homogeneous typical memory list would lead to the priming of the experimenter-chosen category, but a homogeneous atypical memory list would not, since atypical words may also belong better in other categories, rendering the list heterogeneous. The possible effects of a homogeneous typical or atypical copy list would thus be restricted entirely to the typical memory list. This would happen not because of typicality differences as such, but because the subjects had failed to access the category in question when shown atypical memory-list words. To this extent, subjects receiving atypical memory lists would be using different cues from those subjects tested with typical memory lists, making it difficult to interpret the results unambiguously in terms of Cue-Overload.

One way to circumvent the problem described above is to make sure that, with presentation of the memory list, all subjects access the same general category representation (typical and atypical members), so that any effects of a homogeneous copy list are restricted to typicality differences within the same category, i.e., with the same general cue. To this end, the memory lists used in Experiment 3 contained equal numbers of both typical and atypical words, in random order (and hence were heterogeneous with respect to typicality), and were paired with homogeneously typical or atypical copy lists.

The experiment thus looked at whether typical or atypical words load category retrieval cues to the same extent, leading to similar amounts of release from retroactive inhibition being obtained when the category membership of the copy list is changed. Because the memory list was made up of equal numbers of typical and atypical items, it was also possible to test the overall effect of typicality on recall -- an effect which, as mentioned in Section 1.4, has not been very extensively researched.

### 2.4.2 Method

Design and Materials The experiment used the modified version of the Method of Interpolated Attributes described in the Introduction. The length of the copy list was also increased from 6 to 8 words to increase the degree of interference (cf. Experiment 1; see also Section 1.2.1, p.19 ).

A within-subjects design was used, consisting of three experimental conditions obtained by combining the mixed-typicality memory lists with three possible types of copy list: Same category - typical words (SAME-TYPICAL); Same category - atypical words (SAME-ATYPICAL), and Different category (DIFFERENT). The subjects were given six trials in all, two per condition for six different stimulus categories.

Memory-list category membership was yoked across all conditions, such that each memory list was paired an equal number of times with a SAME-TYPICAL, a SAME-ATYPICAL, and a DIFFERENT copy list. The two categories making up the DIFFERENT copy lists (and only these) were fully rotated such that each of them was paired once with each memory list. Copy list composition was kept constant for all subjects, but two versions of each memory list were used. Half the subjects received one random ordering of typical and atypical memory-list words, for each category, and the other half received the complementary sequence, using the serial position of each typicality class as the distinguishing feature between both sets of memory lists (i.e., where a typical word had been, in set "a", set "b" had an atypical word, and vice-versa). Condition presentation order was randomized separately for each subject.

Six categories were selected from Rosch's (1975a) norms of typicality: vegetables, sports, fruit, birds, vehicles and clothing. Twenty-two words were chosen from each category, 11 from the typical end of the scale and 11 from the atypical end. The words from the first five categories were then cross-checked with Hampton's (1976) ratings for degree of category membership, obtained for a British sample of subjects, and appropriate changes were made to allow for British spelling and usage. Where large discrepancies were seen to exist between the two sets of norms in the typicality and category-membership ratings given to particular words, those words were replaced, using words from Hampton's norms whenever Rosch's norms could not supply additional suitable replacements. The sixth category (clothing) was not included in Hampton's norms. Rosch's ratings were used, and changes to allow for British usage were made on an intuitive basis (e.g., "vest" was replaced with "waistcoat"). In the final sample, Rosch's ratings were available for 124 of the 132 words used, and Hampton's ratings for 77 of the total word pool. Seven of the 8 words for which ratings of typicality were not available were in Atypical lists, and were selected on an intuitive basis by the experimenter and two other people. Hampton's ratings were available for only two of these words. The average rated typicality (Rosch, 1975) for typical word lists was 1.65 (S.D.=.47, range=1.02-3.30), and mean Hampton category-membership scores for the same lists (treating the scores for 25 words as missing values) were 98.5 (S.D.=2.31). The corresponding figures for atypical lists were: Rosch, mean typicality 3.49 (S.D.=.99, range=1.90-6.21); Hampton, mean category-membership scores 84.33 (S.D.=12.63). (*)

---

(*) Experiment 4 obtained ratings of typicality for a 531-word corpus. When the two lists of words used in Experiment

Eight words were selected at random from each set of 11 typical or atypical category members, to make up the copy lists for the two SAME conditions (TYPICAL and ATYPICAL). The remaining six words per category were shuffled and used as items for the memory lists. As mentioned earlier, two memory lists were constructed for each category, with two complementary sequences of typical and atypical items for each category.

Two additional categories were used to make up copy lists for the DIFFERENT condition: weapons and insects. Rosch's (1975a) norms were used for the first category, and new ratings were obtained for the second category. DIFFERENT copy lists (8 words each) were made up of a mix of typical and atypical words, with a slightly higher proportion of typical instances. Appendix Three contains copies of the lists used in the experiment.

All lists were printed in one single column on 8" x 5" flashcards, in stencilled capitals, for manual presentation by the experimenter.

Subjects Forty-eight students at The City University and at Trinity College of Music, London, were tested individually and were paid for their assistance. All were native English speakers.

---

(* continued) 3 were re-analysed using the new ratings, they were found to differ as designed in their degree of typicality. See Section 3.5, p.122 .

Procedure The subjects were given verbal instructions describing the basic Method of Interpolated Attributes procedure, similar to those used in Experiments 1 and 2. It was emphasized that the copying task should be given as much attention as the learning task, and the subjects were asked not to rush through it, scribbling illegibly. Prepared booklets were provided, containing alternate pages labelled "Copy" and "Recall", each pair of sheets separated by one page containing a distractor task: one 3-digit number at the top left-hand corner, which the subjects had to read aloud prior to beginning counting backwards by threes from that number, saying each number aloud and writing it down simultaneously. This distractor activity lasted for one minute after recall, for all trials except the last.

After the subjects were fully acquainted with the general procedure, the experimenter initiated the first trial by manually displaying the card containing the first memory list. This card remained visible for 6 seconds, after which the copy-list card was presented for the subject-paced copying task. The removal of the copy-list card was the signal for recall, which was written and unconstrained. It lasted typically about 45 seconds and ended when the subject turned over the page to do the -- timed -- distractor task. The same procedure was followed for all 6 trials. One experimental session lasted about 25 minutes.

### 2.4.3 Results

The results were obtained by calculating separately the percentages of typical and atypical memory-list words recalled in each of the three experimental conditions, collapsed across category. Figure 2.2 (overleaf) shows them in graph form. Each point on the graph is based on a total of 288 observations (48 subjects x 2 categories/ condition x 3 target words). Table 2.6 lists the mean recall percentages and respective standard deviations for each condition, and Table 2.7 shows the mean release percentages for the 4 possible combinations of memory- and copy-list typicality (SAME condition) compared with the same-typicality recall in the DIFFERENT condition.

Table 2.6 - Experiment 3: Mean recall percentages as a function of experimental condition.

| Source | | Condition | | |
|---|---|---|---|---|
| | | SAME TYPICAL | SAME ATYPICAL | DIFFERENT |
| Typical target items | Mean | 62.5 | 54.5 | 70.1 |
| | S.D. | 21.7 | 21.7 | 22.0 |
| Atypical target items | Mean | 54.5 | 58.7 | 62.5 |
| | S.D. | 20.0 | 18.8 | 21.3 |

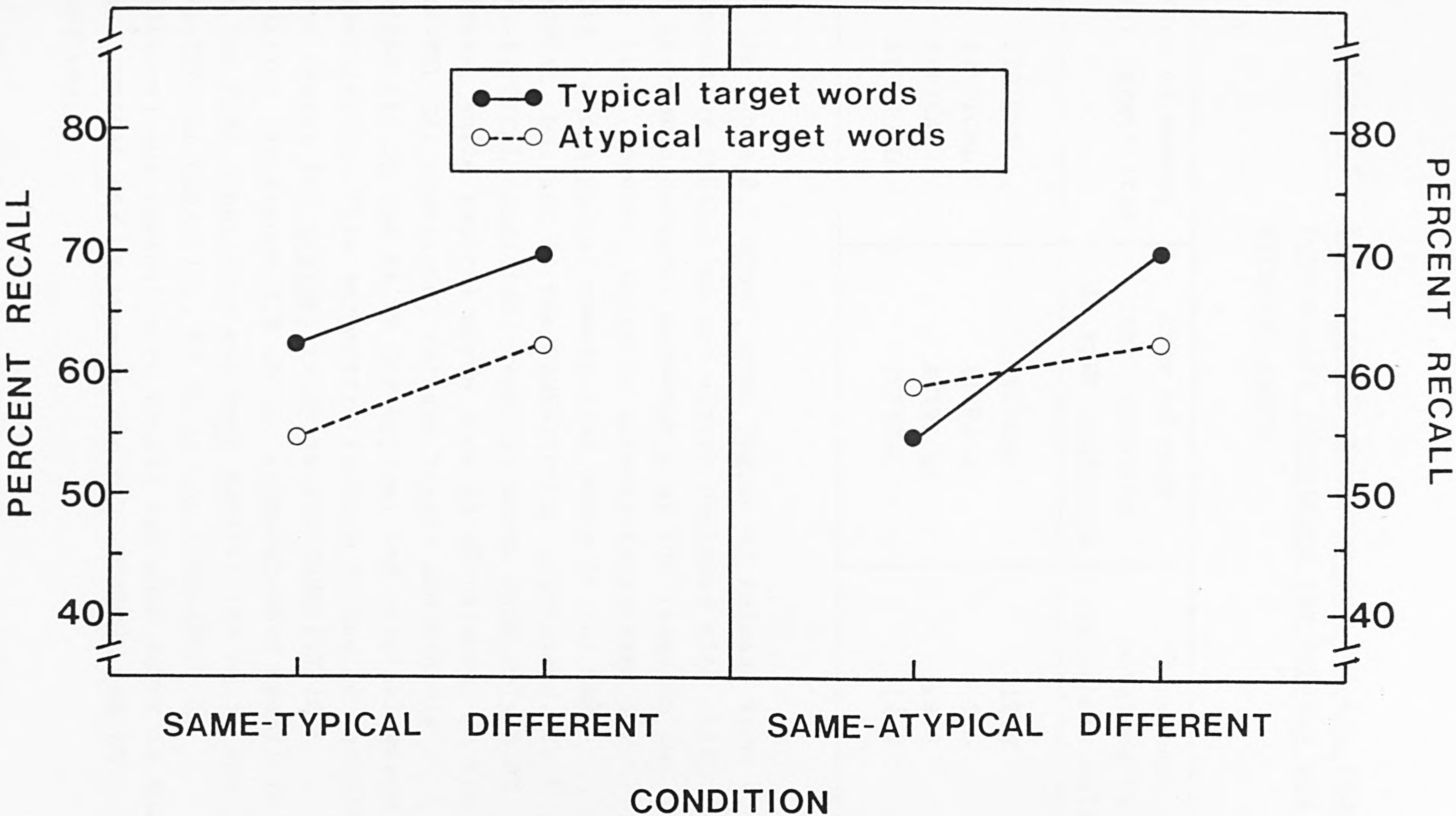



Figure 2.2 - Experiment 3: Percent recall for typical and atypical words as a function of experimental condition.

Table 2.7 - Experiment 3: Mean percent release from retroactive inhibition for typical and atypical items.

| Type of memory list target item | Type of copy list inhibitor in SAME condition | % release obtained by taxonomic shift |
|---|---|---|
| TYPICAL | TYPICAL | 12.2 |
| ATYPICAL | ATYPICAL | 6.5 |
| TYPICAL | ATYPICAL | 28.6 |
| ATYPICAL | TYPICAL | 14.7 |

As Table 2.7 shows, some degree of release from retroactive inhibition was always obtained with shifts in the taxonomic category membership of the items in the copy list. However, there is asymmetric release for typical and atypical memory-list words in the SAME-ATYPICAL, but not in the SAME-TYPICAL conditions. In the SAME-ATYPICAL condition, atypical words show only 6.5% release, while typical words show 28.6% release. In the SAME-TYPICAL condition, release levels are broadly similar (12.2% and 14.7% for typical and atypical words, respectively). This asymmetric release is due to the much lower recall for typical words in the SAME-ATYPICAL condition. As Figure 2.2 shows, atypical-word recall in the two SAME conditions was very similar (54.5% in the SAME-TYPICAL condition, 58.7% in the SAME-ATYPICAL condition) but typical-word recall was much lower in the SAME-ATYPICAL (54.5% than in the SAME-TYPICAL (62.5%) condition.

The results of a 3 x 2 ANOVA (Copy-list type x

Target-word typicality) of the recall scores showed main effects of both factors, and no interaction: Copy-list type, $F(2,94)=11.84$, $MSe=0.90$, $p < .001$; Target-word typicality, $F(1,47)=6.42$, $MSe=1.46$, $p < .025$; Interaction, $F(2,94)=2.09$, $MSe=1.44$, $p > .05$.

Separate analyses of the recall data for conditions SAME-TYPICAL vs. DIFFERENT (left-hand panel of Figure 2.2) and SAME-ATYPICAL vs. DIFFERENT (right-hand panel) revealed the following patterns of results:

(a) <u>SAME-TYPICAL vs. DIFFERENT comparison</u> There were significant main effects of Copy-list type (SAME or DIFFERENT category) and of Target-word typicality: $F(1,47)=17.00$, $MSe=0.62$, $p<.001$ and $F(1,47)=5.55$, $MSe=1.90$, $p< .025$, respectively; and no interaction ($F < 1$). Thus, with a typical copy list in the SAME condition, the release from retroactive inhibition effect was equivalent for both levels of typicality, even though typical words were better recalled than atypical words (typical-word release=12.2%, atypical-word release=14.7%).

(b) <u>SAME-ATYPICAL vs. DIFFERENT comparison</u> There was a significant main effect of Copy-list type, $F(1,47)=16.33$, $MSe=1.28$, $p < .001$, and a marginally non-significant interaction, $F(1,47)=3.87$, $MSe=1.56$, $.05 < p < .10$. The typicality main effect had an F of less than one. Thus the ANOVA results confirm the asymmetric release effect in this condition: Atypical copy-list words interfered more with typical than with atypical memory-list words (26.6% and 6.5% release, respectively).

Taken together with the results of the overall ANOVA, these individual pairwise comparisons also show that a typical-word recall advantage was obtained only in the

SAME-TYPICAL and DIFFERENT conditions, whereas in the SAME-ATYPICAL condition typical and atypical words had an about equal probability of being recalled.

Intrusion data There were a total of 250 intrusions (19% of all words produced). Table 2.8 gives the breakdown of these intrusions by condition (SAME-TYPICAL, ATYPICAL and DIFFERENT) and source (copy-list or extralist intrusions).

Table 2.8 - Experiment 3: Breakdown of intrusion data.

| Condition | Source | | Total |
|---|---|---|---|
| | Copy-list intrusions | Extralist intrusions | |
| SAME-TYPICAL | 86 | 25 | 111 |
| SAME-ATYPICAL | 56 | 36 | 92 |
| DIFFERENT | 4 | 43 | 47 |
| Total | 146 | 104 | 250 |

As Table 2.8 indicates, more copy- than extralist intrusions were produced in the two SAME conditions (TYPICAL, $t(47)=5.00$, $p < .001$, two-tailed; ATYPICAL, $t(47)=1.99$, $p < .05$, two-tailed) and more extra- than copy-list intrusions occurred in the DIFFERENT condition (43 and 4, respectively). The four copy-list intrusions in the DIFFERENT condition were the only intrusions produced which were not category-appropriate. They occurred when one subject mistakenly recalled the copy-list category, instead of the memory-list category.

Further analysis also showed that the two SAME conditions produced equivalent levels of extralist intrusions, $t(47) = 1.45$, $p > .05$, but copy-list words were more intrusive when they were typical than when they were atypical members of the category: SAME-TYPICAL vs. SAME-ATYPICAL, $t(47)=3.84$, $p < .001$, two-tailed.

Table 2.9 gives the breakdown of the extralist intrusions into typicality groups, using a procedure similar to that followed to select list materials: two non-overlapping samples with a cut-off point for typical words at around the 2.0 rating on the scale. Rosch's norms were used to allocate typicality ratings.

Table 2.9 - Experiment 3: Breakdown of extralist intrusions into typicality groups.

| Condition | Typical intrusions | | Atypical intrusions | | Absent from Rosch (1975) |
|---|---|---|---|---|---|
| | Total | Mean | Total | Mean | |
| SAME-TYPICAL | 6 | 1.35 | 14 | 3.04 | 5 |
| SAME-ATYPICAL | 20 | 1.59 | 3 | 2.28 | 13 |
| DIFFERENT | 22 | 1.33 | 11 | 2.94 | 10 |
| Total | 48 | | 28 | | 28 |

As the table indicates (and would be expected), more atypical than typical category instances were produced as intrusions in the SAME-TYPICAL condition, and the reverse was true for the SAME-ATYPICAL condition. However, since the number of words absent from Rosch's norms (and hence for which no typicality ratings could be obtained) was quite high, it is not advisable to place too much emphasis on the above pattern of results, and no analysis was carried out.

### 2.4.4 Discussion

The results of Experiment 3 are somewhat less clear than one would wish. Release from retroactive inhibition was obtained, for both typical and atypical memory-list words, with shifts in the taxonomic category membership of the copy list. However, this release was asymmetrical in that atypical copy-list words produced much greater inhibition (and release) for typical than atypical memory list members, even though typical copy lists produced equivalent levels of release across typicality. Finally, an advantage to typical-word recall was obtained for only two of the experimental conditions (SAME-TYPICAL and DIFFERENT).

The finding that typical memory-list words are better recalled in two of the conditions, coupled with the fact that for these the typicality advantage does not interact with the release effect, makes the pattern of results obtained in the SAME-ATYPICAL condition somewhat suspect. It could be that the large imbalance in the numbers of typical and atypical words involved in this condition made at least some subjects adopt "atypicality" cues (for instance, cues about the oddness of certain words as members of the category) for retrieval of memory-list words: there were a total of 11 atypical and only 3 typical words in the pairs of lists used in the SAME-ATYPICAL condition. These atypicality cues would obviously not be of much use for retrieval of typical category instances, leading to the sharp decline in typical-word recall obtained in this condition. Other subjects may have used the general category cues elicited with memory-list presentation for all condition of the experiment. Since these cues appear not to be very efficient for retrieval of atypical instances (cf. the results for the SAME-TYPICAL and DIFFERENT conditions), for these subjects atypical-word recall in the SAME-ATYPICAL condition would be poor. Hence the final result obtained in this condition, where typical and atypical words had an apparent equal probability of being recalled, could be due to the combined effects of these two recall strategies, one boosting recall of atypical words and reducing typical-word recall, the other having the converse effect.

Though other explanations are possible, the possibility that the asymmetric release effect could be due to the use of different cues in the SAME-ATYPICAL conditions means that the results, as they stand, should not be used to test a Cue-Overload interpretation of the effects of prior associative strength on recall.

The typical-word recall advantage obtained for the SAME-TYPICAL and DIFFERENT conditions is, however, interesting and warrants further research. For these two conditions, it is likely that subjects used the general category cue for recall, since in the SAME-TYPICAL condition there were a total of 11 typical and only 3 atypical words. Typical words are, by definition, closer to the representation of the category concept, or prototype, than atypical words are. If subjects did use a general category cue for recall, it is difficult to see how the typical-word recall advantage can be squared with a Cue-Overload account of recall. More typical than atypical words were subsumed by the category cue so, if anything, typical-word recall should have been lower, not higher, than atypical-word recall. However, there are again problems with using these results to argue against a Cue-Overload interpretation, since other factors, not controlled in this experiment, could have brought about the typical-word recall advantage. Glass and Meany (1978) suggested that there are in fact two kinds of atypical words: those that are well-known but unrepresentative (e.g., TOMATO as a FRUIT), and those which are not well-known and are classified as atypical largely because of this (e.g., PERSIMMON as a FRUIT). Familiarity was not equated across typicality level in the present experiment, so that lack of knowledge about some of the atypical words could have led to their poor recall (cf., too, McCloskey, 1980; and Section 1.4, p.31 of the present thesis). Similarly, if the typical instances could be more easily generated to the category name (cf. Collins & Loftus, 1975; Loftus, 1975), then the superior recall for typical words could have come about because of differences in production frequency.

Tighter control over experimental materials is hence necessary, if the typicality advantage is to be used as evidence against a simple Cue-Overload interpretation of these data. However, there is at present no set of normative data which systematically grades the same set of words on all three variables which could have contributed to the results described above and which, in addition, is based on a British sample of subjects. All available published norms of typicality, familiarity and associative frequency were collected in the USA, at different points in time and place, for different word sets. Translating these ratings for use in this country generally involves a certain amount of guesswork and even then there is no guarantee that the lists thus obtained really do reflect the differences implied by the ratings. A comprehensive set of normative data needs therefore to be collected before pursuing any further the line of research adopted in Experiment 3. Chapter Three describes the collection of such a set of norms.

The second major problem with Experiment 3 (besides materials) is the use of mixed-typicality memory lists. This aspect of the design needs modification before further research can be undertaken. The large imbalance between the numbers of typical and atypical words across condition, using this design, could have led to the changes in strategy described in the first part of this discussion section. It is thus necessary to find some way of ensuring that both typical and atypical words are encoded using equivalent sets of cues, without using mixed-typicality memory (or copy) lists. Chapter Four describes the results of three studies using different designs, intended to take account of this problem.

## 2.5 General discussion and conclusions

The three experiments reported in this chapter show (a) that cue load can be manipulated directly by increasing copy-list length, (b) that provision of non-overloaded cues can reduce the degree of cue overload and improve recall, and (c) that equal amounts of release from retroactive inhibition may be obtained for both typical and atypical words, even though typical words are better recalled than atypical words.

The first two results also show that a Cue-Overload account of release from retroactive inhibition using the Method of Interpolated Attributes needs to take account of differences in discriminability between items in the memory and copy lists. Experiment 1 revealed that part of the effects of increased copy-list length could be due to loss od discriminability between targets and distractors when copy-list length was increased. Experiment 2 showed that the additional, non-overloaded cues may have been used by the subjects to improve recall by allowing them to discriminate better between copy-list items (to which the cues did not apply) and memory-list items (to which they did).

The results of Experiment 3 are less clear cut. A typicality advantage in recall was obtained, and both typical and atypical words showed equivalent levels of release from retroactive inhibition with a taxonomic category shift (SAME-TYPICAL condition). However, before the results can be used to test a Cue-Overload interpretation, tighter control over materials is necessary and the design needs modification, so that it is possible to check whether the effects of copy-list typicality on recall probability are symmetrical for both levels of typicality. In Experiment 3, only the effects of typical copy-list words could be unambiguously interpreted, since it was possible that subjects in the atypical copy-list condition did not consistently use the same category cues used in the other conditions.

Chapter Three will now deal with the first of the problems discussed in Experiment 3. It describes the collection of normative data covering the three possible determinants of the typicality advantage obtained -- typicality, familiarity and production (or associative) frequency.

# CHAPTER THREE : NORMATIVE DATA

## 3.1 Introduction

The results of Experiment 3 highlighted the need to obtain better sources of materials before further research could be carried out into the effects of typicality on recall. This chapter describes the collection of normative data for three interrelated dimensions of internal category structure (Section 3.2) and analyses their interdependence (Section 3.3). Though not entirely germane to the original aims of the thesis, this analysis provides a more systematic test of the factors determining differences in typicality than is possible to do using any of the available published norms. Finally, the normative data collected are used to re-analyse the materials used in Experiment 3 (Section 3.5).

## 3.2 Experiment 4 - Collection of normative data.

### 3.2.1 Introduction

In Experiment 4, normative data for three inter-related dimensions of internal category structure were collected, so that the questions raised in the Introduction and Discussion sections of Experiment 3 can be investigated further. The aim of the experiment (and by extension of this thesis) was to determine the relationship between cue effectiveness and prior associative strength. The measure of associative strength chosen for investigation was the rated typicality of words within a category. This dimension is, in turn, associated with at least two other factors: familiarity with the words being rated (McCloskey, 1980; Malt & Smith, 1982) and associative frequency, the measure of the probability of an item being generated to the category name (Mervis, Catlin & Rosch, 1976; Hampton, 1979). Both these variables -- familiarity and associative frequency -- may affect level of recall independently of typicality differences. It was therefore desirable to determine a priori their respective contributions to recall of the lists used in later experiments (see also Section 1.4, p. 31, for a more detailed discussion of the theories associated with typicality differences).

Collecting normative data for typicality, familiarity and associative frequency for the same set of materials and for the same subject population also permitted a correlational analysis of the interdependence of the three measures to be carried out. This is described in Section 3.3.

### 3.2.2 Materials

Typicality, familiarity and associative frequency data were collected, in this order, for a set of 12 semantic categories. These included 8 of the 10 used by Rosch (1975a; "carpenter's tools" and "toys" were dropped because it was felt they overlapped extensively with other categories, such as "weapons", "vehicles", etc.) plus an additional four categories, chosen from Battig and Montague (1969), which had been used in previous experiments. The final sample contained the following categories: birds, clothing, fish, flowers, food flavourings, fruit, furniture, insects, sports, vegetables, vehicles and weapons. Lists of category items were compiled for each of the 12 categories by selecting between 34 and 55 words from Battig and Montague (1969) and Rosch (1975a), as well as including words used in other experiments and for which normative data were needed. Table 3.1 (overleaf) lists the total number of words in each category. The samples of words included a wide range of typicality and familiarity for each category, but words which, in the author's judgement, were clearly outside the category were not included.

Table 3.1 - Experiment 4: Sample size per category.

| Category | Sample size |
|---|---|
| Birds* | 52 |
| Clothing* | 55 |
| Fish | 37 |
| Flowers | 46 |
| Food flavourings | 40 |
| Fruit* | 43 |
| Furniture* | 41 |
| Insects | 34 |
| Sports* | 48 |
| Vegetables* | 41 |
| Vehicles* | 54 |
| Weapons* | 40 |
| Total | 531 |

(* - also present in Rosch, 1975a.)

### 3.2.3 Method

#### 3.2.3a Typicality norms

Subjects    Ninety-three subjects took part, of whom 71 were students at The City University, London, and 22 were at other London colleges. They were unpaid. About half the subjects were psychology students.

Design and Procedure    Testing booklets were prepared by typing all the items from each category in a random order in a column headed by the category name. Next to each word was the scale from one to six for the rating of typicality. The scale was explained at the top of each page. To reduce testing time to a convenient period, each subject rated six categories, sampled at random from the twelve, and presented in a random order. A paired subject then rated the remaining six categories, also in a random order. The instructions given to the subjects largely followed those used by earlier researchers, with the following modifications, introduced to make the task clearer and less ambiguous for the subjects:

(a) Subjects were given a separate response ("6" on the scale) for denying that an item belonged in a category.

(b) Subjects could also leave a rating line blank if they did not know a word.

(c) The instructions stressed, with an appropriate example, that simple frequency of occurrence should not be used as a basis for the typicality judgement.

The top page of each response booklet contained the following instructions:

"In the following pages you will find lists of items belonging to six different categories. The items are arranged by category and your task will be to rate each word according to how typical or atypical an instance it is of the category it belongs to. In other words, you have to decide whether each word is a good or a bad example of the category named. For instance, most people would say that Churches are very typical examples of the category Buildings; more typical than, say, Telephone boxes, which some people would classify as very atypical examples. The above example also serves to illustrate the fact that, just because a specific word is more typical than another, it does not mean that it occurs more often in your experience than an atypical word. Telephone boxes are probably seen much more often than Churches, but they are still less typical of the category Buildings than Churches are.

"At the top of the next page you will find the key to the rating method you must use. You will have to rate each word along a scale going from 1 to 5, where 1 represents a very typical instance of a category, and 5 represents a very atypical instance. The numbers in between should be used to represent gradations in typicality of the words being rated. In some cases, you may feel that the item being rated simply does not belong to the category you are considering. You should then ring the number 6 on the scale. Borderline cases, such as items which sometimes belong to the category, but not always, should be given a 5. If you do not know the word, leave the rating line blank.

"Proceed as follows: Make sure you know how to use the scale, using the key at the top of the page. Read the category name given below that. You can then start rating each word given under the category name, by ringing the one number, from 1 to 5, which you think best expresses its goodness-of-example. If you do not think that the word belongs to the category used, ring 6."

Subjects completed the booklets in their own time, and were asked to use their own judgement without consulting other people. Because not all booklets were returned, not all categories have the same number of ratings. Of about 130 booklets distributed, 93 (72%) were returned. Of these, 66 were in matched pairs, and 27 were unmatched. Since all booklets used a different random order of 6 randomly selected categories, the varying number of subjects in different categories should have no systematic effect. The number of subjects contributing to each category was as follows: food flavourings, fruit and insects (43); birds and weapons (44); clothing and sports (45),vehicles (46) vegetables (48), fish and flowers (50) and furniture (51).

### 3.2.3b Familiarity norms

Subjects Sixty-three students at The City University, London, served as unpaid volunteer subjects. About one third of them were psychology students. Only five of the subjects who did familiarity ratings had also done typicality ratings (the interval between the two tests was about 8 months).

Design and Procedure The testing booklets used for the typicality ratings were modified as follows: The scale at the top of each page now ranged from 1 (Very Familiar) to 5 (Very Unfamiliar), with 6 now meaning that the word was unknown. For convenience, the order of the words within category lists was the same as for the typicality ratings. It should therefore be stated that familiarity as measured here refers to familiarity within the context of other category items. Since most experimenters normally use words in the context of other category members, this was deemed an appropriate procedure. The following instructions (adapted from McCloskey, 1980) were given as the top page in the response booklet:

> "On the following pages you will find collections of words grouped together by category. There are 6 categories in all. Your task will be to rate the words according to how familiar you are with their meaning. Please make your ratings on a scale from 1 to 5, with 1 meaning the word is highly familiar to you, and 5 meaning that it is highly unfamiliar. There is also an additional number, 6, which you may use to indicate that you do not know the particular word. You will probably encounter few or no words whose meaning you do not know. A highly familiar word is one whose

meaning is immediately obvious to you, while a less familiar word is one that you may have to think about for a moment.

"The words are grouped together by category for our convenience only -- it was easier for us to present the words by category rather than, say, alphabetically. So you should not let yourself be influenced by whether you think that a particular word is a good or a bad example of the category mentioned. Rate the words simply on how familiar they are to you, as words. Finally, try to spread your ratings out over the whole of the scale -- in other words, don't use all 1's or 2's; or 4's and 5's. At the top of the first page, and at the top of any subsequent new category headings, you will find the key to the rating scale, so please always check that you know how to use it."

To reduce testing time, subjects again completed only six of the categories in their own time. The randomization procedure was the same as described for the Typicality norms. In the final sample, there were between 30 and 32 subjects rating each category. Sixty-six booklets were distributed, and 63 were returned, of which all but 3 were matched pairs.

3.2.3c Associative frequency norms

Subjects The subjects were 72 first- and second-year students at The City University, London. They were tested either individually or in large groups, as part of a psychology laboratory class. They were unpaid. Roughly two-thirds were psychology students.

Design and Procedure The subjects were each given a booklet containing 12 pages, each page headed by a category name and otherwise blank. Order of presentation of the pages was randomized for each subject. A cover sheet contained the following instructions (similar to those used by Battig and Montague, 1969, with the provision of 60, rather than 30 seconds per category):

> "We are running this experiment to try to find out which items or objects people commonly give as belonging to various categories. The procedure is very simple. This booklet contains 12 pages. Each page has, written on the top left-hand corner, the name of a category. When I signal "Begin", turn over this page and read the name of the category on the first page. I will then say "Start". You will then have one minute to write down as many words as you can think of which in your opinion belong to that category. When the minute is up, I will say "Stop". This means that you must stop writing and turn over the page and read the name of the next category. You will again be given one minute to write down as many members of that category as you can think of. We will do the same thing for all 12 categories represented in the booklet. Just two final points: Always wait for me to say "Sart" and "Stop" before writing anything

down and turning the page, respectively. Secondly, please write clearly and write each word or phrase in full."

The subjects were given time to read the instructions, and any queries were answered. The procedure was then followed as described in the instructions.

### 3.2.4 Results and discussion

Tables of mean rated typicality, mean rated familiarity and associative frequency for the 531 words in the corpus are presented in Appendix Four. The tables also include three further counts: number of category rejections (the rating of "6" on the typicality scale), number of "unknown" category members (the sum of the number of blank lines in typicality protocols, and the number of ratings of "6" on familiarity protocols), and the number of times a word was generated first in the associative frequency task (given in parentheses after the associative frequency score).

Mean typicality and familiarity scores were calculated including the ratings of 6 as the extreme end of both scales (*). Typicality ratings were calculated only for those subjects who knew the word. Associative frequency scores were derived using a strict criterion, i.e., a perfect match between a given word and the word in the corpus. No synonyms or subvarieties of list words were included, though obvious misspellings were corrected. Words

---

(*) This was done so as to extend the range of the scale and the size of the subject sample. Treating the ratings of 6 as "missing values" and re-calculating the typicality and familiarity ratings yielded scale values which correlated at .94 or better with the values reported in Appendix Four. Thus, employing a separate response for non-members and unknown items served mainly to increase the face validity of the ratings task for the subjects, as well as providing information on how many subjects consider an item not to belong in a category. This information may itself be useful when selecting materials for experiments using semantic dimensions.

generated by more than three subjects in the associative frequency task, but absent from the corpus are listed after each category, in Appendix Four, together with their associative frequency scores and the number of times they were given as the first item in the generated list.

Table 3.2 shows the means and standard deviations across category items for mean typicality and familiarity ratings, and the total number of items generated for each category in the associative frequency task. As can be seen, the two ratings measures had very similar standard deviations, making it possible to carry out a correlational analysis of the interdependence of the three measures with some degree of confidence.

Figures 3.1, 3.2 and 3.3 illustrate the frequency distributions for all three measures. These were examined collapsed across category. All three were positively skewed (typicality, 0.79; familiarity, 1.85; associative frequency, 1.58), with familiarity having the highest kurtosis (3.66), followed by associative frequency (1.78) and typicality (0.04). Thus familiarity scores appear to have clustered at the high familiarity end of the scale (68% of all words being given ratings between 1 and 2, compared with 41% for typicality).

Table 3.2 - Experiment 4: Means and standard deviations for typicality (TYP) and familiarity (FAM) ratings, and number of items (I) generated in the associative frequency task.

| Category | Means | | Standard Deviations | | I |
|---|---|---|---|---|---|
| | TYP | FAM | TYP | FAM | |
| Birds | 2.07 | 1.78 | 0.62 | 0.66 | 129 |
| Clothing | 2.39 | 1.69 | 0.86 | 0.63 | 114 |
| Fish | 2.18 | 2.34 | 0.75 | 0.98 | 112 |
| Flowers | 1.92 | 2.33 | 0.60 | 1.13 | 107 |
| F. flavourings | 2.39 | 2.42 | 0.68 | 1.35 | 116 |
| Fruit | 2.30 | 1.93 | 1.00 | 0.73 | 69 |
| Furniture | 2.90 | 1.73 | 1.50 | 0.52 | 86 |
| Insects | 2.10 | 2.18 | 0.84 | 1.29 | 64 |
| Sports | 2.42 | 1.58 | 0.97 | 0.39 | 108 |
| Vegetables | 2.21 | 1.89 | 0.85 | 1.03 | 74 |
| Vehicles | 3.20 | 1.65 | 1.09 | 0.45 | 107 |
| Weapons | 2.22 | 1.64 | 0.98 | 0.88 | 146 |
| Means: | 2.36 | 1.93 | 0.98 | 0.84 | 103 |

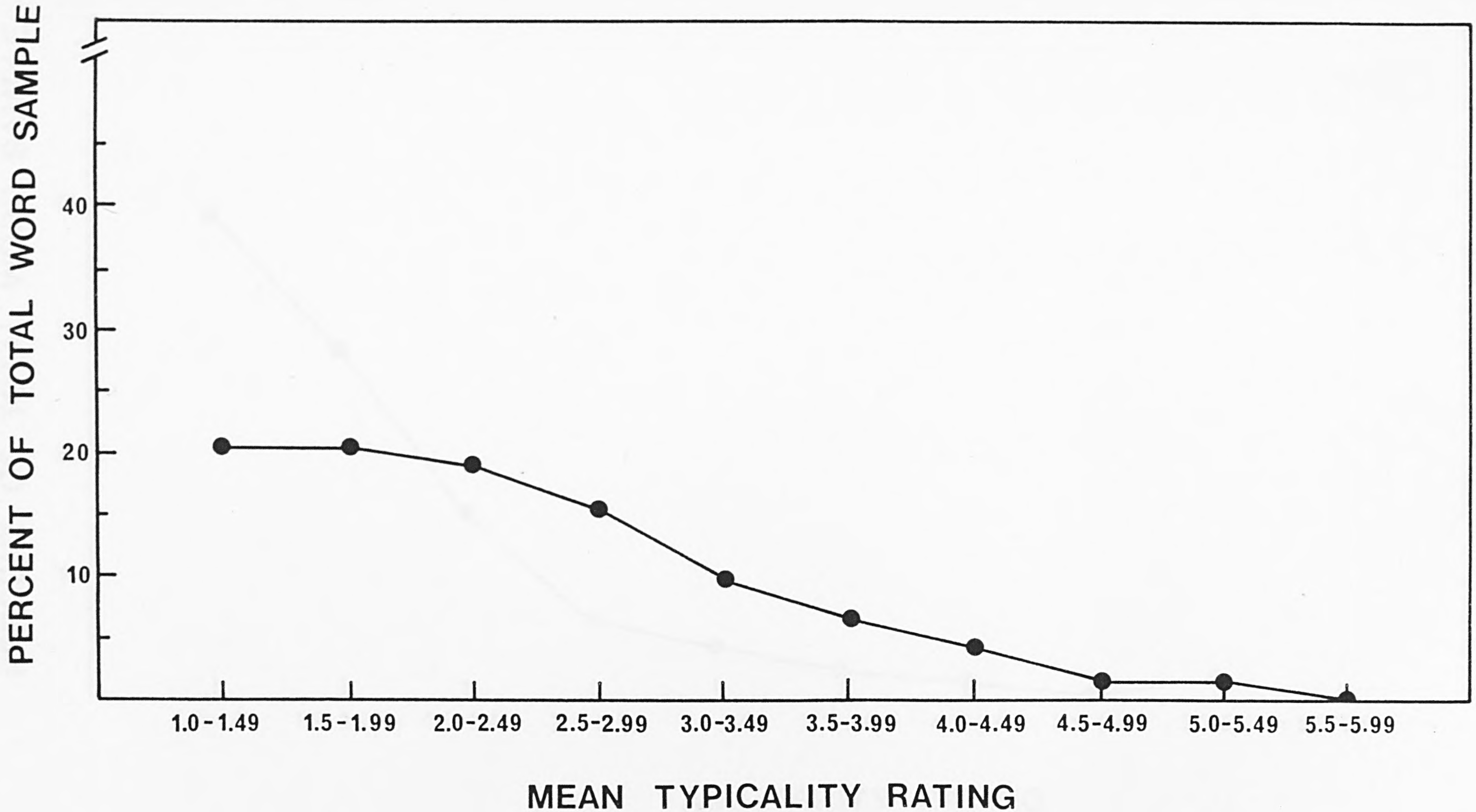



Figure 3.1 - Experiment 4: Frequency distribution for the typicality measure.

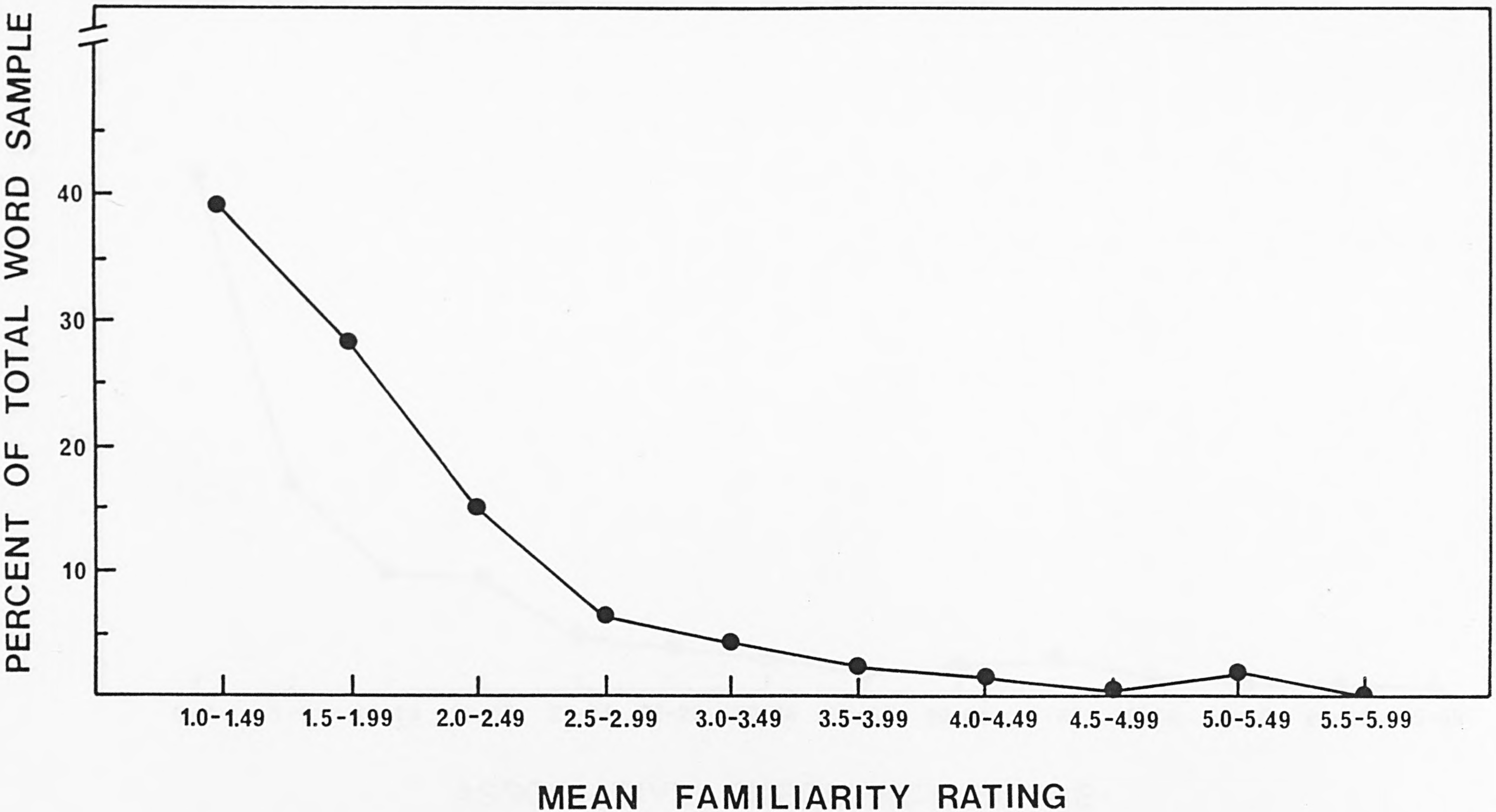



Figure 3.2 - Experiment 4: Frequency distribution for the familiarity measure.

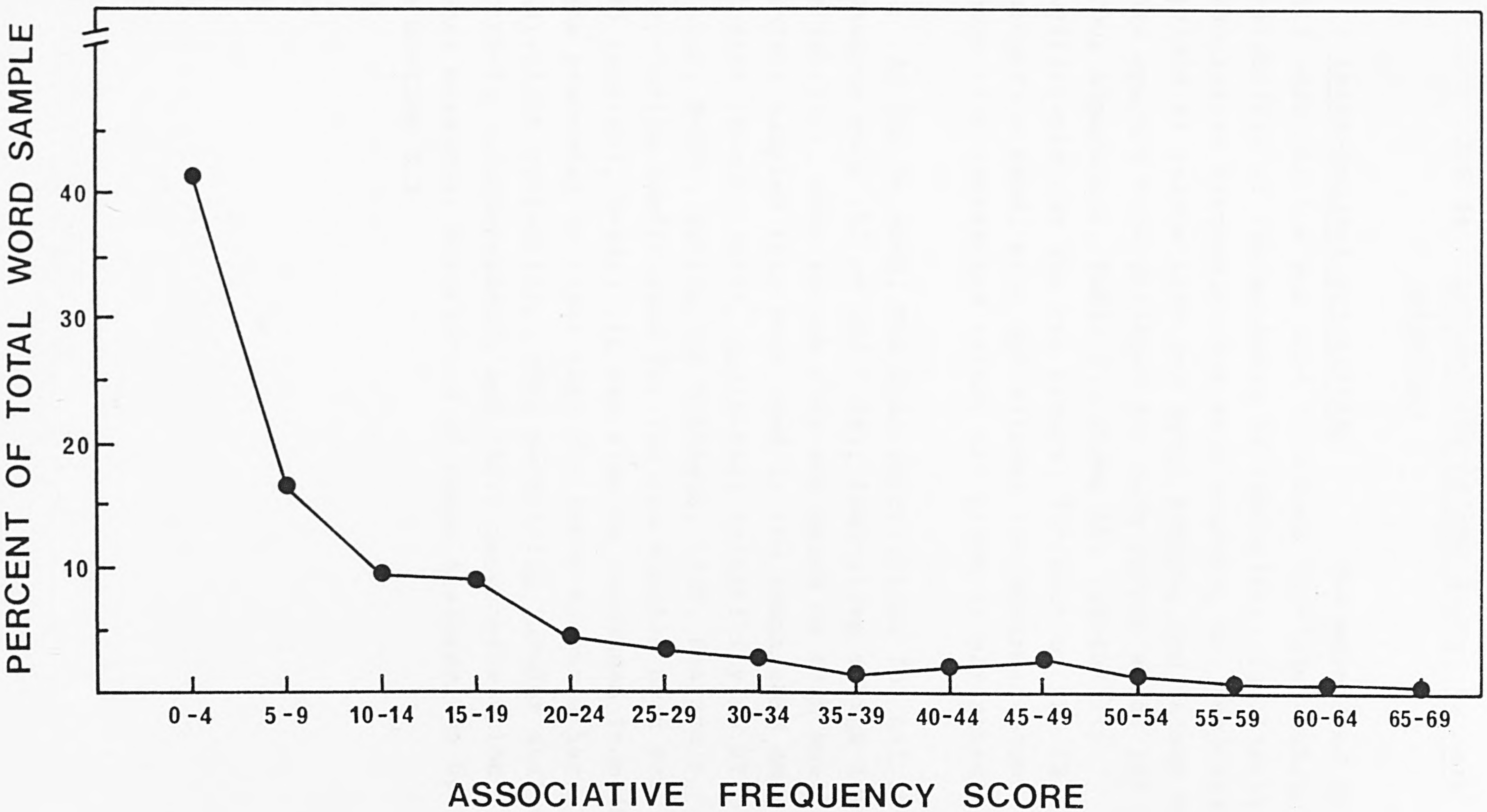



Figure 3.3 - Experiment 4: Frequency distribution for the associative frequency measure.

### 3.2.4a Reliability of the normative data collected

Inter-subject reliability The method of split-half correlations was used to assess the inter-subject reliability of the measures of typicality, familiarity and associative frequency. For each measure, the subjects were divided at random into two equal groups, and values of the same measure were obtained for each corpus word, for each group separately. Table 3.3 shows the reliability coefficients for the two groups, for each of the 12 categories used, with and without the Spearman-Brown correction (corrected values are given in parentheses).

As can be seen, the mean correlations for all three measures were .92 or .93 (.96), indicating a high level of reliability, even though they are based on much smaller subject samples than were used in the comparable American studies (Rosch, 1975a, split-half reliability r=.97 or higher, N=209; Battig and Montague, 1969, Pearson's correlation coefficient for the two samples used was .96 (median), N=442). It can also be concluded from the data presented overleaf that the three measures had equivalent reliability, thus permitting further analysis of their interdependence and their intercorrelation with other measures. Description of these analyses can be found in Section 3.3.

Table 3.3 - Experiment 4: Reliability coefficients for typicality (TYP) and familiarity (FAM) ratings and for associative frequency (AF)

| Category | Reliability r | | |
|---|---|---|---|
| | TYP | FAM | AF |
| Birds | .91 (.95) | .92 (.96) | .91 (.95) |
| Clothing | .95 (.97) | .92 (.96) | .94 (.97) |
| Fish | .92 (.96) | .95 (.97) | .94 (.97) |
| Flowers | .90 (.95) | .96 (.98) | .93 (.96) |
| F. flavourings | .83 (.91) | .96 (.98) | .86 (.93) |
| Fruit | .95 (.97) | .90 (.95) | .93 (.96) |
| Furniture | .99 (.99) | .90 (.95) | .94 (.97) |
| Insects | .88 (.94) | .97 (.99) | .96 (.98) |
| Sports | .95 (.97) | .82 (.90) | .94 (.97) |
| Vegetables | .90 (.95) | .95 (.97) | .96 (.98) |
| Vehicles | .92 (.96) | .90 (.95) | .97 (.99) |
| Weapons | .94 (.97) | .90 (.95) | .91 (.95) |
| Means: | .92 (.96) | .92 (.96) | .93 (.96) |

(Numbers in parentheses represent the reliability coefficient values after application of the Spearman-Brown correction.)

Typicality and familiarity reliability Examination of the reliability coefficients for typicality and familiarity in Table 3.3 shows a connection between the reliability of these two measures across categories. Specifically, it is apparent that the more reliable typicality is, the less reliable familiarity becomes ($r=-.66$, $p<.05$). Although the range of variation in reliability is small (.82 to .99), this significant negative correlation may be taken as further evidence of the independence of the two measures. It seems to indicate that the judgement of typicality for very unfamiliar items is erratic, depending largely on the subjects' depth of knowledge about the item (cf. Glass & Meany, 1978). Thus, those categories with a higher proportion of unfamiliar items will have lower reliability coefficients for typicality judgements, while the greater number of clearly unfamiliar items in those categories will increase the reliability of the familiarity judgements. In support of this interpretation, the standard deviation of familiarity ratings was also found to correlate with the typicality reliability coefficients ($r=-.80$, $p<.002$), and the partial correlation between the two reliability measures, holding familiarity standard deviation constant is near zero ($r=-.08$). No such relationship is apparent in any comparison involving associative frequency ($r=.42$ and $r=-.13$ with typicality and familiarity, respectively). The full correlation matrix illustrating the above results can be found in Table 3.4, while Table 3.5 lists the results of the partial correlational analysis.

Table 3.4 - Experiment 4: Correlation matrix for the typicality and familiarity reliability relation.

| | Typicality standard deviation | Familiarity reliability | Familiarity standard deviation |
|---|---|---|---|
| Typicality reliability | .72* | -.66* | -.81* |
| Typicality standard deviation | - | -.50 | -.58* |
| Familiarity reliability | - | - | .84* |

(* - significant at 5%.)

Table 3.5 - Experiment 4: Summary of the partial correlational analysis on the typicality and familiarity reliability relation

1. Holding typicality reliability constant

| | Familiarity reliability | Familiarity standard deviation |
|---|---|---|
| Typicality standard deviation | -.054 | -.002 |
| Familiarity reliability | - | .702* |

2. Holding typicality standard deviation constant

| | Familiarity reliability | Familiarity standard deviation |
|---|---|---|
| Typicality reliability | -.497 | -.694* |
| Familiarity reliability | - | .784* |

3. Holding familiarity reliability constant

| | Typicality standard deviation | Familiarity standard deviation |
|---|---|---|
| Typicality reliability | .593 | -.630* |
| Typicality standard deviation | - | -.342 |

Table 3.5 (Contd.) - Experiment 4: Partial correlational analysis.

4. Holding familiarity standard deviation constant

| | Typicality standard deviation | Familiarity reliability |
|---|---|---|
| Typicality reliability | .513 | .079 |
| Typicality standard deviation | - | -.023 |

(* - significant at 5%, two-tailed.)

### 3.2.4b Comparison with American norms

As stated in the discussion section of Experiment 3, one of the reasons for collecting these norms was the possibility that American norms might not be directly translatable to situations involving British samples of subjects. Given that the reliabilities of the measures of typicality and associative frequency are the same for the corpus data, a comparison can be made with the previously published American norms of typicality (Rosch, 1975a) and associative frequency (Battig and Montague, 1969). Table 3.6 illustrates the relationship between the two sets of norms (British and American). Note that, for associative frequency, the correlations between Battig and Montague scores and the corpus scores were calculated for the whole of the corpus sample, but for the Rosch comparison only those words common to both sets of norms could be used.

Comparison of Tables 3.3 and 3.6 seems to support the argument for the collection of new sets of norms specifically for a British population -- the mean correlations for a comparison with American norms (Table 3.6) are less than the split-half reliability measures given in Table 3.3.

Interestingly, the correlations for typicality (mean=.85) given in Table 3.6 are higher than those for associative frequency (mean=.76). This difference is significant across the 8 categories for which the typicality comparison could be made ($t(14)=1.84$, $p < .05$). It is a result which is consistent with Rosch's (1978) theory of typicality effects (see Section 1.4, p. 30 ). This assumes that typicality depends more on the family resemblance among items, whereas associative frequency

Table 3.6 - Experiment 4: Correlations between rated typicality and Rosch's (1975a) typicality norms; and between associative frequency and Battig and Montague's (1969) norms.

| Category | Typicality | N | Assoc. Freq. | N |
|---|---|---|---|---|
| Birds | .82 | 38 | .76 | 52 |
| Clothing | .90 | 32 | .76 | 55 |
| Fish | - | - | .48 | 37 |
| Flowers | - | - | .77 | 46 |
| F. flavourings | - | - | .85 | 40 |
| Fruit | .86 | 34 | .90 | 43 |
| Furniture | .92 | 21 | .81 | 41 |
| Insects | - | - | .85 | 34 |
| Sports | .80 | 37 | .53 | 48 |
| Vegetables | .72 | 31 | .75 | 41 |
| Vehicles | .90 | 29 | .76 | 54 |
| Weapons | .92 | 26 | .91 | 40 |
| Means | .85 | | .76 | |

may be expected to reflect local differences in language use and familiarity with specific items. Consequently, typicality should be less sensitive to transatlantic variation. Though other explanations are possible, particularly as the American norms also differ in time and place of collection, it would appear that associative frequency may be more sensitive to cultural differences than is rated typicality.

## 3.3 Theoretical considerations

This section deals with the intercorrelation among the three measures for which normative data were obtained. Before any analysis could be carried out, however, it was necessary to transform the associative frequency scores to take account of the marked non-linearity of this measure with typicality and familiarity (see Figures 3.1-3.3). This non-linearity seemed to be largely due to the positive skewness of all three measures, coupled with the negative correlation between associative frequency and the other two measures (as they were scaled). Since correlations measure the linear relationship between two variables, associative frequency scores were transformed to reduce the curvilinearity by defining them as $Log_{10}(AF+1)$, where AF is the associative frequency. The transformed scores had skewness -.05 and kurtosis -1.13, thus slightly reversing the skew of the distribution (see Figure 3.4). All references to associative frequency in the following sections refer to the transformed variable.

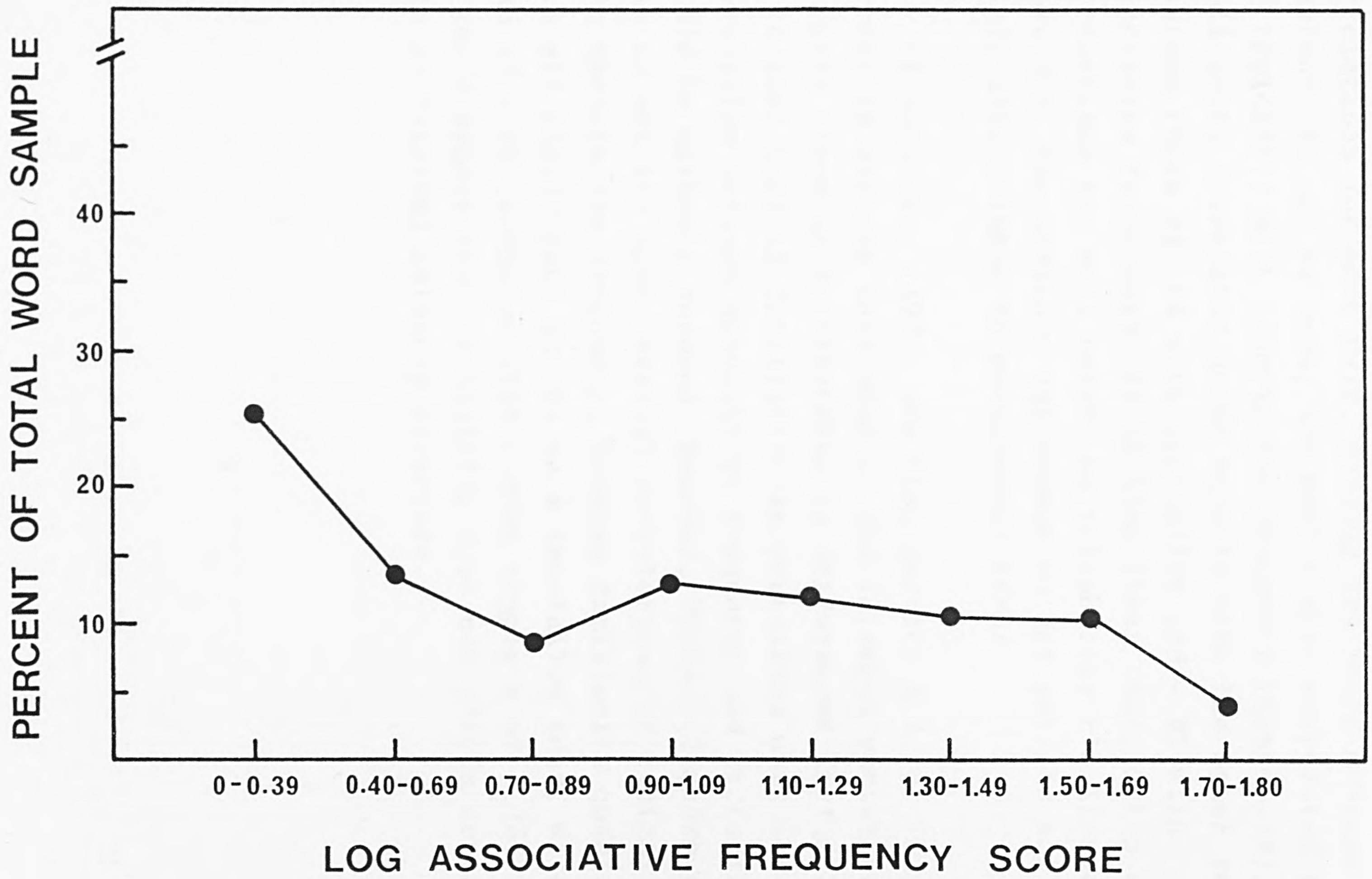



Figure 3.4 - Experiment 4: Frequency distribution for the log associative frequency transformation.

### 3.3.1 Familiarity and internal category structure

Table 3.7 (overleaf) gives the product-moment correlations for each pair of measures, and the partial correlations for each pair, holding the third measure constant. As can be seen, the most highly correlated pair was typicality with associative frequency (mean=-.76), and familiarity correlated about equally with the other two measures (mean of .54 with typicality and -.61 with associative frequency). It is also clear that, since all correlations are well below the reliability r's listed in Table 3.3, the variance not common to any pair is not solely attributable to measurement error.

If McCloskey (1980; see also section 1.4, p. 31) is correct in arguing that most of the internal variation of category items is attributable to differences in familiarity, it follows that if familiarity is partialled out, the correlation between associative frequency and typicality should be markedly reduced. However, Table 3.7 shows that this is not the case. Partial correlations of typicality with associative frequency, holding familiarity constant, were all significant at .01 on a two-tailed test, with a mean of -.63 (compared with a mean simple r of -.76). Thus, it would appear that familiarity does not play a central role in internal category structure.

Table 3.7 - Experiment 4: Simple and first-order partial correlations between typicality (TYP), familiarity (FAM) and associative frequency (AF).

| Category | N | Simple correlations (all sig.<.01, except where marked †, sig.<.05) | | | First-order partial correlations (*sig. <.01) | | |
|---|---|---|---|---|---|---|---|
| | | TYP-FAM | TYP-AF | FAM-AF | TYP-FAM.(AF) | TYP-AF.(FAM) | FAM-AF.(TYP) |
| Birds | 52 | .58 | -.70 | -.75 | .12 | -.49* | -.59* |
| Clothing | 55 | .60 | -.71 | -.63 | .28 | -.54* | -.36* |
| Fish | 37 | .76 | -.79 | -.85 | .27 | -.42* | -.62* |
| Flowers | 46 | .82 | -.90 | -.79 | .41* | -.71* | -.21 |
| F. flavourings | 40 | .51 | -.77 | -.44 | .30 | -.71* | -.09 |
| Fruit | 43 | .52 | -.87 | -.52 | .16 | -.82* | -.16 |
| Furniture | 41 | .40 | -.81 | -.53 | -.06 | -.77* | -.39* |
| Insects | 34 | .51 | -.70 | -.82 | -.16 | -.57* | -.75* |
| Sports | 48 | .27† | -.75 | -.35 | .01 | -.72* | -.23 |
| Vegetables | 41 | .67 | -.86 | -.66 | .27 | -.75* | -.22 |
| Vehicles | 54 | .52 | -.71 | -.55 | .22 | -.59* | -.30* |
| Weapons | 40 | .37 | -.57 | -.47 | .14 | -.48* | -.34* |
| Means : | | .54 | -.76 | -.61 | .16 | -.63* | -.35* |
| "Creatures" : | | .62 | -.73 | -.81 | .08 | -.49* | -.65* |
| Others : | | .52 | -.77 | -.55 | .19 | -.68* | -.26 |

### 3.3.2 Determinants of associative frequency

The partial correlations given in Table 3.7 can be interpreted in another fashion. If the generation task associated with the associative frequency measure is viewed as a dependent variable, the partial correlations indicate the relative contribution of each of the other two measures to the probability of an item being generated.

Typicality seems to be a good predictor of associative frequency (mean partial r=-.63) compared with familiarity (mean partial r=-.35). Table 3.8 gives a more detailed description of this pattern of results. Again, familiarity turns out not to be of crucial importance for the carrying out of a semantic task, though it does have some effect (in categories birds, fish and insects, in particular).

Table 3.8 - Experiment 4: Multiple r's for the relationship between typicality (TYP), familiarity (FAM) and associative frequency (AF), and associated β-weights.

| Category | AF from TYP and FAM | | |
|---|---|---|---|
| | β-weights | | Multiple r's |
| | TYP | FAM | |
| Birds | -.392 | -.519 | .813 |
| Clothing | -.509 | -.330 | .754 |
| Fish | -.333 | -.596 | .878 |
| Flowers | -.903 | - | .903 |
| F. flavourings | -.769 | - | .769 |
| Fruit | -.866 | - | .866 |
| Furniture | -.717 | -.241 | .844 |
| Insects | -.374 | -.635 | .885 |
| Sports | -.750 | - | .750 |
| Vegetables | -.860 | - | .860 |
| Vehicles | -.582 | -.247 | .741 |
| Weapons | -.458 | -.303 | .635 |

### 3.3.3 Basic- and superordinate-level categories

The pattern of intercorrelations described in the previous section does not hold equally well for all the categories involved in the norms. Three of the categories (fish, birds and insects) seem to depart from the general pattern of results in the degree to which familiarity can predict associative frequency holding typicality constant. For these three categories, both typicality and familiarity jointly determine the probability of an item being generated to the category name, whereas for the remaining categories only typicality seems to be involved. What makes this difference intriguing is the fact that the three "anomalous" categories are all types of "creature" and are all at what Rosch termed a "basic" level of abstraction (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976). According to Rosch et al., basic-level categories (as opposed to superordinate-level categories) have a high degree of inter-item similarity, can be visualised as a generic image (it is easier to form an image of what fish look like than of what furniture looks like), and are the first level of a semantic hierarchy children learn. All the remaining categories in the corpus, with the exception of flowers, are at the superordinate concept level.

One possible interpretation of these partial correlation results is that, because of their higher inter-item similarity, basic-level categories have less marked typicality differences (cf. Rosch & Mervis, 1975: typicality is seen to reflect the extent of featural overlap, or family resemblance, among items. Typical words show a greater degree of featural overlap than atypical instances of the same category). It is thus possible that familiarity with the different members of basic-level categories may be more crucial than typicality in determining the probability of their being generated. By comparison, superordinate-level categories are more heterogeneous, so that featural overlap (and hence typicality) is more crucial in a generation task.

If the above interpretation is correct, it then remains to be explained why _flowers_ (also a basic-level category) did not fall in with the three other basic-level categories in the norms. The answer may lie in the strong collinearity of the typicality, familiarity and associative frequency measures for this category (r's of -.79, .82 and -.90), which renders partial correlations particularly unstable. It is the case that, on simple correlations, the four highest values for the correlation of familiarity with associative frequency occur for all four basic-level categories (including _flowers_), even though these are _not_ the categories with the highest familiarity variance. Overall, there is some difference in the level of familiarity variance between basic- and superordinate-level categories (mean basic-level S.D.=1.02; mean superordinate-level S.D.= .75). But of the four categories with the highest variance (_food flavourings_, 1.35; _insects_, 1.29; _flowers_, 1.13; and _vegetables_, 1.03) two are at the superordinate level, and the remaining two basic-level categories (_birds_, .66; _fish_, .98) have familiarity standard deviations at least on a par with those for superordinate-level categories. (See Table 3.2, p.96 ).

### 3.3.4 Familiarity, number of properties and typicality

The final analysis investigated the relationship between the measure of rated typicality used in these norms, and an alternative way of conceptualizing item familiarity, through the number of properties, or features, subjects are able to report they associate with the given items.

Ashcraft (1978a and b) reported experiments which showed that the number of properties a person can generate for a word was a good predictor of that word's rated typicality. In a follow-up study, Malt and Smith (1982) argued that thus number-of-properties measure could be a better way of estimating item familiarity than the rating procedure used by McCloskey (1980) and in these norms, since the number-of-properties measure estimates familiarity with the referred-to object, whereas the rating procedure measures the subject's familiarity with the object's name. Malt and Smith (1982) provide values of the number-of-properties variable for seven of the 12 categories in the corpus, based on a sample of 240 subjects. They also provide mean typicality ratings. It is thus possible to compare their data with those presented here, to test the value of this fourth measure.

For analysis purposes, one of Malt and Smith's categories (<u>furniture</u>) was dropped, because the degree of sample overlap was very small and the range of number of properties within the overlap that existed was too small to permit reliable analysis. In the remaining six categories, the overlap sample ranged from 9 to 13 words, and the range of number of properties was at least 2.5. Correlations were calculated for each category between all pairs of the following five variables:

i) Malt and Smith's typicality ratings
ii) Malt and Smith's number-of-properties measure
iii) rated typicality as reported here
iv) rated familiarity
v) associative frequency

Table 3.9 (overleaf) shows the results of this analysis. The two sets of typicality measurements correlated at .85 (cf. the similar result in Table 3.6), and each correlated with familiarity and associative frequency in an equivalent way (Malt and Smith's typicality ratings with familiarity, mean -.49; typicality with familiarity, mean .56; Malt and Smith's typicality ratings with associative frequency, mean .67; typicality with associative frequency, mean -.68). The correlations of Malt and Smith's typicality ratings with the number-of-properties measure are also broadly the same on these reduced sample sizes as those reported for the complete sample used by Malt and Smith (1982: original mean .49, present mean .43). The reduction in sample size does not therefore appear to have introduced any new sampling bias.

The number-of-properties measure correlated highest with familiarity (-.65) and associative frequency (.65) and least with typicality (-.35). Flowers was the only category for which the number-of-properties generated correlated significantly with typicality. Partial correlational analysis confirmed the lack of involvement of this measure in typicality. The mean correlation of number-of-properties with typicality, holding familiarity

Table 3.9 - Experiment 4: Correlations between rated typicality (TR) and number of properties (NP) taken from Malt and Smith (1982), and typicality (TYP), familiarity (FAM) and associative frequency (AF) from the present study. (Note: TR and TYP were scored in opposite directions.) (*=p $<$ .05, two-tailed)

| Category | N | TR correlated with: | | | | NP correlated with: | | | TYP with: | | FAM with: |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NP | TYP | FAM | AF | TYP | FAM | AF | FAM | AF | AF |
| Birds | 12 | .63* | -.71* | -.67* | .86* | -.14 | -.67* | .66* | .42 | -.70* | -.72* |
| Clothing | 10 | .12 | -.82* | -.09 | .15 | -.42 | -.38 | .59 | .35 | -.53 | -.93* |
| Fish | 9 | .07 | -.89* | -.37 | .38 | .07 | -.87* | .84* | .27 | -.25 | -.94* |
| Flowers | 12 | .66* | -.86* | -.64* | .88* | -.82* | -.89* | .81* | .84* | -.94* | -.84* |
| Fruit | 13 | .49 | -.94* | -.59* | .87* | -.47 | -.42 | .40 | .74* | -.93* | -.65* |
| Vehicles | 9 | .60 | -.89* | -.90* | .88* | -.33 | -.66 | .57 | .72* | -.71* | -.99* |
| Means: | | .43 | -.85 | -.49 | .67 | -.35 | -.65 | .65 | .56 | -.68 | -.85 |

constant, was .04, whereas the mean correlation of number-of-properties with familiarity, holding typicality constant was -.55 (significant in 3 of the 6 categories). Thus, number-of-properties does not appear to be a confounding variable in typicality ratings, and its apparent correlation with typicality can be entirely attributed to its correlation with rated familiarity. The data presented indicate that a direct rating of familiarity, such as that used in the norms, may be a better measure of the effects of familiarity on typicality than the number of properties people can generate.

## 3.4 General discussion

The results of Experiment 3 (Chapter Two) highlighted the need to obtain new sets of normative data for a British sample of subjects, before carrying out further research on possible factors affecting cue loading. Because all the available published data were collected in the USA, at different points in time and location, and involved different sets of materials, it was possible that lists selected using such norms might not accurately reflect the differences implied by the ratings. Experiment 4 was meant to provide normative data for three different measures of category internal structure (typicality, familiarity and associative frequency), for the same set of materials, using a British sample of subjects. Correlational analysis showed that collection of new normative data was indeed warranted, since the new norms correlated with the equivalent American norms at a level below that of their internal reliability. It was also found that typicality was slightly less affected by cultural differences than associative frequency was, a finding which is consistent with Rosch's (1978) interpretation of typicality effects.

The results of Experiment 4 also allowed further analysis to be carried out concerning the interdependence of the three dimensions used. It was hoped that such an analysis might provide more information about the nature of typicality in general. The patterns of intercorrelations described in Section 3.3 indicate that the three dimensions used, though, intercorrelated, reflect different sources of variance and do not derive from one single underlying factor. In addition, a number of interesting patterns emerged from the analysis of results reported. These can be summarized as follows: (i) familiarity was shown not to be the critical variable in internal category structure, since rated typicality was still well correlated with associative frequency when rated familiarity was partialled out; (ii) rated typicality was found to be a better predictor of associative frequency than familiarity was, but (iii) this did not hold for basic-level categories, where both variables independently predicted associative frequency. (iv) When rated familiarity was held constant, the number of properties people can generate for an item was not related to typicality or associative frequency, so that the rating procedure used here is to be preferred to the property-generation task as a measure of a word's familiarity.

It is important to note, however, that rated familiarity as described above refers to familiarity with an item in the context of other category members, since this was the way such ratings were meant to be used for the present research. The relation of this particular way of measuring familiarity to familiarity within a broader context will require further empirical clarification.

## 3.5 Re-analysis of Experiment 3's materials using the new normative data

Using the normative data collected in Experiment 4, it is possible to re-examine the materials used in Experiment 3 and to check whether, in fact, they reflect the differences intended when they were selected and whether any, or both, of the other two dimensions for which normative data were collected also differed in a systematic way across typicality level.

The mean typicality value for typical list words, using Rosch's norms, was 1.65 (range 1.02-3.30). The equivalent figure for atypical instances was 3.49 (range 1.90-6.21). Using the new normative data, typical-word mean typicality was 1.44 (range 1.00-3.02), and atypical-word mean typicality was 2.87 (range 1.51-4.67). Ratings for both types of word were still significantly sifferent on an independent t-test: $t(130)=11.00$, $p < .0001$, one-tailed. Thus, using the new normative data, the degree of overlap between typical and atypical-word ratings was slightly higher but both sets of ratings still differed as designed in overall typicality.

More importantly, when the new norms were used to estimate the degree to which the other two variables differed across typicality, it was found that typical words were not only more familiar than atypical words (mean typical familiarity 1.27; atypical 1.72, range 1.00-3.00 and 1.10-4.47, respectively; $t(130)=3.78$, $p < .0005$, one-tailed), but they had also much higher associative frequency scores (mean typical 35.22, range 1-69; mean atypical 5.98, range 0-32). Thus the alternative explanations of the typical-word recall advantage given in the Discussion section of Experiment 3 cannot be ruled out. The typicality effect could have come about because atypical words were not easily retrieved from the category cue to begin with, and were also unfamiliar to the subjects.

# CHAPTER FOUR : TYPICALITY AND THE CUE-OVERLOAD PRINCIPLE

## 4.1 Introduction

Experiment 3 tested tested the Cue-Overload Principle's ability to predict or explain recall performance in a situation where inter-item and item-cue strength of association was manipulated. A design problem, and uncertainty as to the appropriateness of the materials used in this experiment, meant that its results could not be used confidently to test Cue-Overload theory. Chapter Three described the collection of new normative data, to overcome the problems encountered when selecting materials for Experiment 3. The three experiments now described in Chapter Four use these normative data, and a new design format, to pursue the same line of enquiry adopted in Experiment 3.

## 4.2 Experiment 5 - Typicality and within-category release from retroactive inhibition

### 4.2.1 Introduction

This experiment investigated the effects of typicality on cue overload, using the Method of Interpolated Attributes. It involved shifting copy-list typicality within the same semantic category, as in conditions SAME-TYPICAL and ATYPICAL of Experiment 3, but using homogeneously typical or atypical memory and copy lists, and materials selected so as to differ in typicality but not in rated familiarity, as per the norms collected in Experiment 4. Thus the materials took account of McCloskey's (1980) criticism concerning the nature of typicality effects (see Section 1.4, p.31 ). In order to make sure that both typical and atypical lists were encoded with respect to the same general category (without using mixed-typicality lists), a between-subjects "naming" factor was introduced into the basic design. Half the subjects were given the category name for each pair of lists, and had the categorical nature of the lists emphasized in the instructions read prior to the experiment, while the other half were given standard instructions -- including that lists were organised by category -- and were not given any category names. Each group's recall data were then looked at separately, so see if any systematic differences had occurred in the patterns of results obtained.

The reasoning behind the experiment was that, if typicality is a functional retrieval cue, then Cue-Overload would predict release from retroactive inhibition with shifts in the rated typicality of the copy-list words. In addition, because equivalent numbers of typical and atypical words were used in each condition, the Principle would also predict the same level of release for both types of word.

It is interesting to note that this latter prediction is at variance with the results Keller and Kellas (1978) obtained, in a similar experiment within the release from proactive inhibition paradigm (Brown-Peterson procedure). Keller and Kellas (1978) showed that it is possible to obtain release from proactive inhibition with shifts in typicality within the same semantic category. Their study also argued against an interpretation of typicality effects in terms of differential encoding of atypical words with respect to the category used (Loftus, 1975; Collins & Loftus, 1975) because symmetrical proactive inhibition build-up was obtained. The differential encoding hypothesis would have predicted less pronounced decline in recall for atypical words (see, for instance, Loess, 1967; Brown & Atkinson, 1974). However, Keller and Kellas found that, though build-up was symmetrical, proactive inhibition release was not. A shift from atypical to typical words produced only about half the release obtained with the converse shift. They explained this finding in terms of the overall similarity between the category representation elicited by the pre-shift items (the prime) and the category members used for the shift triad. Similarity was functionally defined in terms of the overlap of defining and characteristic features of the words used (see, for instance, Smith, Shoben & Rips, 1974): defining features are common to all members of a category and are essential for category inclusion; characteristic features are exclusive to typical members. Minimal release was obtained with a shift from atypical to typical items because the prime was very similar to the typical (or shift) items.

Greater release was obtained with a typical to atypical shift because of the lack of common characteristic features between the prime and the shift items. (In other words, typical pre-shift triads lead to the abstraction of a prime which is very similar to a restricted category representation; atypical pre-shift triads lead to the setting up of a more general category representation.)

If the effects obtained within the release from proactive inhibition paradigm can be carried over to the release from retroactive inhibition paradigm, then one would expect release to be lower when the recall for typical words is being tested (SAME and DIFFERENT conditions with typical memory lists). This is because the category representation elicited by the typical words will also contain all the features common to the atypical words and the category label. The converse also applies: release should be higher for the conditions involving atypical memory lists (SAME and DIFFERENT conditions) where the shift words have characteristic features which are not common to the prime.

For both sets of predictions (Cue-Overload and Keller and Kellas) naming the category prior to list presentation would not be expected to greatly affect the pattern of results obtained, particularly since Keller and Kellas demonstrated that atypical members were encoded as members of the same category as typical members (they obtained symmetrical proactive-inhibition build-up). Naming the category prior to study thus helps to make sure that the same general category cue is being used, for typical and atypical words, and allows for interpretation of the results in terms of Cue-Overload.

4.2.2 Method

Design The Method of Interpolated Attributes was used, with the following modifications, designed to control presentation more tightly and to increase the effectiveness of the interference list:

(a) List presentation was via the VDU of a Commodore-PET microcomputer, and not manual, using decks of cards, as in the previous experiments.

(b) The copy list became part of an experimenter-paced task (as also used by Parkin, 1980), with each word being presented individually at a pre-determined rate.

(c) A verbal response was required from the subject on presentation of every single list word. This also applied to the copy list, so that the subjects had to both read aloud and write down the copy-list words.

The design was a 2 x 2 x 2 factorial (Naming x Typicality x Condition) with category-naming the only between-subjects factor. The eight experimental conditions yielded by this design are summarized in Table 4.1. All typicality shifts were within the same semantic category, and both materials and condition ordering were the same for the two naming groups.

For the within-subjects factors, the subjects were divided at random into two groups and a different set of four categories were allocated to each group. The balancing used ensured that half the subjects received typical word

Table 4.1 - Experiment 5: Summary of experimental conditions.

| Category naming factor | Memory list word typicality | Copy list word typicality | Condition type |
|---|---|---|---|
| Named | TYPICAL | TYPICAL | SAME |
| Named | TYPICAL | ATYPICAL | DIFFERENT |
| Named | ATYPICAL | ATYPICAL | SAME |
| Named | ATYPICAL | TYPICAL | DIFFERENT |
| Unnamed | TYPICAL | TYPICAL | SAME |
| Unnamed | TYPICAL | ATYPICAL | DIFFERENT |
| Unnamed | ATYPICAL | ATYPICAL | SAME |
| Unnamed | ATYPICAL | TYPICAL | DIFFERENT |

lists first, and half atypical lists, and within these groups, half the subjects were tested first under the SAME condition and half under the DIFFERENT condition. Category membership was fully rotated across condition, so that each category was used equally often under each condition. List membership (copy or memory) and word order within list were kept constant throughout. Each subject received one trial on each within-subjects condition, a different category per trial.

A distractor task followed each recall test. This consisted of an alphabetical list of 96 names of countries (from the Appendix to the International Driving Licence), which the subjects had to work through, saying aloud the name of each country and writing down next to each name whatever they could think of first about that country.

Materials Twenty-four words were selected from each of 8 categories represented in the corpus (Chapter

Three). The eight categories were divided into two groups of four, viz: Group I -- weapons, birds, clothing and fruit; Group II -- furniture, vehicles, sports and vegetables. Within each category, 12 of the words selected were from the typical end of the scale, and 12 from the atypical end. The words were further selected so that mean familiarity did not vary across the two typicality levels. Table 4.2 gives the mean typicality scores for each category, and collapsed across category, for typical and atypical words. Also in the table are the standard deviations and ratings ranges for each measure, for each category. As can be seen, across category there was some overlap between the ranges for typical and atypical words. Nevertheless, overall, both sets of words differed significantly on typicality alone -- $t(190)=16.8$, $p < .0005$. Associative frequency could not be equated across typicality because the number of words needed, per category, for this particular design exceeded those for which equal familiarity and associative frequency average scores could be obtained, within each typicality level. Mean associative frequency for typical words was 17.1 (range 0-50) and for atypical words 5.32 (range 0-37).

Each group of 12 typical or atypical words were subdivided at random into two separate lists, arbitrarily labelled "copy" and "memory". Copies of the lists used can be found in Appendix 5.

Subjects Sixty-four students at The City University served as paid volunteers. They were all native English speakers, for the most part reading subjects other than psychology, and were tested individually.

Procedure Response booklets were prepared containing alternate pages for the copying and recall tasks and appropriately labelled. Separate booklets were

Table 4.2 - Experiment 5: Summary description of the materials used. (S.D. = Standard deviation)

| Category | Typical lists | | | | | |
|---|---|---|---|---|---|---|
| | Typicality | | | Familiarity | | |
| | Mean | SD | Range | Mean | SD | Range |
| Birds | 1.49 | .22 | 1.19-1.84 | 1.49 | .19 | 1.23-1.90 |
| Clothing | 1.56 | .31 | 1.02-1.96 | 1.42 | .28 | 1.10-2.10 |
| Fruit | 1.62 | .22 | 1.30-2.09 | 1.63 | .25 | 1.28-2.09 |
| Furniture | 1.55 | .33 | 1.10-2.12 | 1.64 | .33 | 1.26-2.07 |
| Sports | 1.70 | .31 | 1.18-2.04 | 1.53 | .24 | 1.19-2.07 |
| Vegetables | 1.55 | .28 | 1.10-1.98 | 1.49 | .35 | 1.03-2.07 |
| Vehicles | 2.08 | .61 | 1.17-2.98 | 1.47 | .29 | 1.13-2.13 |
| Weapons | 1.67 | .31 | 1.07-2.09 | 1.64 | .29 | 1.23-2.09 |
| Overall | 1.65 | .39 | 1.02-2.98 | 1.54 | .29 | 1.03-2.13 |
| | Atypical lists | | | | | |
| Birds | 2.55 | .34 | 2.16-3.23 | 1.54 | .26 | 1.13-1.97 |
| Clothing | 2.93 | .35 | 2.20-3.51 | 1.44 | .14 | 1.30-1.76 |
| Fruit | 3.33 | .91 | 2.05-5.02 | 1.84 | .23 | 1.56-2.31 |
| Furniture | 4.04 | .82 | 2.39-5.14 | 1.62 | .22 | 1.29-2.03 |
| Sports | 3.01 | .52 | 2.47-4.16 | 1.53 | .17 | 1.23-1.81 |
| Vegetables | 2.75 | .61 | 1.96-3.57 | 1.54 | .45 | 1.00-2.33 |
| Vehicles | 3.98 | .54 | 3.13-4.89 | 1.52 | .20 | 1.19-1.94 |
| Weapons | 3.12 | .67 | 2.16-4.21 | 1.59 | .33 | 1.10-2.17 |
| Overall | 3.22 | .82 | 1.96-5.14 | 1.58 | .33 | 1.00-2.33 |

prepared for the distractor task, containing 3 pages, each page listing 32 different countries, in alphabetical order.

The subjects read typewritten instructions, which were then supplemented by a verbal explanation by the experimenter, who demonstrated the use of both experimental booklets -- response and distractor. The basic Method of Interpolated Attributes procedure was described, as outlined in the earlier experiments in this thesis. The subjects were additionally told that the words in both lists would be presented one at a time, and that they had to read each word aloud when it appeared on the VDU screen. For the copying task, they would have to read aloud and immediately copy down each word as it appeared. The fact that both lists belonged in the same category was emphasized to subjects in both experimental groups, and subjects in the Named group were further told that they would be given the category name prior to each list presentation.

All subjects were also required to read aloud the list titles which preceded list-word presentation. These were as follows:

(a) "Named" group:
- ** (Category name) **, followed by the memory list;
- "REPEAT ALOUD AND WRITE DOWN THESE WORDS", followed by the copy list.

(b) "Unnamed" group:
- ** "MEMORY LIST: SAY THESE WORDS ALOUD" **, followed by the memory list;
- "REPEAT ALOUD AND WRITE DOWN THESE WORDS", followed by the copy list.

The end of the memory list was indicated by a row of asterisks and was the signal for the subjects to pick up their pens and get ready to write down the copy-list words. Recall was signalled by the words "NOW RECALL THE FIRST LIST" printed at the top of the screen after the end of the copy list.

Each memory-list word remained on the screen for 1.5 seconds, and was followed by a .5 second blank gap before the next word was flashed on. Two seconds were allowed to elapse between each list title and the first list word, during which time the screen was always blank.

Each word in the copy list remained on the screen for 2.5 seconds, with a .5 second blank gap immediately afterwards. Pilot tests revealed that this time sequencing for the copy list was enouth to allow the subjects to perform the task accurately, but fast enough to prevent them from rehearsing the memory list.

Recall was written and unconstrained and lasted typically for about 45 seconds. After each recall test, the subjects were handed the distractor-task booklet and for a period of one minute were required to read aloud and write down the name of each country and its associated response. Blanks were allowed, if the subject could not think of anything, but the country name was always read aloud. The subjects started a fresh page in their distractor booklets after each recall.

All trials were started by the experimenter, after a warning to the subject, and each session (4 trials) lasted approximately 15 minutes.

### 4.2.3 Results

Figure 4.1 shows the mean recall percentages for all conditions, collapsed across category (left panel), and the mean percent release associated with each typicality level, for each of the two independent groups (right panel). Each data point on the graph is based on a total of 192 observations (32 subjects x 6 words per memory list). Table 4.3 lists the mean recall percentages and corresponding standard deviations across the 8 experimental conditions, and the percent release for each typicality level.

Table 4.3 - Experiment 5: Mean recall and percent release scores as a function of experimental condition.(DIFF = DIFFERENT condition.)

| | NAMED | | | | UNNAMED | | | |
|---|---|---|---|---|---|---|---|---|
| | TYPICAL | | ATYPICAL | | TYPICAL | | ATYPICAL | |
| | SAME | DIFF | SAME | DIFF | SAME | DIFF | SAME | DIFF |
| Mean % recall | 54.7 | 59.9 | 52.1 | 54.7 | 58.9 | 64.0 | 50.5 | 59.0 |
| Standard deviation | 20.5 | 17.1 | 21.1 | 20.9 | 21.6 | 17.2 | 19.8 | 15.0 |
| % release | 9.51 | | 4.99 | | 8.66 | | 16.83 | |

As Table 4.3 indicates, typical words were on average better recalled than atypical words, and shifting copy-list word typicality in the DIFFERENT conditions led to better recall for memory-list words. Thus both a typicality effect and a release effect appear to have obtained, though release levels differ quite markedly across condition.

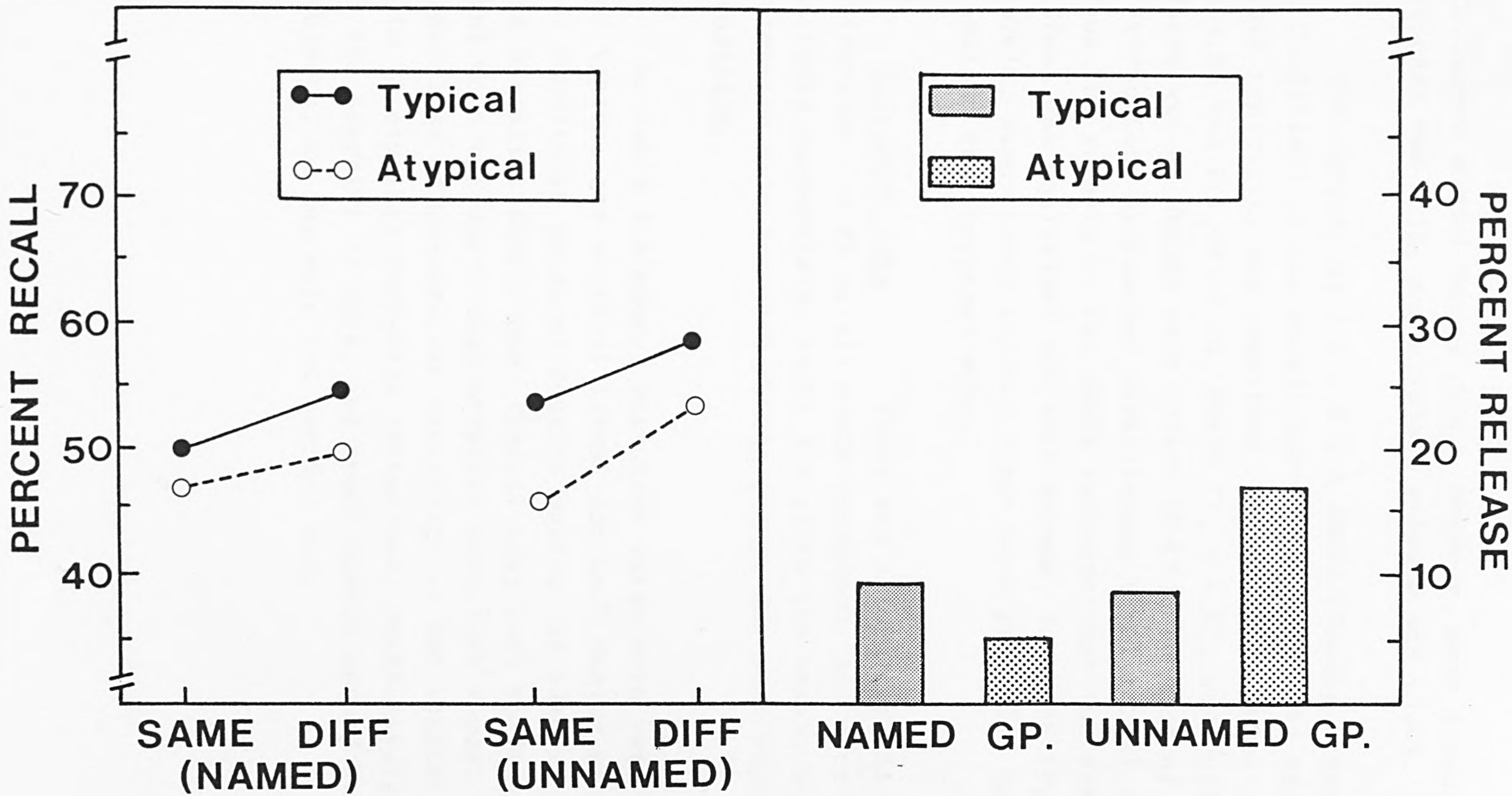



Figure 4.1 - Experiment 5: Percent recall and release from retroactive inhibition for typical and atypical words as a function of experimental condition and Naming group.

Collapsed across Naming group, however, mean typical word release was 9.1%, and atypical release was 10.9%.

The results of a 2 x 2 x 2 ANOVA (Naming x Typicality x Condition) of the recall scores revealed main effects of both typicality and condition, $F(1,62)=5.43$, $MSe=1.21$, $p < .05$ and $F(1,62)=9.62$, $MSe=0.71$, $p < .01$, respectively. There was no Naming main effect ($F < 1$) and none of the interactions approached significance ($F < 1$ in all cases). Thus the results of the ANOVA indicate that the release effect was equivalent for both groups, for both typicality levels, even though typical words were generally better recalled than atypical words.

Intrusion data There was a total of 221 intrusions (20.6% of all words produced). All were category-appropriate. Table 4.4 gives the breakdown of intrusion data for each Naming group, for each experimental condition.

As Table 4.4 shows, intrusion rates were very similar for typical and atypical lists, for both Naming groups, and all conditions produced roughly equivalent numbers of copy- and extralist intrusions. Typical copy-list words were somewhat more intrusive than atypical copy-list words, regardless of memory-list typicality -- 100 typical compared with 72 atypical copy-list intrusions, respectively. Because of the scarcity of data, and their uneven spread across subjects, no analysis was carried out.

Table 4.4 - Experiment 5: Breakdown of intrusion data. (DIFF = DIFFERENT condition)

| | NAMED GROUP | | | | TOTAL |
|---|---|---|---|---|---|
| | TYPICAL | | ATYPICAL | | |
| | SAME | DIFF | SAME | DIFF | |
| Copy-list intrusions | 25 | 10 | 23 | 22 | 80 |
| Extralist intrusions | 5 | 6 | 4 | 3 | 18 |
| Total | 30 | 16 | 27 | 25 | 98 |
| | UNNAMED GROUP | | | | |
| Copy-list intrusions | 24 | 20 | 19 | 29 | 92 |
| Extralist intrusions | 6 | 10 | 10 | 5 | 31 |
| Total | 30 | 30 | 29 | 34 | 123 |

### 4.2.4 Discussion

The results of this experiment demonstrate that an advantage in recall for typical words (a typicality effect) can still occur in the absence of familiarity differences (cf. McCloskey, 1980) and when both sets of words are encoded as members of the experimenter-chosen categories (no Naming main effect and no interaction with Naming, and all intrusions were category-appropriate). Thus two uninteresting explanations of typicality differences failed to receive support.

The data also show significant release from retroactive inhibition with shifts in list-word typicality. The levels of release are of the same magnitude for both typical and atypical words. These two main findings (two non-interacting effects, one of typicality, the other a release effect) are only partly predictable from a Cue-Overload viewpoint. The Principle can account for the release effect, since this shows that, in addition to the category cue, typicality was used as a cue for retrieval and became overloaded in the two SAME conditions (TYPICAL and ATYPICAL), leading to poorer recall by comparison with the DIFFERENT conditions. Cue-Overload has more difficulty predicting the typicality advantage. The results show that all words were encoded as members of the experimenter-chosen categories and that typical and atypical words load the cue to the same extent (equivalent release was obtained for both types of category member). Hence, for typical and atypical lists, the category-plus-typicality retrieval cues subsumed equal numbers of words at any given time. Yet, equivalent loading was translated into significantly different recall levels. The Principle cannot predict this effect. However, since it does not specify the relationship between the effectiveness of a given cue and the number of events it subsumes, the Principle can explain the data by arguing that different levels of recall may be associated with the two typicality cues used to retrieve both types of word, even though both subsumed equivalent numbers of items. Nevertheless, this explanation still leaves unaddressed the question of the direction of such a difference.

It is also possible that the typicality advantage came about as the result of differences in associative strength between the category cue and the list words (as opposed to differences in the extent of featural overlap between a word and another word, as determined primarily by typicality ratings). As mentioned in the Materials section, it was not possible to equate associative frequency across typicality level because the number of words of either typicality level needed to make up lists in this paradigm exceeded those for which both familiarity and associative frequency could be equated. Keller and Kellas (1978, Experiment 2) found that differences in associative frequency (Battig and Montague, 1969, scores) holding typicality constant did lead to differences in recall, with high associative-frequency words being better recalled. But in their case there was no release from proactive inhibition with shifts in associative frequency. Since symmetrical release was obtained at both associative frequency levels in Experiment 5, it is debateable whether associative frequency differences alone can account for the typicality effect, though they may have played a part. From a Cue-Overload point of view, however, recall differences between typical and atypical words would be expected if atypical words could not be retrieved easily via the category recall cue, even before it became overloaded. Before these data can be used to argue against a Cue-Overload account, it must first be demonstrated that an equivalent pattern of results can be obtained with associative frequency scores equated across typicality level.

Comparing now the results of Experiment 5 with those obtained by Keller and Kellas (1978, Experiment 1), it is obvious that only one aspect of their data (the typicality effect) was replicated here. The typicality effect, which occurs in both studies, was explained by Keller and Kellas in terms of the better encoding of typical items compared with atypical words, which are sufficiently unlike the category prototype that their encoding is inhibited. This explanation could also be applied to the results of Experiment 5 and indicates that, in order to account for the effects of typicality on recall, Cue-Overload needs to consider such factors as differences in accessibility which are not simply determined by the level of overload in any given cue. Chapter Five of this thesis discusses this suggestion in more detail.

The second finding of the Keller and Kellas study -- asymmetric release for typical and atypical words with shifts in typicality -- was not replicated here, even though total percent release was lower (though not significantly so) for typical than for atypical words in the Unnamed group (8.66% and 16.83%, respectively), but not in the Named group (typical word release 9.51%, atypical release 4.99%). Since Keller and Kellas did not provide the category names at any point in their experiment, comparison with the Unnamed group alone is probably fairer than comparison with both Naming groups in this experiment. However, the interaction between typicality and release is still non-significant ($F < 1$) when the results are analysed separately for the Unnamed group. So it would appear that the asymmetric release effect obtained by Keller and Kellas (1978) may be specific to the proactive inhibition paradigm. It could also be due to the greater sensitivity of this paradigm compared with the retroactive inhibition procedure (see, too, Chapter One, Section 1.2.1, p. 24).

## 4.3 Experiment 6: Typicality and part-list cueing

### 4.3.1 Introduction

Experiment 5 demonstrated that shifts in list-word typicality within the same semantic category can lead to significant release from retroactive inhibition using the Method of Interpolated Attributes. It also showed that release was equivalent for both levels of typicality, even though typical words were better recalled than atypical words. With the exception of the typical recall advantage, the pattern of results obtained could be explained in terms of Cue-Overload. Experiment 6 was designed to test the generality of these findings within another experimental paradigm, and included one further control over materials which had been lacking in the previous experiment: word associative frequency (as well as familiarity) was not allowed to vary across typicality level. This meant that any differences in recall observed across typicality could not be due solely to systematic differences in the strength of association between the category label and the list words.

The paradigm chosen for Experiment 6 was that of part-list cueing, first used by Slamecka (1968) and by Brown (1968). The following section describes the development of this testing method, and the findings of major experiments which utilized it.

#### 4.3.1a The part-list cueing paradigm

The initial studies using this paradigm (e.g., Slamecka, 1968) were designed to test the theory that items from lists just learned are stored in memory in a dependent fashion, being organised into chunks through horizontally-formed associations (cf. Wickelgren, 1976), a theory which can be traced back to Ebbinghaus (1885). Such a storage process would facilitate recall by enabling the subjects to retrieve words via the "chunks", each one of which would access several list words.

The theory described above would lead one to expect that if at recall some portions of the lists were re-presented, then recall for the remaining words in the list ought to be facilitated by comparison with the recall for those same lists by subjects receiving no such prompting. Re-presentation of some of the list words should enable subjects to gain access to at least some of the chunks which they might otherwise not have retrieved, thus improving their recall.

However, instead of the predicted recall facilitation, Slamecka (1968) found that his results suggested, if anything, poorer recall for cued, than for control subjects. It seemed that, rather than accessing otherwise unobtainable memory chunks, cueing actually blocked retrieval. Since these studies, several other researchers have reported similar retrieval blocks (e.g., Roediger & Neely, 1982).

A similar effect was reported by Brown (1968), this time within the domain of semantic memory. His subjects were asked to generate names of US states. Brown found that supplying some of the names required, prior to the generation task, tended to depress recall for the remaining states, compared with the performance of subjects who received no cueing.

Subsequent research concentrated on tightening the experimental procedure and testing for generality of the phenomenon. The method of scoring recall protocols from cued subjects became much stricter: rather than simply comparing gross recall percentages for cued and control subjects (which meant that the comparison was across different groups of words), later researchers compared the recall for the same words across both conditions (which required discarding at least some of the words produced by the control subjects). The inhibition effect was demonstrated for lists made up of rhyme sets (Mueller & Watkins, 1977, Experiment II), intuitive sets obtained by the same sorting technique used by Mandler (1967; Mueller & Watkins, Experiment III), and when the assignment of items to sets was totally arbitrary and acquisition was incidental, in a variation of the A-B, A-C paradigm (the A items served as the set labels, the responses as set instances, and assignment of responses to stimulus set was random). Inhibition with part-list cues also occurs for recognition, rather than recall (Todres & Watkins, 1981), and is present and not reduced even when the cues are extralist category members (e.g., Watkins, 1975; Mueller & Watkins, 1977; Roediger, Stellon & Tulving, 1977). This latter finding has led to its being re-christened "part-set cueing" effect.

It has been argued, nevertheless, that the inhibition in recall obtained with part-set cueing only occurs in those situations where subjects would have recalled some items even without the external cue. Where the part-list cue enabled subjects to recall categories which would not otherwise have been recalled, recall for the cued subjects was actually facilitated (cf. Tulving & Pearlstone, 1966). Roediger (1973) operationalized this finding by proposing that (1) retrieval cues will improve recall as compared with non-cued conditions when they allow access to more higher-order units than could be recalled unaided, and (2) retrieval cues will impair recall when they provide more information than is necessary to simply gain access to the higher-order unit.

Several uninteresting interpretations of the part-set cueing effect have been put forward. For instance, it was suggested that providing cues simply confused the subjects, or disrupted their normal order of output in free recall, thus lowering their performance (e.g., Basden, Basden & Galloway, 1977). Alternatively, cues were assumed to provide "non-specific interference" or simply to lower recall by delaying it. These arguments have all been criticized on theoretical grounds (e.g., Raaijmakers & Shiffrin, 1981; Roediger, Stellon & Tulving, 1977), but perhaps the strongest evidence against an artefactual explanation of the effects comes from a study by Mueller & Watkins (1977, Experiment I) in which, in addition to the usual category-appropriate cues, on one condition they also provided category-inappropriate cues. If the effect can be wholly explained in terms of delayed recall, disruption of normal output order, etc., then one would expect the category-inappropriate cues to have at least as much disruptive power as the category-appropriate ones. In fact, only the latter had any effect on recall -- the usual, inhibitory, effect.

### 4.3.1b Part-list cueing, Cue-Overload and predictions for Experiment 6

Inhibition with part-set cueing is one of the memory phenomena which can be interpreted in terms of Cue-Overload (cf. Chapter One, Section 1.2.1, p. 21 ). The Principle explains the recall impairment observed in terms of the effects of increased list-length for subjects receiving part-set cues. Thus, it assumes that re-presenting a list word prior to recall is functionally equivalent to adding that word to the study list, so that subjects in the cued condition have more words attached to their retrieval cues (category labels or "list") than subjects in the control condition.

If the results of Experiment 5 can be generalized across paradigms, than one would expect a part-set cueing effect to occur when groups of subjects are given lists of typical or atypical words to study and are then asked to recall them in the presence or absence of part-set cues. Furthermore, the inhibitory effects of providing part-list cues should be equivalent across typicality level, since the release effect observed in Experiment 5 did not interact with the typicality of the words used -- typical and atypical words loaded the cue to the same extent.

Also, since associative frequency was not allowed to vary across the two typicality levels, in Experiment 6, any advantage to typical word recall (typicality effect) observed in this experiment will not be due simply to differences in the strength of association between list items and the category label.

Encoding equivalence for typical and atypical words was ensured by providing the category labels before list presentation and always (for both conditions) at recall. This also permitted tight control to be exercised over category output order.

A study by Greenberg and Bjorklund (1981), using free recall of lists of typical and atypical words, needs to be considered here, in order to formulate a hypothesis concerning the replication of the typicality effect obtained in Experiment 5. Greenberg and Bjorklund found that provision at the learning stage of the category label for <u>each word</u> presented led to a drastic reduction of the typicality effect. They also equated associative frequency across typicality level. They interpreted this result as showing that at least some of the typicality effects reported in the literature may be due to differential encoding of atypical words with respect to the experimenter-selected category. Their results also show that significant typicality effects may be obtained, in the absence of associative frequency differences, so long as the category name is not present. At first sight, then, it would appear that Greenberg and Bjorklund's data should lead to the prediction of the loss of the typicality effect in the present study, given that category labels were provided. However, the design of Experiment 6 differs from Greenberg and Bjorklund's in at least one important respect: the category labels were not presented item by item. Rather, they were shown in a block <u>before</u> list presentation, and then at recall subjects had to follow the order indicated by the category labels in their recall protocols. Thus emphasis on the category label was not as heavy here as it was in Greenberg and Bjorklund's study. An alternative interpretation of their results, which they also report, is that emphasizing the category cues may have led to differences in recall strategy for subjects in the cued condition, who may have resorted to a simple generate-recognise strategy. No-cue subjects may instead have utilized typicality, not category, as a cue. The fact that the emphasis on category labels is much less pronounced in the present study may thus reinstate, at least in part, the typicality effect.

4.3.2 Method

Design The design used followed closely that of previous recent experiments (e.g., Todres & Watkins, 1981). Subjects were presented with long lists of words in which several natural categories were represented, with category membership being fully randomized rather than blocked. (Previous research -- e.g., Dong, 1972 -- showed that blocked presentation led to superior recall of the first category recalled when it was cued with the category name, compared with recall to successively cued categories.) Following list presentation, the subjects were required to recall as many words as possible, category by category, in the presence or absence of additional intralist cues. Control over category recall order was achieved by providing at recall, regardless of experimental condition, the appropriate category labels and requiring the subjects to follow a strict one-way order of recall

A 2 x 2 factorial design was used, with typicality between subjects and condition (Cued or Free recall) within subjects. The four experimental conditions were labelled as follows: TYPICAL-CUED and TYPICAL-FREE RECALL; ATYPICAL-CUED and ATYPICAL-FREE RECALL.

Materials Six categories were selected from the norms described in Chapter Three. From each category, a total of 16 words were selected such that each group of 16 could be split into two sets of 8 words, with non-overlapping typicality ratings and equivalent familiarity and associative frequency ratings. Table 4.5 (overleaf) lists the individual category means on all three variables, as well as their associated ranges and standard deviations. The categories selected were fish, sports, birds, vehicles,

Table 4.5 - Experiment 6: Summary of the typicality (TYP), familiarity (FAM) and associative frequency (AF) scores for the 6 experimental categories used.

| Category | TYP | | | FAM | | | AF | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | S.D. | Range | Mean | S.D. | Range | Mean | S.D. | Range |
| TYPICAL LISTS | | | | | | | | | |
| Sports | 1.62 | 0.38 | 1.00-2.00 | 1.46 | 0.31 | 1.13-2.06 | 12.5 | 5.43 | 5-19 |
| Birds | 1.52 | 0.22 | 1.18-1.77 | 1.53 | 0.19 | 1.29-1.90 | 11.6 | 6.32 | 8-23 |
| Vehicles | 2.28 | 0.62 | 1.17-2.98 | 1.48 | 0.33 | 1.09-2.13 | 9.1 | 7.40 | 3-23 |
| Insects | 1.48 | 0.13 | 1.35-1.65 | 1.73 | 0.35 | 1.37-2.34 | 17.1 | 8.89 | 8-29 |
| Vegetables | 1.65 | 0.22 | 1.15-1.94 | 1.37 | 0.25 | 1.03-1.70 | 16.3 | 3.41 | 10-20 |
| Fish | 1.69 | 0.18 | 1.26-1.92 | 1.85 | 0.46 | 1.35-2.45 | 13.8 | 8.14 | 8-31 |
| Overall | 1.71 | 0.41 | 1.00-2.98 | 1.57 | 0.35 | 1.03-2.45 | 13.4 | 7.02 | 3-31 |
| ATYPICAL LISTS | | | | | | | | | |
| Sports | 2.34 | 0.19 | 2.09-2.58 | 1.71 | 0.54 | 1.23-3.00 | 11.3 | 4.33 | 6-17 |
| Birds | 2.59 | 0.15 | 2.16-3.23 | 1.61 | 0.85 | 1.19-2.00 | 11.0 | 4.24 | 8-21 |
| Vehicles | 3.26 | 0.29 | 3.02-3.89 | 1.54 | 0.28 | 1.19-1.97 | 10.6 | 5.32 | 3-19 |
| Insects | 2.11 | 0.30 | 1.67-2.67 | 1.71 | 0.70 | 1.03-2.84 | 16.3 | 12.6 | 5-30 |
| Vegetables | 2.31 | 0.42 | 1.96-3.06 | 1.72 | 0.53 | 1.00-2.70 | 16.1 | 7.68 | 11-33 |
| Fish | 2.49 | 0.38 | 1.98-3.24 | 2.15 | 0.74 | 1.16-3.32 | 13.9 | 8.95 | 6-34 |
| Overall | 2.51 | 0.49 | 1.67-3.89 | 1.74 | 0.55 | 1.00-3.32 | 13.2 | 7.72 | 3-34 |

insects and vegetables. As can be seen, there is some overlap in the typicality ranges for typical and atypical words collapsed across category, though not within category. However, the difference in the typicality ratings between the two groups was still significant on a t-test ($t(94) = 8.76$, $p < .001$).

In addition to the six experimental categories, a seventh category -- countries -- was selected, this time from Battig and Montague (1969), to provide buffer words which, when inserted at the beginning and end of each list, would minimize the effects of primacy and recency recall advantages. Again, 16 words were chosen, at random, from this category and were arranged arbitrarily into two groups of 8 words.

Two lists were prepared, one Typical and one Atypical. Each list contained a random ordering of all 48 experimental words (8 words x 6 categories) and each started and ended with a group of four names of countries. Total list-length was thus 56 words. The lists were printed, column-wise, on acetate rolls, for presentation via an overhead projector fitted with one single, word-size, "window" over the light. Appendix Six contains copies of the lists used.

Response booklets were prepared as follows: Three different groupings of list words were selected to act as cues. This was done by taking each category in turn and selecting 4 of its eight words to be re-presented at recall. This ensured that each word was used as a cue and as a target at least once. These cue sets were then perfectly balanced across the 12 possible combinations of cueing and free recall condition, with the only constraint that conditions could not simply alternate, and that the same condition was not tested more than twice in succession. Of the 7 categories used, countries was always used as a practice category, always in the cued condition (though subjects were not aware of its being a practice test). Three of the remaining 6 categories were tested under cued and 3 under free recall conditions. Order of testing of each category was rotated across cueing order and cue sets. Thus, for every 12 subjects, all possible orderings of condition were used, yoked across two equal rotations of category order, for one set of category cues. Each category was used an equal number of times under each experimental condition. The balancing described above was the same for typical and for atypical lists. Half the subjects were tested with typical, and half with atypical lists.

Experimental response booklets were individually assembled, since the balancing of conditions and materials meant that each subject received a unique combination of condition order, category order and cue set. Each booklet contained, topmost, a detailed set of typewritten instructions, followed by the practice page (countries tested under the cued condition), followed by two pages, each subdivided into 3 different sections, for the recall of the 6 experimental categories.

The pages meant for the recall of words from the experimental categories contained, at the top of each section, the name of one of the categories presented, followed by (a) the phrase "Words recalled" and a space for written recall (free recall condition) or (b) four list words from that category, typewritten in upper case, followed by two pairs of dotted lines (the purpose of which will be described later), followed by the phrase "Remaining words" and a space for written recall (cued condition).

Subjects The subjects were 72 undergraduate volunteers at The City University, London. Approximately two-thirds were psychology undergraduates from all 3 years, who participated, unpaid, as part of a class test. The remaining subjects were drawn from other departments within the university, were paid for their assistance, and were tested either individually or in groups of up to 4 people. All had English as their first language.

Procedure The subjects were asked to read the instructions written on the topmost page of their response booklets, and these instructions were then supplemented by the experimenter with the help of dummy transparencies to illustrate the types of recall task being tested.

The subjects were told that they would be shown a long list of words, 56 in all, drawn from several categories (making it 8 words per category). Words from each category were randomly combined in the list with words from the other categories, but the category names would be provided before the list was presented. The subjects were further told that the list words would be presented, one at a time, at a fairly fast rate, and that they would be tested on their memory for the list after it had been presented.

The layout of the response booklets was then explained, and a dummy page was projected, to help explain the two types of task being tested. The subjects were informed that they would have to recall the words from each category separately, following strictly a pre-determined order -- the order in which the category names appeared on each page. In some cases, they would find only the category name, and had simply to write down as many od the 8 list words from that category as they could. In other cases, however, they would find, in addition to the category name, four of the list words presented, which belonged in that category. All they had to do then was to provide as many as they could of the remaining four list words for that category. At this point, the experimenter added that, so that she could be sure the subjects actually read the cues provided, the subjects would have to perform a little task on them. Specifically, they had to subdivide the 4 cues into two pairs on a "go-together-best" basis (as used by Todres & Watkins, 1981). An example was given, using a dummy category (names of towns), utilizing both an inclusion rule (A and B are X; C and D are Y), and an exclusion rule (A and B are X, C and D are not). The two rules were used so that subjects would find it easier to perform the task without wasting too much time. Subjects were asked to write down each pair of words on the dotted lines provided under the cue words, before going on to recall the remaining category instances. It was emphasized that the pairing-off task should be completed quickly and that roughly the same time should be devoted to each category.

List presentation followed, preceded by presentation of the seven category labels. These were shown in one block and remained on display for one minute. List words were shown individually at the rate of one word every 2½ seconds.

Recall was initiated by the instruction (given by the experimenter) to turn over the instructions sheet. The experimenter made sure, as far as possible, that the subjects did not flit backwards and forwards between categories, and that the pairing-off task was completed before category recall, in the cued condition. Recall was unconstrained and lasted typically about 10 minutes.

### 4.3.3 Results

Cued condition scores were derived from the total number of remaining words correctly recalled, and expressed as a percentage of the maximum number recallable for any one category (i.e., out of 4 possible words). Free recall condition scores were computed using two different criteria: The subjects were divided into matching pairs, on the basis of the order of cueing received, and the identity of the categories cued. Subject "a" within any given pair received a certain order of cueing for a given order of categories; subject "b" received the complementary order of cueing for the same order of categories. Thus the three categories cued for subject "a" were free recall categories for subject "b" and vice-versa. Free recalled words were then divided into two groups: those which had been used as cues for the matched subject, and those which had been targets for the matched subject. The strict scoring criterion considered only the latter group, and ignored the remaining words recalled. Recall was proportionalized on the maximum number of words recallable per category, defined in this way (i.e., out of 4 possible instances). The lenient scoring criterion considered both types of free recalled words, and expressed them as a proportion of the total words recallable per category (i.e., 8 words). Data were then collapsed across category, yielding one single score for each condition, for each subject. Figure 4.2 illustrates the mean recall percentages for typical and atypical words, as a function of cueing condition, using both scoring criteria. Each point on the left-hand panel is based on a total of 432 observations (4 words x 3 categories x 36 subjects), while in the right-hand panel the free recall data points are based on 864 observations (8 words x 3

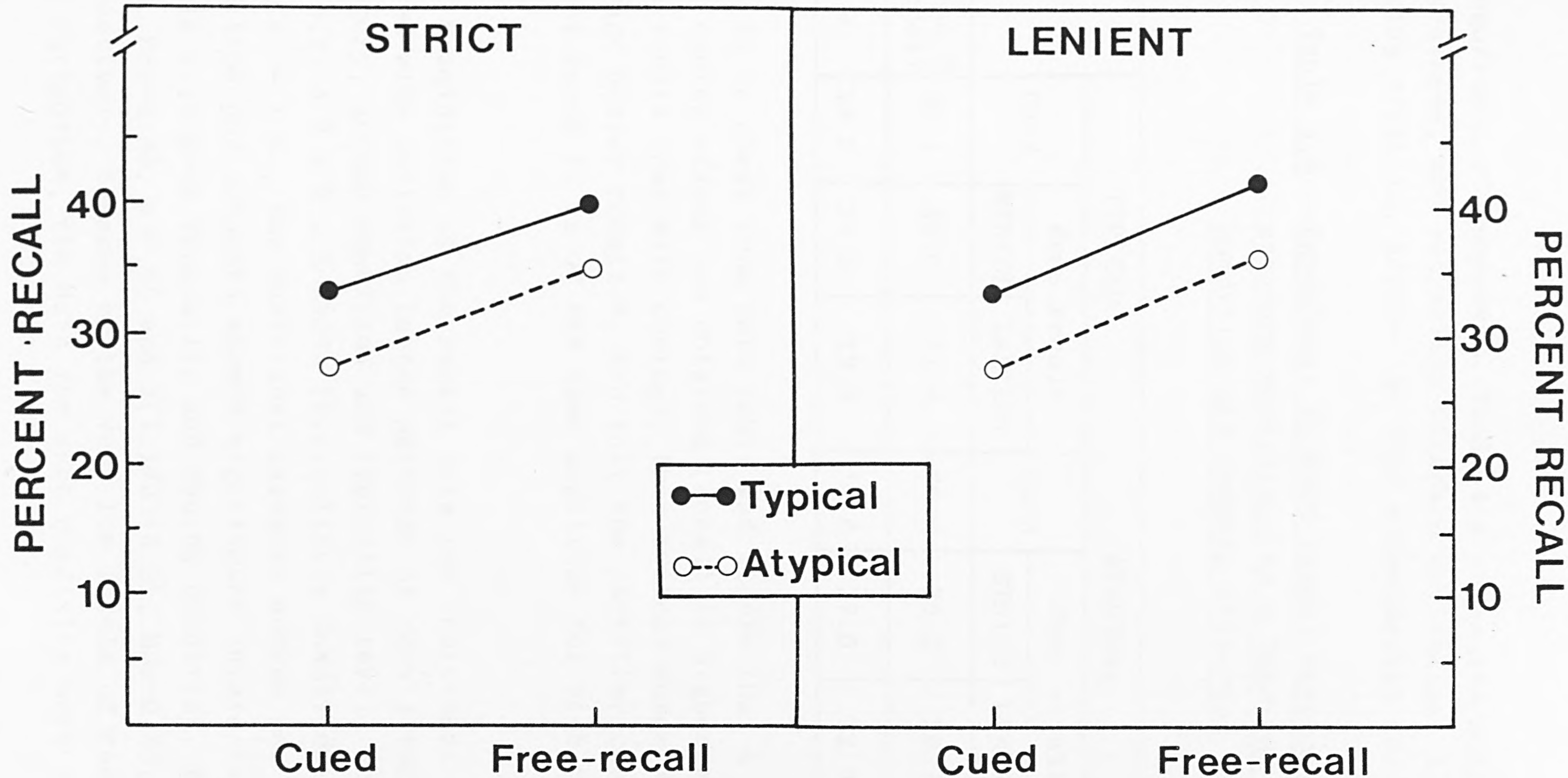



Figure 4.2 - Experiment 6: Percent recall for typical and atypical words as a function of experimental condition and scoring criterion.

categories x 36 subjects). Table 4.6 shows the mean recall percentages, and respective standard deviations, for both scoring criteria, across the four experimental conditions.

Table 4.6 - Experiment 6: Mean recall percentages and standard deviations as a function of condition and scoring criterion.

| | TYPICAL | | | ATYPICAL | | |
|---|---|---|---|---|---|---|
| | Cued | Free recall | | Cued | Free recall | |
| | | STRICT | LENIENT | | STRICT | LENIENT |
| Mean % recall | 33.1 | 40.0 | 41.4 | 28.0 | 35.2 | 36.7 |
| S.D. | 14.5 | 14.5 | 13.0 | 12.0 | 17.0 | 12.0 |

It is clear from both table and figure that a part-list cueing effect was obtained (recall is higher with free recall than with cueing), that typical words were on average better recalled, and that the part-list cueing effect seems to be of the same magnitude for each typicality level.

Examination of the recall data for individual categories showed wide variation in the patterns of recall for each category, across condition and typicality level, for each subject. A 2 x 2 x 3 ANOVA (Typicality x Condition x Trials -- i.e., the individual category scores per condition per subject) showed significant interactions of Trials with both Typicality and Cueing condition, $F(1,140)=3.03$, $MSe=0.88$, $p < .05$ and $F(1,140)=3.21$, $MSe=0.88$, $p < .05$, respectively. Because of the very low levels of recall for some categories, the data for each condition were collapsed across consecutive pairs of subjects and further analyses were carried out only on the macrosubject data.

The results of a 2 x 2 ANOVA (Typicality x Condition) of the strictly scored data show a main effect of condition, $F(1,34)=23.19$, $MSe=2.33$, $p < .001$, and a marginally non-significant main effect of typicality, $F(1,34)=3.84$, $MSe=6.69$, $.05 < p < .10$. The interaction is not significant, $F < 1$.

Lenient scoring of the data only affected the reliability of the typicality advantage -- the results of a 2 x 2 ANOVA similar to that carried out above showed significant main effects of both main factors: Typicality, $F(1,34)=5.25$, $MSe=83.38$, $p < .05$; Condition, $F(1,34)=36.38$, $MSe=35.89$, $p < .001$. The interaction was again non-significant, $F < 1$.

This latter result, with lenient scoring, is important, since it means that it is possible to obtain a reliable advantage in typical-word recall, which cannot be attributed to differences in familiarity and associative frequency, and when the subjects receive the category names before list presentation and again prior to recall. The fact that the effect is only significant with lenient scoring does not weaken the argument, since this type of scoring is closest to that used by Greenberg and Bjorklund (1981), who also considered all the words produced in their free recall condition. Thus, the results appear to support the interpretation of the reasons for the loss of the typicality effect in Greenberg and Bjorklund's study, described in Section 4.3.2. These were that the overemphasis on the category names produced differences in recall strategy in their two experimental conditions. Subjects cued with the category name used this as a cue for recall, whereas subjects who were not cued used instead typicality as their recall cue.

### 4.3.4 Category differences

Post-hoc examination of the data for the 6 experimental categories revealed one further interesting effect. By chance, three of the experimental categories used belonged to what Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976) labelled a "basic level", whereas the remaining three categories were at a "superordinate level" (see also Section 3.3.3, p.114 ). The three basic-level categories were birds, fish and insects. Analysis of the individual category (strict scoring) data for each category, for each subject, revealed a difference between the two types of category as regards the typicality advantage in recall. Specifically, the recall advantage to typical words was present for categories at the superordinate level (cued condition, $t(106)=2.19$, $p < .025$; free recall condition, $t(106)=1.44$, $p < .075$, marginally non-significant), but was totally absent for categories at the basic level ($t < 1$ in both cases). As was reported in Section 3.3.3, partial correlations indicated that for basic-level categories both typicality and familiarity were involved in determining the associative frequency (probability of generation) of a word, whereas for superordinate-level categories only typicality was an important predictor. Controlling word familiarity and associative frequency in the present experiment appears to have largely wiped out the typicality component for basic-level categories, leaving nevertheless the superordinate categories unaffected. Figure 4.3 illustrates this effect. Thus the results of Experiment 6 provide further, converging evidence for a distinction between the two types of category.

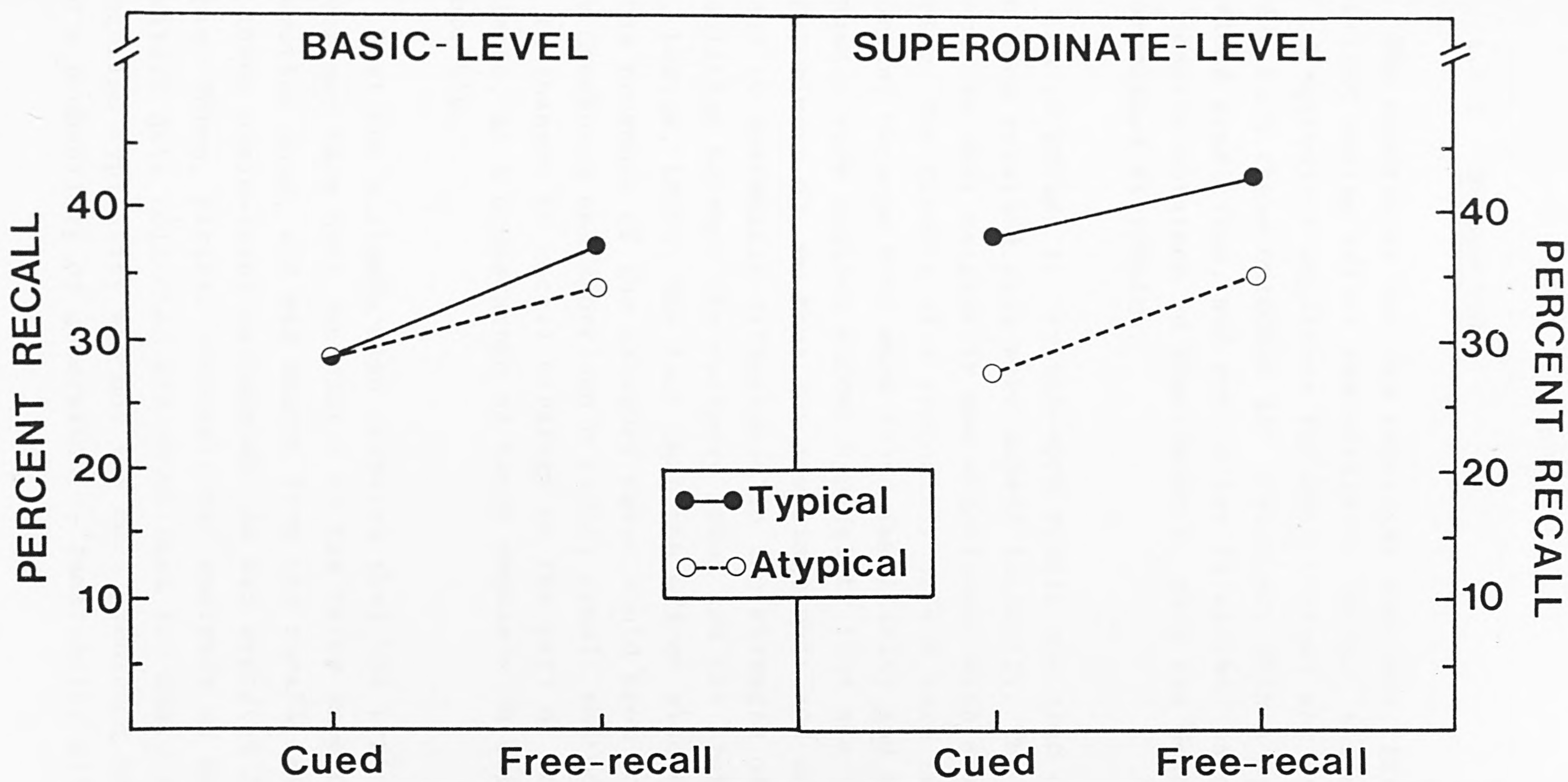



Figure 4.3 - Experiment 6: Percent recall for typical and atypical words as a function of experimental condition and abstraction level.

### 4.3.5 Discussion

The experiment had two important outcomes. First, a part-list cueing effect was obtained. Second, the effect was of equivalent magnitude for both typical and atypical words. Both these outcomes are consistent with a Cue-Overload prediction, and generalize to another paradigm the results obtained in Experiment 5, with the Method of Interpolated Attributes.

An advantage to typical-word recall was also obtained, when free recalled data were scored leniently. The same effect was only marginally non-significant with strict scoring. The finding of a typicality effect here is important because both word rated familiarity and associative frequency were equated across typicality (not the case in Experiment 5), so that the advantage obtained could not be due to systematic differences in the strength of association between the category label and the list words (cf. Loftus, 1975). The fact that the effect also occurred in the presence of the category names would seem to indicate that Greenberg and Bjorklund's (1981) result was possibly due to changes in recall strategy on the part of their cued subjects, as a consequence of heavy emphasis on category membership.

Post hoc analysis also revealed that the typicality effect may have been restricted to the three superordinate categories used, and was absent from the recall data for the three basic-level categories. As was reported in Chapter Three, partial correlational analysis of the normative data collected had shown that for basic-level categories typicality was not the sole predictor of a word's probability of generation -- familiarity also independently predicted associative frequency. This is in contrast with the pattern of correlations for superordinate categories, where typicality was the major predictor of associative frequency. In the present experiment, both familiarity and associative frequency were not allowed to vary with typicality. This seems to have had the effect of eliminating the effects of typicality at the basic level. This outcome suggests that the typicality effects normally observed for basic-level categories may be due to either familiarity or associative frequency differences. Though an interesting outcome, since it provides further, converging evidence for the distinction proposed by Rosch and associates, the effects of level on category recall need to be further investigated before any suggestions can be made about the nature of the recall difference.

None of the typicality differences described above could have been predicted by the Cue-Overload Principle. It could not predict the overall effect, since conditions and list length were equivalent across the two typicality groups, and it could not predict the differences in typicality effects across concept level. In both situations, the total number of word presentations subsumed under the category cue was the same. The Principle can only explain the effect of typicality on recall by arguing that, in spite of the category label being present, subjects used different cues for retrieval of typical and atypical words. Since the Principle does not specify the relationship between degree of overload on a given cue, and that cue's effectiveness for retrieval, different levels of recall may be associated with equivalent levels of overload for different cues. Though logical, this is a less than informative or parsimonious explanation. Cue-Overload appears to fail at two levels. First, it cannot specify which cue, or cues, will be used in any given situation. Whether a cue was used is determined entirely post hoc. In the present study, however, since typicality was a between-subjects factor and for both typical and atypical words the category labels were all that was supplied at the learning and recall stages, it would seem reasonable to suppose that, faced with a long list of assorted category items, and with a list of 6 category names, subjects in both groups should somehow have linked the two together and used category cues. And since overload on each category cue was equivalent for both groups, similar recall levels should have been obtained. This clearly did not happen.

One can, however, take a different stance. Maybe the category cue was simply not a good retrieval prompt for some of the words, even though the same number of word presentations was subsumed under the same cue, for both groups of subjects. This is the second shortcoming with a Cue-Overload interpretation -- it fails to take account of the manner in which a given cue, overloaded or not, can be utilized to gain access to the words it subsumes. It may be more profitable to use Cue-Overload as a convenient heuristic device to predict decline in recall with cueing, while at the same time accepting that in its present form it cannot predict or gainfully explain the results of experiments where cue efficiency is linked to differences in purely semantic memory components.

### 4.3.5a Theories of part-list cueing and the results of Experiment 6

The results of this experiment also have some bearing on the theories associated with the part-list cueing effect. Two of those theories, in particular, would have problems interpreting the pattern of results obtained. These are the models proposed by Rundus (1973) and by Raaijmakers and Shiffrin (1980; 1981).

Rundus (1973; see also Shiffrin, 1970; Roediger, 1973; 1974; 1978) proposed what has come to be known as a "sampling with replacement" model to account for retrieval blocks with part-list cueing. The model assumes a three-tiered organisation for categorized lists. Each tier is supposed to represent a "control element" (Estes, 1972). Within this hierarchical system, "list" represents the topmost level, the list-wide context or episode. This subsumes a second level, made up of the individual category names represented in the list. The control elements in this level determine access to the third level, where individual items are stored. According to Rundus, list items in this third level are linked to the category names in the level above by vertical associations which very in degree of strength, so that each third-level element is connected to its fellow list items via the control unit they all share at the next highest level in the hierarchy. Recall is assumed to take place through a sampling-with-replacement process. Order of recall is determined by the strength of association between the items and their immediate superordinate. In addition, the act of retrieving an item is assumed to increase that strength of association, thus increasing, too, the probability that the same item will be retrieved again. The memory system outlined above is supposed to contain a stopping mechanism, which ensures that recall is stopped after a certain, unspecified number of unsuccessful attempts to produce a new target item. Note that, since there are no horizontal links among items at any one level, the only way a cue can facilitate recall is by providing access to new, non-sampled, second-level control units (cf. Roediger, 1974). The inhibitory effects of part-list cueing come about because re-presentation of some of the list words at recall is assumed to increase the strength of association between these cues and the category names stored at the second level (re-presentation is thus equivalent to recall). This in turn means that the cues will have a higher probability of being retrieved than the target items, whose recall will consequently be poor -- the stopping rule will prematurely stop recall after the pre-set number of unsuccessful retrievals. Though the model does not provide fine detail as to what exactly is behind the strength of association between items at the two lower levels of the hierarchy, it would be reasonable to suppose that a dimension such as associative frequency would at least in part reflect such a relationship.

The second model, proposed by Raaijmakers and Shiffrin (1980; 1981) includes both vertical inter-item associations (as Rundus' model does), and horizontal associations. It aspires to be a general model of memory retrieval through Search of Associative Memory (SAM for short). It uses the concept of "probe cues" to initiate a memory search across a network of horizontal and vertical associations. Probe cues may be virtually anything, from items recalled to cues provided by the experimenter, to simple context cues associated with the list presentation. Recall is assumed to be initiated by the probe cues associated with the list context and then, as words start to be produced, each word in turn is added to the context cues and is used as a probe in its own right. Cues are used repeatedly, until they lose their effectiveness, or up to a preset parameter (labelled Lmax). As with Rundus' model, the act of recalling a word is seen to increase that word's strength of association, this time with the context cues used for recall. This process is labelled "incrementing". There is also a re-checking process, where all the cues are utilized again, until a Kmax parameter of unsuccessful retrieval is met, when recall stops.

Raaijmakers and Shiffrin's model is much more convoluted than Rundus' model when it comes to accounting for part-list cueing effects, even though it was specifically formulated to do so. It involves a series of parameters and countervailing processes to account for the inhibition effect, mostly because it includes horizontal inter-item associations. These mean that it has to explain how provision of part-list cues should not only nullify, but actually override the positive effects of part-list cueing due to these inter-item, horizontal associations, so that the net effect is negative (i.e., the end result is inhibition. In an attempt to simplify the argument, computer simulations were carried out (Raaijmakers and Shiffrin, 1981) to try to determine which, if any, of the many parameters postulated by the model were <u>not</u>, in effect, involved in the inhibition produced by part-list cueing. These turned out to be (somewhat interestingly) the very two processes which play such a central role in Rundus' account of the same effect: the incrementing rule and the stepping rule (the number of retrieval attempts allowed by the system resulting in no new recovered items). Retrieval inhibition with part-list cueing is then explained in terms of the recall strategies used in the two experimental conditions (cued and free recall). Subjects in the free recall condition are assumed to start off with a "richer" environment for recall, since they start with the context cues and will recall first those items which are most strongly associated with each other and with the list context. These items, in turn, will be those most likely to lead to retrieval of other list items in the same higher-order unit, or category. Cued subjects, on the other hand, do not have such an optimal start since, by definition, they will have at least one of the items in the cluster they begin their search with. Cueing is thus seen to disrupt recall order.

Raaijmakers and Shiffrin's model is far less efficient than Rundus' model in accounting for the part-set cueing effect. For example, inhibition with extralist cues, which presumably are not encompassed by the context cues used throughout the recall procedure, should be less than inhibition with intralist cues. The fact that the two effects have equal magnitude (demonstrated by Mueller and Watkins, 1977) is seen simply as a lucky coincidence -- an unparsimonious explanation at the best of times.

Both models, however, rely on differences in strength of association between items and superordinate control units, with stronger associations being recalled first and blocking weaker associations. Strength of association, though largely undefined by the models, appears to come about either through differences in associative frequency, or through incrementing processes linked to retrieval or re-presentation. Since in Experiment 6 associative frequency was not allowed to vary, the only other mechanism which could lead to increased strength of association would be the incrementing process via recall or re-presentation.

As the same numbers of words and cues were presented for both typical and atypical word lists, it then becomes difficult to interpret the differences in recall for these lists under the cued condition in terms of storage alone, as the models try to do. The only one of the two models which may be consistent with the data is the very same one which is most inconsistent with other part-set cueing evidence. Because they postulate horizontal inter-item associations, Raaijmakers and Shiffrin could account for the better recall of typical instances in terms of the strength of association between an item and another item. According to Rosch (e.g., Rosch and Mervis, 1975), typical words share more features in common with the category label and with each other than atypical words do, so their strength of association may also be higher. However, given these premises, one would predict a bigger effect of part-list cueing for atypical words in this experiment: Upsetting the richness of the recall environment for atypical words, by cueing, when that environment was not very rich to begin with, should have more severe consequences for recall than doing the same thing for typical words. Since typicality did not interact with the magnitude of the effect, one must conclude that neither of these two models can convincingly explain the results obtained in Experiment 6.

### 4.3.6 Conclusions

The results of Experiment 6 may be summarized as follows: A part-list cueing effect was obtained, of equal magnitude for typical and atypical words, even though recall for typical words was reliably higher with at least one method of scoring. Basic-level categories showed no advantage to typical word recall on either condition, but superordinate-level categories produced a typicality effect. This discrepancy accounts for the fact that the typicality main effect only just failed to reach significance on the analysis of the strictly scored data. These recall data provide further support for the distinction proposed by Rosch between basic- and superordinate-level categories, and are intriguing enough to warrant further research to try to establish the reasons behind the recall differences obtained. The results replicate and extend those obtained in Experiment 5, since they show (a) that the pattern of results found in the latter experiment can be generalized across paradigms, (b) that a typicality effect can still be obtained in the absence of associative frequency differences and (c) that provision of a category label does not necessarily lead to loss of the typicality effect. Thus Greenberg and Bjorklund's (1981) results may have come about because of changes in recall strategy due to the experiment's heavy emphasis on the category name.

A Cue-Overload account cannot predict the effects of typicality on recall, both at the general level (typicality main effect) and at the more detailed level (basic- and superordinate-category differences), since in both cases the category label subsumed equivalent numbers of words. As it stands, the Principle cannot predict (a) which cues will be utilized at any given time; (b) whether factors other than the total number of words presented may affect the way the cue is utilized. Both of these factors have to be taken into account in any comprehensive explanation of the results of the last two experiments.

## 4.4 Experiment 7: Cue-Overload and taxonomic category shifts within the same typicality level.

### 4.4.1 Introduction

Experiment 3 showed that release from retroactive inhibition, with a category shift, could be higher for typical than for atypical words (SAME-TYPICAL vs. DIFFERENT conditions, 12.2% release; SAME-ATYPICAL vs. DIFFERENT conditions, 6.5% release). However, there were some problems with interpreting the results of the SAME-ATYPICAL condition, because of the large imbalance in the ratio of typical to atypical words (3:11) in this condition. This might have led some subjects to adopt different cues (atypicality cues) for retrieval of the memory-list words. Experiment 7 tries to determine whether the asymmetric release effect is indeed reliable, using a design similar to that of Experiment 5. The design involved shifts in taxonomic category within two different levels of typicality -- i.e., the converse of the design used in Experiment 5. A Cue-Overload interpretation of the release effects normally found with shifts in category would lead one to expect equivalent levels of release for typical and atypical words, so long as the two types of item are encoded as members of the same set. In both cases the same number of word presentations would be subsumed under the category label and, in addition, both Experiments 5 and 6 showed that typical and atypical words load the cue by the same amount (typicality was always independent of the experimental manipulation used). As with Experiment 5, encoding equivalence was ensured by providing the category name prior to list presentation and, again as with that experiment, this Naming factor was introduced as a between-subjects factor, to make sure that subjects did indeed encode list words as members of the experimenter-chosen categories.

In order to support a Cue-Overload interpretation, then, one would predict a significant effect of condition (SAME or DIFFERENT category) for both levels of typicality, and no interaction.

4.4.2 Method

Design and Materials The Method of Interpolated Attributes was used, with taxonomic category as the shift variable. The design was a 2 x 2 x 2 factorial, with Typicality and Condition (SAME or DIFFERENT) as within-subjects factors and Naming (Named or Unnamed) as a between-subjects factor. This yielded 8 experimental conditions, labelled as follows: TYPICAL-SAME (NAMED) -- typical memory and copy lists, with category name provided, both lists belonging in the same category; TYPICAL-SAME (UNNAMED) -- the equivalent condition, without the category name being made explicit; TYPICAL-DIFFERENT (NAMED) and TYPICAL-DIFFERENT (UNNAMED) -- both memory and copy lists were made up of typical words, a different category being used for each, with and without a category label prior to the presentation of each list; ATYPICAL-SAME (NAMED), ATYPICAL-SAME (UNNAMED), ATYPICAL-DIFFERENT (NAMED) and ATYPICAL-DIFFERENT (UNNAMED) were the equivalent conditions for atypical lists.

The balancing used ensured that half the subjects received typical word-lists first, and half atypical lists; and half the subjects started with SAME, and half with DIFFERENT conditions. Categories were rotated across subjects and conditions such that each one was used an equal number of times under each condition. Two categories were used exclusively as copy lists for the DIFFERENT conditions and these, too, were rotated so that half the subjects had one category first, and the other half had it second.

The materials used were identical to those for Experiment 5, with the addition of the two categories used to make up DIFFERENT copy lists. Fish and insects were selected for this purpose. Twelve words were chosen from each of these last two categories, 6 typical and 6 atypical, with non-overlapping typicality scores. Mean rated typicality for the two lists of names of fish was 1.51 (typical) and 2.65 (atypical). The equivalent figures for insects were 1.44 and 2.68. Familiarity scores were 1.73 and 1.79, and 1.84 and 1.67, respectively.

Response and distractor booklets were made up, as for Experiment 5, and the list presentation was again via the VDU of a Commodore-PET microcomputer, with the following modifications to the title sequencing: "Named" group subjects were shown the category name of each list, memory and copy, prior to list presentation. "Unnamed" group subjects had the following titles introduced before memory and copy lists, respectively: "MEMORY LIST" and "COPY LIST".

Subjects Sixty-four students at The City University, London, volunteered to take part in the experiment and were paid for their assistance. Subjects were tested individually and were all native English speakers.

Procedure The procedure was exactly the same as for Experiment 5. Subjects in the Named group were forewarned that in some cases both memory and copy lists would share the same category, whereas in other cases two different categories would be represented. They were asked to read aloud the category label regardless of condition. Subjects in the Unnamed group also had to read the list titles aloud, but were not informed of any differences in the categorical membership of the lists.

### 4.4.3 Results

Figure 4.4 shows the mean recall percentages for each condition, and the final percent release for each "naming" group. Table 4.7 gives the standard deviations associated with each condition's recall data.

Table 4.7 - Experiment 7: Mean recall and release percentages as a function of experimental condition. (DIFF = DIFFERENT condition)

| | NAMED | | | | UNNAMED | | | |
|---|---|---|---|---|---|---|---|---|
| | TYPICAL | | ATYPICAL | | TYPICAL | | ATYPICAL | |
| | SAME | DIFF | SAME | DIFF | SAME | DIFF | SAME | DIFF |
| Mean % recall | 51.0 | 65.1 | 59.4 | 60.9 | 54.2 | 73.4 | 58.3 | 65.1 |
| S.D. | 22.8 | 20.2 | 15.0 | 19.3 | 16.2 | 15.5 | 16.7 | 16.8 |
| % release | 27.6 | | 2.6 | | 35.6 | | 11.6 | |

Each condition's score is based on 192 possible observations (32 subjects x 6 words). The right-hand panel in Figure 4.4 seems to indicate that release was substantially higher for typical than for atypical words across both Naming groups, as can also be seen in the very different slopes of the recall data shown on the left-hand panel, for both typicality levels.

The results of a 2 x 2 x 2 ANOVA (Naming x Typicality x Condition) showed a main effect of Condition (SAME or DIFFERENT), $F(1,62)=18.94$, $MSe=1.32$, $p<.001$; and a

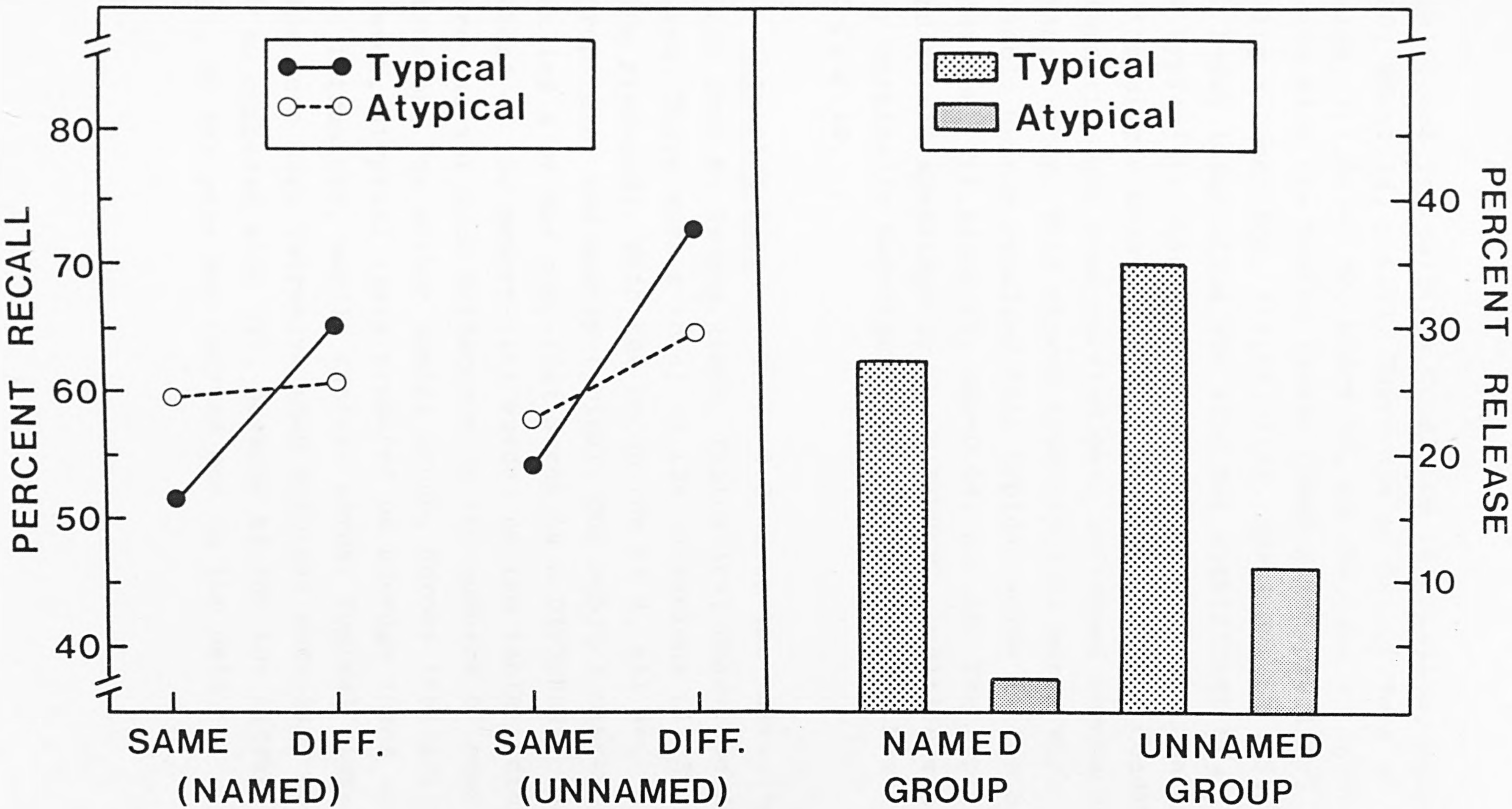



Figure 4.4 - Experiment 7: Percent recall and release from retroactive inhibition for typical and atypical words as a function of experimental condition and Naming group.

significant Typicality x Condition interaction, $F(1,62)=7.89$, $MSe=1.14$, $p<.01$. There was no main effect of Naming, $F(1,62)=2.39$, $MSe=1.28$, $p>.05$, and no interaction with the Naming factor (Naming x Typicality, $F<1$; Naming x Condition, $F(1,62)=1.19$, $MSe=1.32$, $p>.05$), and the 3-way interaction was also not significant ($F<1$). The Typicality main effect was not significant. An analysis of the simple main effects of the Typicality x Condition interaction was then carried out, collapsed across the Naming factor. This showed that atypical words were reliably better recalled than typical words in the SAME condition, $F(1,62)=4.79$, $MSe=0.94$, $p<.05$. The typical-word recall advantage in the DIFFERENT condition was only marginally non-significant, $F(1,62)=3.63$, $MSe=1.24$, $.05<p<.10$.

Intrusion data Table 4.8 shows the intrusion data broken down by Naming group, Typicality, Condition and Source. There were a total of 134 intrusions (12.5% of all words produced). With the exception of 4, all were category-appropriate, and mostly typical. Ons subject mistakenly recalled 4 of the copy-list words in a DIFFERENT condition, instead of the memory-list words. As the table indicates, there was not much difference in the numbers of copy-list intrusions for either Naming group, across typicality. However, atypical lists produced on average fewer extra-list intrusions, mostly typical words. Typical copy-list words were more intrusive than atypical copy-list words (49 as compared with 39). Because of the low intrusion rate, no analysis was carried out on the data.

Table 4.8 - Experiment 7: Breakdown of intrusion data.

| SOURCE | NAMED | | | | TOTAL |
|---|---|---|---|---|---|
| | TYPICAL | | ATYPICAL | | |
| | SAME | DIFFERENT | SAME | DIFFERENT | |
| Copy-list intrusions | 23 | 0 | 17 | 4 | 44 |
| Extralist intrusions | 10 | 9 | 3 | 3 | 25 |
| Total | 33 | 9 | 20 | 7 | 69 |
| UNNAMED | | | | | |
| Copy-list intrusions | 26 | 0 | 22 | 0 | 48 |
| Extralist intrusions | 4 | 7 | 2 | 4 | 17 |
| Total | 30 | 7 | 24 | 4 | 65 |

#### 4.4.4 Discussion

The results of Experiment 7 show that, with a taxonomic category shift, release from retroactive inhibition is much higher for typical than for atypical category members. Atypical lists show only minimal release. This asymmetric release effect was associated with a loss of the typicality recall advantage in the SAME condition, where atypical words were actually better recalled than typical words. This result is at odds with the results of Experiments 5 and 6, and with the theories which have been proposed to explain typicality effects on recall -- Greenberg and Bjorklund's (1981) and Keller and Kellas' (1978) theories, for example. Greenberg and Bjorklund (1981) explain the typicality advantage in recall in terms of the featural overlap among typical category instances, which is assumed to create featural-overlap determined retrieval pathways among category items. Since atypical words do not have such extensive featural overlap, they cannot be retrieved through these pathways and must be recalled independently via the category name. Keller and Kellas (1978) interpret the typicality advantage in terms of the poorer encoding of atypical category instances, because they are so different from the category prime. Both these theories would predict that atypical words would always be worse recalled than typical words. They clearly were not, in this experiment.

The results of Experiment 7 are, however, consistent with an interpretation which assumes that atypical words were retrieved via many more cues than typical words, in the SAME condition of Experiment 7. Since these cues were much less overloaded than the category cue used for retrieving typical category members, recall for atypical words was porportionately higher in this condition. Why subjects should opt for such a strategy is not altogether clear. It could be that the category shifts associated with half the trials they did served to emphasize the oddness of atypical words as members of the experimenter-chosen categories, and this, added to the large number of atypical words in the SAME condition, made the subjects reason that those features of the atypical words which make them poor category members (e.g., atypicality features or cues referring to an item's membership in a category subset -- "flightless birds", "salad vegetables", etc.) would be much better cues for recall than the category label -- as indeed they were. In this respect, the ATYPICAL-SAME condition of Experiment 7 was very similar to the equivalent (SAME-ATYPICAL) condition of Experiment 3: there, too, when large numbers of atypical words were present, subjects preferred to use atypicality cues to the general category cues they used to retrieve typical category members. In the ATYPICAL-DIFFERENT condition, the shift in category (together with the smaller number of atypical words from one single category) may have promoted recall via the category cue, possibly by calling attention to the same-category membership of the memory-list words.

This interpretation of the interaction obtained (that it was due to the abnormally high recall of atypical words which were recalled via multiple cues) is more consistent with a Cue-Overload account. Overload can explain the release effect obtained with typical lists, and the use of additional cues to retrieve atypical words in the ATYPICAL-SAME condition would mean that overload on any one cue was much lower, facilitating retrieval. However, the principle still cannot <u>predict</u> the effect, largely because it cannot predict which cues will be used.

Note, too, the effects of <u>context</u> on choice of cues for retrieval. The design used in Experiments 5 and 7 means that the SAME conditions in both experiments overlapped exactly, both in terms of materials and of presentation order, differing only in the general experimental context. In one case (Experiment 7) they were tested in the context of shifts in category within typicality, in the other (Experiment 5) they were tested in the context of shifts in typicality within category. The pattern of results obtained for both SAME conditions, across experiments, was strikingly different, as Figure 4.5 illustrates. Cue-Overload makes no provision for context. Atypical words were obviously not retrieved via the same set of cues in Experiments 5 and 7, even though the lists were identical. In one case they were poorly recalled (Experiment 5), in the other they were well recalled (Experiment 7), depending on whether emphasis was placed on typicality or category (whichever was shifted in the DIFFERENT condition). This pattern of results was significant on a 2 x 2 x 2 ANOVA (Experiment x Naming x Typicality): The Experiment x Typicality interaction was significant, $F(1,178)=9.30$, $MSe=0.83$, $p < .01$. All the other factors and interactions failed to reach significance ($F < 1$ in all cases bar the Typicality x Naming interaction, $F(1,178)=1.70$, $MSe=0.83$, $p > .05$).

Taken together with the findings of Experiments 3, 5 and 6, the results of Experiment 7 illustrate the need to "flesh out" the original formulation of the Cue-Overload Principle if it is to be of any use in further research using typicality or other semantic dimensions. In its present form, the Principle cannot predict either the effects of typicality on recall, or the effects of context on choice of retrieval cues. Both these factors are critical for a comprehensive interpretation of the results described so far.

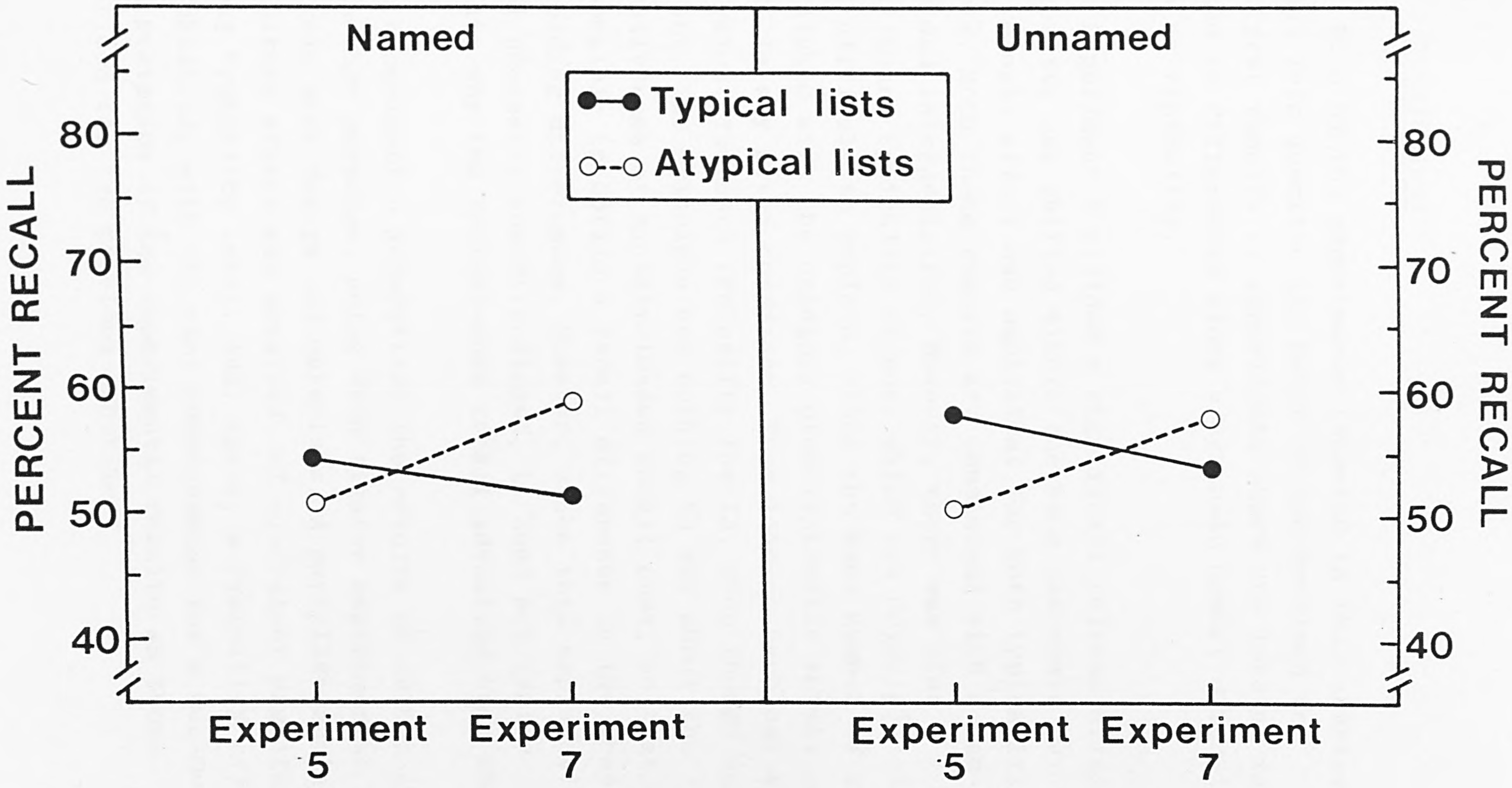



Figure 4.5 - Experiment 7: Percent recall for typical and atypical words in the overlapping conditions of Fxperiments 5 and 7.

## 4.5 Conclusions

Each of the experiments reported in this chapter brings into question the power of Cue-Overload to interpret results of experiments where cue loading is linked to differences along a semantic memory dimension such as typicality.

Experiment 5 yielded a significant release effect when typicality was shifted within the same taxonomic category. The release effect was equivalent for both typicality levels. Both these results are consistent with a Cue-Overload interpretation. However, there was also a significant typicality effect, which the Principle finds more difficult to explain, since the same number of words was linked with the category-plus-typicality recall cues used, in any given condition. Thus load on each cue was equivalent for both typicality levels, even though recall was not. The Principle has nothing to say about the effectiveness of equally-loaded recall cues, so that it is possible to obtain a recall difference in the absence of loading differences. However, since this explanation is of necessity non-directional, it does not really explain why the typical-word recall advantage came about.

Experiment 6 generalized the results described above to another paradigm, using even tighter experimental controls over design and materials. A part-list cueing inhibitory effect was obtained, of equivalent magnitude across typicality level. And, again, a typicality effect was obtained, with the same consequences for a Cue-Overload interpretation of the experimental results as those described for the previous experiment.

Finally, Experiment 7 showed that the effect of release from retroactive inhibition with a category shift interacted with list-word typicality -- it was much bigger for typical than atypical category members. This again highlighted the lack of predictive power of the Cue-Overload Principle (this time its inability to predict the effects of context on choice of cues for retrieval). The Principle could, however, accommodate this pattern of results by assuming that category cues were not the main cues used for recall in one of the experimental conditions, possibly because the large number of clearly atypical items from the same category made subjects resort to atypicality cues for retrieval. A similar effect was also obtained in the equivalent condition of Experiment 3, where atypical words outnumbered typical category members.

The inability of the Cue-Overload Principle to predict, or usefully explain, the results obtained stems from two different sources: It cannot specify which cue will be used other than by post hoc examination of recall data; and it does not take into account the manner in which a cue may be utilized, even if overloaded (for instance, it does not consider the possibility that strong associates may be better retrieved, even through an overloaded cue, than weak associates). Chapter Five now outlines one possible way in which Cue-Overload could be extended to be able to cope with the results of the experiments described above.

# CHAPTER FIVE : CONCLUSIONS AND THEORETICAL IMPLICATIONS

"Law of Superiority:

The first example of [a] superior principle is always inferior to the developed example of [the] inferior principle."

" Murphy's Law and other reasons why things go ¡ƃuoɹʍ" - A. Bloch (Ed.), p.78

## 5.1 Introduction

In this chapter, the main findings of the experimental work reported are summarized (Section 5.2) and their implications for a Cue-Overload interpretation of memory phenomena are considered (Section 5.3). A proposed Similarity/Accessibility Hypothesis is outlined (Section 5.5), based on Tversky's structural model of similarity (Section 5.4), as an additional, complementary process to Cue-Overload theory. This extended version of the theory is then evaluated (Section 5.6) and its consequences for future research are outlined (Section 5.7). Finally, the major theoretical conclusions reached on the basis of the research reported are described (Section 5.8).

## 5.2 Summary of the experimental work reported

Chapter One described the aim of this thesis as testing the generality of the Cue-Overload Principle, both at a broad level, and in the specific case where the strength of association among to-be-recalled items, and between these and the retrieval cues used for their recall, was systematically varied. The first two experiments laid the groundwork for this latter line of research, by establishing the replicability of the effects of cue overload, as well as testing for the predictive power of the Cue-Overload interpretation of memory phenomena. Both these studies involved the release from retroactive inhibition paradigm, using the Method of Interpolated Attributes described by Watkins and Watkins (1976; see Section 1.5, p.33 ).

Experiment 1 replicated the results of a taxonomic category shift experiment described by Watkins and Watkins (1976) and showed, too, that increasing the length of the interference list produces a larger release from retroactive inhibition effect. This result supports a Cue-Overload interpretation of the effects of retroactive inhibition. However, the intrusion data suggested that the Method of Interpolated Attributes also involves a discrimination process (between memory and interference lists), so that it is possible that part of the recall deficits observed in this paradigm, which are attributed simply to lack of accessibility through cue-overload, may be due as well to lack of discriminability.

Experiment 2 replicated two of the conditions of the Gardiner, Craik and Birtwistle (1972) release from proactive inhibition study, using shifts in category subsets. Gardiner et al. showed that release was only obtained when subjects were cued with the category-subset names. Experiment 2 generalized these results to the release from retroactive inhibition paradigm, thus supporting a common interpretation of the two paradigms in terms of Cue-Overload. However, the intrusion data also suggested that poorer recall for non-cued lists might have been due, in part, to the subjects' inability to discriminate well between memory and interference lists, to which the general category cues applied equally well.

Experiment 3 looked at the effects of varying the degree of relatedness between interference-list words and the category recall cue, again using the Method of Interpolated Attributes. It also tested for release from retroactive inhibition, with a category shift, at two levels of memory-list word typicality. It was noted that the design used may have promoted the use of different recall strategies, preventing a comprehensive interpretation of the data in terms of Cue-Overload. For the two conditions where the same recall strategy appears to have been used, release from retroactive inhibition with a category shift was equivalent for both typical and atypical target words, and typical items were recalled better than atypical items. Only the first of these results was predictable in terms of Cue-Overload.

The results of Experiment 3 also demonstrated the need for tighter control over experimental materials. Because factors such as familiarity and associative frequency had not been equated across typicality level, it was not possible to say for sure what lay behind the typical-word recall advantage obtained. It became clear that, before further research could be carried out, better sources of materials had to be obtained, since all the available published norms not only refer to different word samples, collected at widely different points in time, but are also based on American subject samples. Chapter Three described the collection of such normative data.

Experiment 4 provided norms of rated typicality, rated familiarity and associative frequency for a 531-word corpus, divided among 12 semantic categories. Comparison of these data with the equivalent American norms of typicality (Rosch, 1975a) and associative frequency (Battig and Montague, 1969) supported the argument for collection of new norms. It was found that the correlations between the corpus and the American norms were lower than the split-half correlations carried out to ascertain the inter-subject reliability of the corpus data. Further, typicality seemed to be less affected by differences in cultural background and collection time than associative frequency was.

It was also possible to analyse in a systematic way the intercorrelations among the three dimensions of internal category structure. In particular, it was possible to test the relative importance of rated familiarity in determining typicality, and to check which of these two variables -- typicality or familiarity -- was most predictive of the probability of generating a word to the category label. It was found that, contrary to McCloskey's (1980) argument, familiarity was not the crucial factor behind an object's rated typicality. Familiarity was also not the major determinant of associative frequency, at least for superordinate categories. It did, however, predict generation probability at the basic level at least as well as typicality did. The findings reported in Chapter Three thus extend our understanding of the nature of typicality and test some of the theories put forward to account for typicality effects. They also provided a stable data-base from which materials could be selected for further experimentation to test the generality and predictive power of the Cue-Overload Principle. The results of these experiments were reported in Chapter Four.

The three experiments reported in Chapter Four improved on the design used in Experiment 3, while asking essentially the same questions: First, does the strength of association between an item and its recall cue, and among items subsumed by the same cue, affect the degree to which that cue becomes overloaded? And, second, can Cue-Overload predict the patterns of results obtained when associative strength is systematically varied?

Experiment 5 tested for release from retroactive inhibition with shifts in rated typicality within the same semantic category. The results showed that typicality was used as a functional cue for recall and that typical and atypical category members placed equivalent loads on the category cue. However, typical words were consistently better recalled than atypical words. The experiment also showed that typicality differences could be obtained when typical and atypical words were encoded with respect to the experimenter-chosen categories, and when familiarity was not allowed to vary systematically across typicality level. Thus two uninteresting accounts of typicality differences were not supported (see Chapter One, Section 1.4, p. 31). Cue-Overload could predict the release effect obtained -- typicality was used as a functional retrieval cue -- but it could not predict the typicality effect.

Experiment 6 extended the results of Experiment 5 across paradigms, with even tighter controls over the experimental materials. The part-list cueing paradigm was used, and both a part-list cueing inhibitory effect and a typicality effect were obtained. Further, the typical-word recall advantage occurred in the absence of encoding, familiarity and associative frequency differences across typicality level. It was also found that the typicality effect obtained only for the superordinate-level categories used in the experiment. The three basic-level categories showed no difference in recall for typical and atypical words. As was the case with Experiment 5, these results offer only partial support to Cue-Overload theory. Whilst it can predict and explain the effects of part-list cueing, it cannot predict any of the effects of typicality on recall.

Experiment 7 used the Method of Interpolated Attributes and tested for release from retroactive inhibition with shifts in category at two levels of rated typicality. The results showed release only for typical words -- atypical words showed very little effect of shifts in the category-membership of the interpolated list. Again, Cue-Overload cannot predict the effects of typicality on recall (a typicality effect was obtained in the shift conditions), nor can it predict the interaction between typicality and release. It was suggested, however, that in one of the conditions atypical list words might have been retrieved via multiple cues (for instance, atypicality cues), since the large number of atypical words from the same category, in that condition, could have emphasized the usefulness of such a retrieval strategy. Interpreted in this fashion, the results can be explained in terms of Cue-Overload. Nevertheless, though logical and well-supported by the data, this explanation was post hoc. The Principle still cannot predict the results obtained, which appear to have been brought about by context determining the cues to be used for retrieval.

## 5.3 Critical analysis of the Cue-Overload Principle in the light of the results described.

As formulated by Watkins and Watkins (1976; see, too, Watkins, 1979), the Cue-Overload Principle argues that recall probability is an inverse and monotonic function of the number of word presentations subsumed by a recall cue. In this very general form, the Principle can account for data from a variety of paradigms (see Chapter One, pp.19-24). The results summarized in the previous section highlight two main shortcomings of the Cue-Overload Principle, which greatly undermine its predictive power.

First, they show that even though the prior strength of association between an item and other items subsumed by the same cue does not affect the degree to which that cue gets overloaded, this equivalent loading is translated into consistently different recall levels. Equating the strength of association between each item and the recall cue, while systematically varying inter-item associative strength also produces the same pattern of results, so that the recall difference cannot be due simply to the prior lack of effectiveness of a cue, even when not overloaded, for retrieval of target items. While these results cannot be predicted by Cue-Overload, they can be made to fit the Principle by postulating different cues for retrieval of typical and atypical words. Since the Principle does not actually specify the relationship between cue-effectiveness and cue-loading, for equally-loaded cues (beyond saying it is inverse and monotonic) it is then perfectly consistent with the finding of differences in recall associated with equivalent degrees of overload. However, this is a less than informative or parsimonious explanation, particularly in the case of Experiment 6, where typicality was a between-subjects factor and only the category names were provided to each group of subjects.

Second, as the results of Experiment 7 demonstrate, experimental context can lead to the reversal of the typicality effect which was obtained, with the same materials and conditions, in Experiment 5. As stated at present, Cue-Overload does not have anything to say about the effects of context on recall, even though it is possible to argue that context may determine which cues are used for retrieval in any given situation. In this respect, the Principle is highly circular: certain cues are assumed to have been overloaded if recall for words which could potentially be linked with them is poor. There is in the Principle's formulation no provision for an independent means of testing which cues will, in effect, be used. While this circularity means that a wide variety of data can be made to fit the Principle, it greatly undermines its predictive power.

It may, however, be slightly unfair to try to use the original formulation of the Cue-Overload Principle to try to predict the effects on recall of variables about which the Principle, as stated, has nothing to say. Cue-Overload was not formulated with factors such as rated typicality specifically in mind. It could be that, just as the Levels-of-processing approach required the addition of concepts like elaboration and distinctiveness before it could become a comprehensive theoretical approach (see Cermak & Craik, 1979), so Cue-Overload will need to be similarly extended before it can gainfully be used to account for the kind of data presented here. It is, after all, the case that most of the results described support a Cue-Overload interpretation. Section 5.5 will outline one such possible addition to the Cue-Overload formulation, based on Tversky's contrast model of similarity, which will now be described.

## 5.4 Similarity, context and typicality

Rosch's theory of typicality (e.g., Rosch, 1975a; 1978; Rosch & Mervis, 1975) was initially formulated as a structural, rather than processing model. Rosch viewed it as placing constraints over processing models rather than as providing a comprehensive account of the uses of typicality for recall.

According to Rosch, categories are analogue in nature, that is, a category's internal structure represents the degree of relatedness of various category members to the category itself. Category members are assumed to be organized around prototypes, and to vary in their degree of proximity to the prototype. This degree of proximity is seen to reflect the family resemblance among items and between items and the prototype. Thus Rosch veers away from the traditional view of categorization as simply the compilation of objects sharing the same attributes, and instead presents a view of category structure where many of the attributes which are reported as relevant for category membership are in fact not common to all category members (Rosch & Mervis, 1975). Rosch's theory is innovative in yet another respect. It takes account not only of an item's membership in a given category, but also of its lack of membership in contrasting categories. According to her, the more an item is judged to be prototypical of a given category, the more attributes it has in common with other category members, and the fewer attributes in common with members of contrasting categories. The converse relation also applies: the more atypical of a category an item is judged to be, the fewer attributes it shares with members of its ascribed category and the more attributes it has in common with members of other categories. Differences in "family resemblance" or "attribute overlap" among category members have been more formally defined by Tversky as reflecting the degree of "similarity" among category members (e.g., Tversky, 1977; Tversky & Gati, 1978). Like Rosch's theory of typicality, Tversky's "contrast model of similarity" is a structural (as opposed to processing) model, and it overlaps considerably with that theory. It differs from Rosch's account of typicality in that it relies heavily on the effects of context in determining the perceived similarity among objects. Since one of the criticisms raised in the previous section concerning the Cue-Overload Principle was that it could not explain the effects of context on choice of cues for retrieval, Tversky's model was used here, instead of Rosch's theory, to try to extend the original Cue-Overload formulation and make it more consistent with the results described. The following is a very brief outline of Tversky's contrast model of similarity.

Tversky argues that the ability to sort things, events or situations into distinct classes or categories is determined by one's ability to recognise the similarities among those objects or events. Similarity is thus seen as an organizing principle behind classification and concept formation, and as such encompasses most aspects of a person's interaction with his or her environment. Tversky adopted a feature-theoretical approach to the analysis of similarity relations. He specifically rejected the traditional geometric model of similarity, having shown that, at least for some types of materials, the geometric model's three basic axioms -- minimality, symmetry and triangle inequality -- could be systematically broken. Instead, Tversky chose to express similarity between objects as a matching function, i.e., a function that measures the degree to which two sets of features match each other. Choice of which features are to be matched is determined by the task itself. Hence, prior to task performance, sets of task-appropriate features are selected from the total feature domain representing an item or set of items. Similarity among the objects being considered is then defined in terms of the common features they share, as well as in terms of the items' distinctive features (i.e., those that are common to some but not to others, and vice versa). The two sets of features are linearly combined, and the product of this combination is termed "scale f", which defines the similarity among items in the feature space defined by the task. Because both types of features are combined, the model is known as a "contrast model". Logically, similarity increases with the measure of the common features and decreases with the measure of the distinctive features, just as in Rosch's theory (Rosch & Mervis, 1975) family resemblance was seen to be highest for those instances with the most attributes in common and lowest for items with few common features. Typical words are thus generally considered to be very similar to one another, whereas atypical words seem more dissimilar.

Context plays a large part in Tversky's model. Not only does it determine which features are selected, from the total feature domain, to be brought into the matching process, it also influences directly the value on scale "f" of the item pairs whose similarity is being estimated (the task the model was specifically designed to explain). According to Tversky, depending on task context, so a pair of items may be considered similar or dissimilar, reflecting the sampling procedure carried out before task performance. Tversky proposed what he termed a "focusing hypothesis" to explain the effects of context. According to this hypothesis, in tasks requiring judgements of similarity, subjects will attend mainly to the common features among items; whereas in tasks involving judgements of dissimilarity mainly distinctive features will be sampled. The focusing process just described affects the "relative salience" of the features used to perform the task. The relative salience of a feature is related to its diagnosticity, or classificatory significance. Thus a feature may have diagnostic value in one context if it serves as a basis for classification in that particular context; in another task, it may not even be sampled, if it does not have diagnostic value. An example will help demonstrate these context-determined effects. East and West Germany will be judged to be very similar countries in the context of "countries with a common language", but very dissimilar countries in the context of "countries with a common style of government". In the former case, the commonality of features referring to language is emphasized; in the latter case, the fact that both countries share the same language may not even be considered. The differences between the two countries' styles of government are more important. Similarly, atypical words are very dissimilar in the context of typicality judgements, since these concentrate on judgements of the "family resemblance", or similarity, between items; yet in another context, for instance, where subjects have to discriminate among different natural categories, those features which make atypical words belong in the same category may be emphasized and atypical words may be considered more similar than in the previous task.

Tversky's model thus captures Rosch's distinction between typical and atypical words (in the context of typicality judgements, typicality reflects differences in similarity among category members) and also includes a process (the focusing hypothesis) whereby context may determine an item's perceived similarity to other items. The following section now outlines one possible way in which Tversky's contrast model (a structural model) and the Cue-Overload Principle (a processing account) may be seen to complement each other, through a proposed "Similarity/Accessibility Hypothesis". In addition, it will be shown that the addition of this hypothesis to the original Cue-Overload formulation greatly improves the Principle's ability to account for the experimental work reported.

## 5.5 The Similarity/Accessibility Hypothesis

This section describes one way in which Cue-Overload may be extended, through a proposed Similarity/Accessibility Hypothesis based on Tversky's model of similarity. Simply put, the Similarity/Accessibility Hypothesis suggests that, for a given degree of cue-overload, the more similar the to-be-retrieved items are, as a group, the more accessible they will be to recall. Like Tversky's model, the Hypothesis allows for the degree of similarity among items to vary depending on task context. Thus words that are very dissimilar when viewed in connection with a particular set of features, may be more similar when viewed from another angle. In the context of "flightless birds", for instance, "penguin", "hen" and "ostrich" may be seen as very similar (and typical category instances), but in the context of "birds" they become dissimilar (and atypical).

Thus the Hypothesis attempts to stipulate more precisely the relationship between cue-effectiveness and recall, by proposing that this is determined not only by the cue's degree of overload, but also by the kind of item it is being used to retrieve: the more similar, as a group, the items loading a particular cue, the more accessible they are to retrieval by that cue. Intuitively, this is an appealing suggestion. If a single cue is used for retrieval, then the fewer the differences among the items being retrieved the more likely that a retrieval process using that cue will hit upon them -- they are "closer together" in the semantic space the cue is attempting to access.

One other aspect of the Similarity/Accessibility Hypothesis, which follows directly from the notion that inter-item similarity increases accessibility, is the suggestion that differences in accessibility brought about by similarity cannot exist independently of differences in discriminability. A series of papers by Craik and Jacoby (for instance, 1979; Jacoby & Craik, 1979; Lockhart, Craik & Jacoby, 1976) argues that the more similar items are within a group, the less distinctive each individual item becomes, relative to its associated items, and hence the more difficult it will be to discriminate between very similar target words and distractors within the same retrieval group. Thus the Hypothesis proposes that increased accessibility due to inter-item similarity may be bought at the expense of reduced inter-item discriminability.

The interrelation between accessibility and discriminability may itself refer to differences in the types of attributes used to determine inter-item similarity. Tversky suggested that both common and distinctive features are involved in judgements of similarity, so that the perceived similarity of a pair of items is the weighted measure of the two types of feature. Underwood (1969) proposed a similar kind of distinction. According to him, two types of attribute may be involved in retention tests: discriminative attributes and retrieval attributes. It could be that common features or attributes, or the information about the commonality of certain features, may determine accessibility; whereas distinctive features may determine discriminability.

It is important to note, however, that the Similarity/Accessibility Hypothesis does not attempt to specify the exact mechanisms involved in the retrieval of similar or dissimilar items. In its present form, its generality matches that of the Cue-Overload formulation, which also does not provide any fine detail as to the processes involved in cue-loading or retrieval. Its value may lie in the constraints it imposes on the generality of the original formulation of the Cue-Overload Principle. As the following section will now illustrate, used together with the Similarity/Accessibility Hypothesis, Cue-Overload is much more compatible with the data reported in this thesis, and is able to account for it in a much more straightforward manner, than it was when used on its own. But clearly, since this "test" of the Hypothesis is completely post hoc, much more research will still need to be done before the power and usefulness of the extended version of Cue-Overload can be determined.

## 5.6 Cue-Overload Principle, the Similarity/Accessibility Hypothesis, and the experimental results reported.

The three experiments reported in Chapter Four demonstrated that the Cue-Overload Principle breaks down when it is asked to predict the effects of typicality on recall (the typicality effect) and the effects of context on recall (the reversal of the typicality effect, for identical sets of words, with changes in experimental context).

Though it could not predict the typicality effect, Cue-Overload could explain it by assuming that different (typicality) cues were used for retrieval of typical and atypical words. Because the Principle does not specify the relationship between degree of overload and cue-effectiveness, for equally-loaded cues, it can then accommodate the finding of recall differences for both types of category instance. The use of different recall cues for typical and atypical items seemed plausible in the context of Experiment 5, which also demonstrated that typicality could be used as a functional retrieval cue: release from retroactive inhibition was obtained with shifts in list-word typicality. But in Experiment 6 typicality was a between-subjects factor, and subjects were given only the names of the general categories from which typical and atypical words had been drawn. In this context, the argument of different cues for recall is much less plausible.

The addition of the Similarity/Accessibility Hypothesis greatly improves the Principle's ability to account for the data. In this form, Cue-Overload can explain the typical-word recall advantage in terms of the greater inter-item similarity among typical words which, by definition, share many more attributes with one another and with the category prototype than atypical words do. Greater inter-item similarity would make typical words more accessible to recall by an equally-loaded cue, whether it be a typicality cue (Experiment 5) or a category cue (Experiment 6). The composite version of Cue-Overload thus obviates the need to postulate different cues for typical and atypical word recall in this latter experiment.

The extended version of the Principle then has to explain why the effects of experimental manipulation were equivalent across typicality level, in both these experiments. It could be argued that, if atypical words are so dissimilar that their accessibility is impaired (compared with typical words), then the effects of increasing cue overload should be even more marked for them than for the readily accessible typical items. Thus both the SAME condition of Experiment 5 and the CUED condition of Experiment 6 should have been associated with a greater drop in performance for atypical than for typical lists. However, as the Hypothesis also states, it could be that the very dissimilarity which renders atypical words difficult to retrieve, also makes it easier for subjects to discriminate between targets and intrusions and slightly improves their recall in these conditions. Typical target words, because they are so similar to one another, may be less easily discriminated from typical intrusions (also very similar), slightly lowering recall in the interference conditions. The apparent independence of the effects of typicality and experimental manipulation could hence be due to the antagonistic effects of discriminability at each typicality level. Had the intrusion data been more abundant, and evenly spread across subjects, it might have been possible to look there for independent evidence of the differential effects of discriminability on recall. However, they were not and, in addition, subjects may not have written down all the words they retrieved but which they were not sure were target items or intrusions.

Experiment 6 also showed that the typicality effect obtained only for superordinate-level categories, and not for basic-level categories. This was an interesting finding, in that it provided further converging evidence for the distinction proposed by Rosch, and again it could not be predicted by Cue-Overload. To explain the equivalent recall probability for typical and atypical words at the basic-level, the Principle could either claim it was a lucky coincidence (two different, equally-loaded cues producing the same level of recall), or that it reflected a change in recall strategy for basic-level category recall: At the basic level subjects used the general category cue for retrieval, and hence obtained the same level of recall for the same level of cue loading; and at the superordinate level they did not, and used instead two different cues. As was discussed earlier, the Principle can accommodate different recall levels for the same cue loading provided they are associated with different cues.

Addition of the Similarity/Accessibility Hypothesis produces a more straightforward interpretation of these data. At both levels, the same general category cue was used but, because by definition basic-level categories have a higher degree of internal similarity (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976), and because two of their internal sources of variation -- familiarity and associative frequency -- had been equated across typicality in this experiment, the degree of similarity among typical list words may have matched that for atypical words. This would have rendered basic-level atypical instances more accessible to recall than atypical words from superordinate categories. (Note that this need not mean that typical and atypical words are more similar to one another at the basic-level of abstraction. Rather, it means that for each group of typical and atypical words, independently, inter-item similarity may have been equally high. For example, typical birds may be similar to one another because they all fly, are comparatively small, etc.; whereas atypical birds may be similar to one another because they generally do not fly, tend to be big, etc. Sources of distinctive features within each typicality group may thus be fewer at the basic- than at the superordinate-level.)

Thus a combination of Cue-Overload and the Similarity/ Accessibility Hypothesis produces a more comprehensive and plausible interpretation of the effects of typicality on recall than the simple version of Cue-Overload was able to provide. It now remains to be determined whether the composite version of Cue-Overload can also explain the effects of context on recall, observed in the SAME conditions of Experiments 5 and 7.

In Experiments 5 and 7, exactly the same materials were used to make up lists for the interference (SAME) conditions. Yet the typicality effect obtained in Experiment 5 was reversed in Experiment 7. It was suggested that in the ATYPICAL-SAME condition atypical words had probably been retrieved via many more cues than had been used for retrieval in the remaining three conditions. Hence each of the several cues used to retrieve atypical words in the SAME condition was less overloaded than the category cue used in the remaining conditions, and was more effective for recall. This interpretation also helps to explain why no release from retroactive inhibition was obtained for atypical lists when category was shifted in the DIFFERENT condition and, in addition, renders the pattern of results obtained consistent with a Cue-Overload interpretation. However, it leaves unaddressed the question of why subjects chose to adopt such different recall strategies when faced with the task of recalling exactly the same words.

The Similarity/Accessibility Hypothesis does allow for the attributes used to determine the similarity between items to vary depending on the context in which words are encountered. However, at its present stage of development, the Hypothesis still cannot predict the effects of context on perceived similarity and, by implication, on recall. It could be that the shifts in category involved in Experiment 7 primed subjects to concentrate on the differences among words, so that they could classify them accurately into different categories. This in turn could have made them more aware of the category subsets to which the atypical words could be assigned equally well. These subset labels were then in turn used as retrieval cues in addition to, or instead of, the general category name. For instance, subjects could have reasoned that, as well as being names of items of furniture, some of the words in the atypical lists could also be classified as "garden furniture" (e.g., "garden-swing", "deck-chair", "trolley"), that some were names of "fittings" (e.g., "ashtray", "wall-mirror", etc.) and so on. In Experiment 5, on the other hand, the shifts in typicality, always within the same general category domain, could have primed the subjects to concentrate instead on the similarities among words within the same category, since these determine their typicality, according to Rosch's theory.

In Experiment 3, too, context seemed to determine which cues were used for retrieval. When atypical words outnumbered typical items, subjects adopted multiple cues for retrieval of atypical items. Where typical words outnumbered atypical ones, the category label was used instead. Clearly, more empirical work is necessary (for instance regarding the mechanisms for computing similarity in different contexts) before the Similarity/Accessibility Hypothesis can be used to make detailed predictions about experimental outcomes.

## 5.7 Implications of the Hypothesis for further research.

Sections 5.5 and 5.6 described one way in which Cue-Overload could be extended, through combination with a proposed Similarity/Accessibility Hypothesis, and evaluated the ability of this composite version of Cue-Overload to account for the data presented. It was argued that the Hypothesis greatly improved the Principle's ability to explain the effects of typicality on recall, and so that the results of Experiments 5 and 6 could now be fully accounted for in terms of Cue-Overload. The Hypothesis had more difficulty accommodating the results of Experiment 7. One possible account of the change in recall strategy observed in this experiment (compared with Experiment 5) was offered. However, more empirical work needs to be done before the Hypothesis can be used to predict the effects of context on choice of cues for retrieval.

Further research designed to test the predictive power of the composite model of Cue-Overload might concentrate on three separate areas. First, it will obviously be necessary to find some way to quantify similarity, so that inter-item similarity can itself be used to predict accessibility. Orthogonal comparisons of the effects of typicality and similarity on recall would be one useful way to test the interpretation of the results presented here.

Second, more information is needed concerning the relative contributions of accessibility (similarity) and discriminability (distinctiveness) to the patterns of results obtained. Using a recognition procedure could be a good way to start separating out the effects of these two components.

Finally, in order to test the Similarity/Accessibility Hypothesis fully, one must be able to define a priori the exact relationship between context and perceived similarity. How such a task might be achieved is not altogether clear at present. Systematically obtaining ratings of similarity for a variety of pairs, or sets, of words would seem a promising course of action and would, in addition, provide a way of quantifying similarity. But one would need to be careful to stipulate the context in which judgements were to be made. And even than translation of these similarity ratings to an experimental situation (another context) might itself pose interpretation problems.

However, until such controls are forthcoming, it is difficult to see how Cue-Overload can become more than the "loose framework for research" discussed in Chapter One of this thesis.

## 5.8 Conclusions

The experimental work reported in this thesis tested the generality of a Cue-Overload account of memory phenomena. At a broad level, the results support well such an interpretation. However, they also suggest that the Cue-Overload Principle is insufficient in some respects and an additional process needs to be postulated to fully explain the experimental data. To this end, a Similarity/ Accessibility Hypothesis was proposed. It was shown that in this extended form the Principle's ability to account for the data was greatly improved.

The thesis also provided some insight into the effects of typicality on recall, and demonstrated that a typical-word recall advantage could be obtained under circumstances in which it could not be attributed to either familiarity, associative frequency or encoding differences between typical and atypical category members.

Finally, the work reported here also included a comprehensive set of normative data for three measures of internal category structure, obtained for the same set of words, using similar subjects samples. The analysis of the intercorrelation of these three measures, based on the normative data collected, further extends our understanding of the nature and characteristics of typicality effects.

# FOOTNOTE

Experiment 4 has been reported in a different form in a paper entitled "Measures of internal category structure: A correlational analysis of normative data" published in the British Journal of Psychology (1983), 74, 491-516 and co-authored by Dr. James A. Hampton and myself.

I would also like to submit as subsidiary material the following two papers, co-authored by Drs. John M. Gardiner and Vernon H. Gregg, and myself:

Gardiner, J.M., Gardiner, M.M. & Gregg, V.H. (in press). The auditory recency advantage in longer-term free recall is not enhanced by recalling prerecency items first. Memory & Cognition.

Gardiner, J.M., Gregg, V.H. & Gardiner, M.M. Concerning some more evidence of an auditory advantage in prerecency as well as recency recall. (Manuscript submitted for publication in the American Journal of Psychology.)

Copies of these two papers are enclosed at the end of the thesis.

REFERENCES


Anderson, J.R. (1972) FRAN: A simulation model of free recall. In: G.H. Bower (Ed.) The psychology of learning and motivation, vol. 5, New York: Academic Press.

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

Anderson, J.R. & Bower, G.H. (1973) Human associative memory, New York: Wiley & Sons.

Anderson, J.R. & Bower, G.H. (1974) A propositional theory of recognition memory. Memory and Cognition, 2, 406-412.

Ashcraft, M.H. (1978a) Property norms for typical and atypical items from 17 categories: A description and discussion. Memory and Cognition, 6, 227-232.

Ashcraft, M.H. (1978b) Property dominance and typicality effects in property statement verification. Journal of Verbal Learning and Verbal Behavior, 17, 155-164.

Bahrick, H.P. (1969) Measurement of memory by prompted recall. Journal of Experimental Psychology, 79, 213-219.

Basden, D.R., Basden, B.H. & Galloway, B.C. (1977) Inhibition with part-list cueing: Some tests of the item strength hypothesis. Journal of Experimental Psychology: Human Learning and Memory, 3, 100-108.

Battig, W.F. & Montague, W.E. (1969) Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph 80 (3 pt. 2)

Bilodeau, E.A. (1967) Experimental interference with primary associates and their subsequent recovery with rest. Journal of Experimental Psychology, 73, 328-332.

Brown, J. (1968) Reciprocal facilitation and impairment in free recall. Psychonomic Science, 10, 41-42.

Brown, W.P. & Atkinson, J.T.O. (1974) Categories, subcategories and the attenuation of proactive inhibition in free recall. Memory and Cognition, 2, 127-129.

Cermak, L.S. & Craik, F.I.M. (1979) Levels of processing in human memory, New York: Erlbaum.

Collins, A.M. & Loftus, E.F. (1975) A spreading-activation theory of semantic processing. Psychological Review, 82, 407-428.

Craik, F.I.M. & Jacoby, L.L. (1979) Elaboration and distinctiveness in episodic memory. In: L.G. Nilssen (Ed.) Perspectives on memory research, Hillsdale, N.J. Erlbaum.

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

Dong, T. (1972) Cued partial recall of categorized words. Journal of Experimental Psychology, 93, 123-129.

Ebbinghaus, H. (1885) Memory: A contribution to experimental psychology. New York: Dover (1964 edition).

Estes, W.K. (1972) An associative basis for coding and organization in memory. In: A.W. Melton & E. Martin (Eds.) Coding processes in human memory. Washington, D.C.: Winston.

Gardiner, J.M., Craik, F.I.M. & Birtwistle, J. (1972) Retrieval cues and release from proactive inhibition. Journal of Verbal Learning and Verbal Behavior, 11, 778-783.

Gardiner, J.M., Klee, H., Redman, G. & Ball, M. (1976) The role of stimulus material in determining release from proactive inhibition. Quarterly Journal of Experimental Psychology, 28, 395-402.

Glass, A.L. & Meany, P.J. (1978) Evidence for two kinds of low-typical instances in a categorization task. Memory and Cognition, 6, 622-628.

Glenberg, A.M. (in press) A retrieval account of the long-term modality effect. Journal of Experimental Psychology: Learning, Memory and Cognition.

Greenberg, M.S. & Bjorklund, D.F. (1981) Category typicality in free recall: effects of feature overlap or differential category encoding? Journal of Experimental Psychology: Human Learning and Memory, 7, 145-147.

Hampton, J.A. (1976) An experimental study of concepts in language. Unpublished Ph.D. thesis. University of London.

Hampton, J.A. (1979) Polymorphous concepts in semantic memory. Journal of Verbal Learning and Verbal Behavior, 18, 441-461.

Hampton, J.A. (1981) An investigation of the nature of abstract concepts. Memory and Cognition, 9, 149-156.

Hudson, R.L. & Austin, J.B. (1970) Effect of context and category name on the recall of categorized word lists. Journal of Experimental Psychology, 86, 43-47.

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

Keller, D. & Kellas, G. (1978) Typicality as a dimension of encoding. Journal of Experimental Psychology: Human Learning and Memory, 4, 78-85.

Lewis, M.G. (1972) Cue effectiveness in cued recall. Paper presented at the annual meeting of the Psychonomic Society, St. Louis.

Lockhart, R.S., Craik, F.I.M. & Jacoby, L.L. (1976) Depth of processing, recognition and recall. In: J. Brown (Ed.) Recall and recognition, New York: Wiley.

Loess, H. (1967) Short-term memory, word class and sequence of items. Journal of Experimental Psychology, 74, 556-561.

Loftus, E.F. (1975) Spreading activation within semantic categories: Comments on Rosch's "Cognitive representations of semantic categories". Journal of Experimental Psychology: General, 104, 234-240.

Malt, B.C. & Smith, E.E. (1982) The role of familiarity in determining typicality. Memory and Cognition, 10, 69-75.

Mandler, G. (1967) Organization and memory. In: K.W. Spence & J.T. Spence (Eds.) The Psychology of learning and motivation: Advances in research and theory (vol. 1). New York: Academic Press.

Mandler, G. & Pearlstone, Z. (1966) Free and constrained concept learning and subsequent recall. Journal of Verbal Learning and Verbal Behavior, 5, 126-131.

Mayhew, A.J. (1967) Interlist changes in subjective organization during free-recall learning. Journal of Experimental Psychology, 74, 425-430.

McCloskey, M. (1980) The stimulus familiarity problem in semantic memory research. Journal of Verbal Learning and Verbal Behavior, 19, 485-502.

McCloskey, M. & Glucksberg, S. (1979) Decision processes in verifying category membership statements: Implications for models of semantic memory. Cognitive Psychology, 11, 1-37.

Mervis, C.B., Catlin, J. & Rosch, E. (1976) Relationships among goodness-of-example, category norms, and word frequency. Bulletin of the Psychonomic Society, 7, 283-284.

Mueller, C.W. & Watkins, M.J. (1977) Inhibition from part-set cueing: A cue-overload interpretation. Journal of Verbal Learning and Verbal Behavior, 16, 699-709.

Murdock, B.B. (1960) The immediate retention of unrelated words. Journal of Experimental Psychology, 60, 222-234.

Parkin, A.J. (1980) Levels of processing and the cue overload principle. Quarterly Journal of Experimental Psychology, 32, 427-434.

Postman, L., Stark, K. & Fraser, J. (1968) Temporal changes in interference. Journal of Verbal Learning and Verbal Behavior, 7, 672-694.

Raaijmakers, J.G.W. & Shiffrin, R.M. (1980) SAM: A theory of probabilistic search of associative memory. In: G. Bower (Ed.) The psychology of learning and motivation, vol. 14, New York: Academic Press.

Raaijmakers, J.G.W. & Shiffrin, R.M. (1981) Search of associative memory. Psychological Review, 88, 93-134.

Rips, L.J., Shoben, E.J. & Smith, E.E. (1973) Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 12, 1-20.

Rips, L.J., Smith, E.E. & Shoben, E.J. (1975) Set-theoretic and network models reconsidered: A comment on Hollan's "Features and semantic memory". Psychological Review, 82, 156-157.

Roediger, H.L. III (1973) Inhibition in recall from cueing with recall targets. Journal of Verbal Learning and Verbal Behavior, 12, 644-657.

Roediger, H.L. III (1974) Inhibiting effects of recall. Memory and Cognition, 2, 261-269.

Roediger, H.L. III (1978) Recall as a self-limiting process. Memory and Cognition, 6, 54-63.

Roediger, H.L. III (1980) Memory metaphors in cognitive psychology. Memory and Cognition, 8, 231-246.

Roediger, H.L. III & Nealy, J.H. (1982) Retrieval blocks in episodic and semantic memory. Canadian Journal of Psychology, 36, 213-242.

Roediger, H.L. III, Stellon, C. & Tulving, E. (1977) Inhibition from part-list cues and rate of recall. Journal of Experimental Psychology: Human Learning and Memory, 3, 174-188.

Rosch, E. (1973) On the internal structure of perceptual and semantic categories. In: T.E. Moore (Ed.) Cognitive development and the acquisition of language. New York: Academic Press.

Rosch, E.H. (1974) Universals and cultural specifics in human categorization. In: R. Brislim, W, Lonner & S. Bochner (Eds.) Cross-cultural perspectives on learning. London: Sage Press.

Rosch, E. (1975a) Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.

Rosch, E. (1975b) Cognitive reference points. Cognitive Psychology, 1, 532-547.

Rosch, E. (1978) Principles of categorization. In: E. Rosch & B.B. Lloyd (Eds.) Cognition and categorization. Hillsdale, N.J.: Erlbaum.

Rosch, E. & Mervis, C.B. (1975) Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. & Boyes-Braem, P. (1976) Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Rosch, E., Simpson, C. & Miller, R.S. (1976) Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2, 491-502.

Rundus, D. (1973) Negative effects of using list items as recall cues. Journal of Verbal Learning and Verbal Behavior, 12, 43-50.

Shiffrin, R.W. (1970) Memory search. In: D.A. Norman (Ed.) Models of human memory. New York: Academic Press.

Slamecka, N.J. (1968) An examination of trace storage in free recall. Journal of Experimental Psychology, 76, 504-513.

Slamecka, N.J. (1969) Testing for associative storage in multitrial free recall. Journal of Experimental Psychology, 81, 557-560.

Slamecka, N.J. (1972) The question of associative growth in the learning of categorized materials. Journal of Verbal Learning and Verbal Behavior, 11, 324-332.

Smith, E.E., Shoben, E.J. & Rips, L.J. (1974) Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81, 214-241.

Todres, A.K. & Watkins, M.J. (1981) A part-set cueing effect in recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 7, 91-99.

Tulving, E. (1962) Subjective organization in free recall of "unrelated" words. Psychological Review, 69, 344-354.

Tulving, E. (1964) Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. Psychological Review, 71, 219-237.

Tulving, E. & Pearlstone, Z. (1966) Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 5, 381-391.

Tversky, A. (1977) Features of similarity. Psychological Review, 84, 327-352.

Tversky, A. & Gati, I. (1978) Studies of similarity. In: E. Rosch & B.B. Lloyd (Eds.) Cognition and categorization. Hillsdale, N.J.: Erlbaum.

Underwood, B.J. (1948) Retroactive and proactive inhibition after five and forty-eight hours. Journal of Experimental Psychology, 38, 29-38.

Underwood, B.J. (1969) Attributes of memory. Psychological Review, 76, 559-573.

Watkins, M.J. (1975) Inhibition in recall with extralist cues. Journal of Verbal Learning and Verbal Behavior, 14, 294-303.

Watkins, M.J. (1979) Engrams as cuegrams and forgetting as cue-overload: A cueing approach to the structure of memory. In: C.R. Puff (Ed.) Memory organization and structure. New York: Academic Press.

Watkins, M.J. & Watkins, O.C. (1976) Cue-Overload theory and the method of interpolated attributes. Bulletin of the Psychonomic Society, 7, 289-291.

Watkins, O.C. & Watkins, M.J. (1975) Build-up of proactive inhibition as a cue-overload effect. Journal of Experimental Psychology: Human Learning and Performance, 104, 442-452.

Wickelgren, W.A. (1976) Network strength theory of storage and retrieval dynamics. Psychological Review, 83, 466-478.

Wickens, D.D. (1970) Encoding categories of words: An empirical approach to meaning. Psychological Review, 77, 1-15.

Wickens, D.D. (1972) Characteristics of word encoding. In: A.W. Melton & E. Martin (Eds.) Coding processes in human memory. New York: Winston.

Wickens, D.D., Born, D.G. & Allen, C.K. (1963) Proactive inhibition and item similarity in short-term memory. Journal of Verbal Learning and Verbal Behavior, 2, 440-445.

## APPENDIX ONE - MATERIALS FOR EXPERIMENT 1

| Category | Words |
|---|---|
| Fruit | Plum, raspberry, melon, fig, banana, tangerine, peach, grape, cherry, grapefruit, lemon, apricot, strawberry, coconut, lime, pineapple, pomegranate, avocado. |
| Insects | Moth, cockroach, gnat, termite, centipede, dragonfly, mantis, wasp, flea, butterfly, caterpillar, worm, hornet, cricket, tick, silverfish, lice, locust. |
| Animals | Lion, wolf, goat, mule, pig, deer, cow, bear, rabbit, sheep, donkey, tiger, zebra, monkey, fox, lamb, bull, elephant. |
| Flowers | Violet, marigold, daffodil, buttercup, carnation, orchid, lily, dandelion, iris, geranium, peony, azalea, hyacinth, lilac, gladioli, gardenia, aster, magnolia. |
| Food flavours | Nutmeg, cloves, parsley, vinegar, mustard, sage, sugar, garlic, paprika, mint, chives, oregano, cinnamon, sesame, rosemary, thyme, curry, ginger. |
| Birds | Hawk, crow, flamingo, parrot, seagull, vulture, starling, woodpecker, owl, thrush, falcon, pheasant, peacock, raven, swan, swallow, wren, pigeon. |

| | |
|---|---|
| Vegetables | Beans, spinach, celery, peppers, onion, turnip, corn, lettuce, cabbage, beetroot, asparagus, sprouts, peas, radishes, cauliflower, kale, broccoli, cucumber. |
| Trees | Birch, spruce, willow, cedar, mimosa, poplar, redwood, larch, fir, ash, sycamore, yew, palm, beech, holly, cypress, elder, maple. |
| Sports | Golf, hockey, soccer, badminton, volleyball, fencing, ping-pong, boxing, sailing, squash, skiing, racing. |
| Countries | Spain, France, Japan, Israel, Greece, Portugal, China, Argentina, Italy, Austria, Canada, Ireland. |
| Musical instruments | Trombone, harp, organ, clarinet, bugle, oboe, cornet, violin, harmonica, trumpet, fiddle, saxophone. |
| Occupations | Bricklayer, mechanic, clerk, electrician, salesman, policeman, nurse, architect, fireman, secretary, plumber, engineer. |

## APPENDIX TWO - MATERIALS FOR EXPERIMENT 2

| | (a) | | (b) |
|---|---|---|---|
| Fish | Freshwater | | Saltwater |
| | Salmon | | Tuna |
| | Pike | | Plaice |
| | Minnow | | Mackerel |
| | Perch | | Sole |
| | Stickleback | | Haddock |
| | Trout | | Sardine |
| Trees | Evergreen | | Deciduous |
| | Cedar | | Ash |
| | Holly | | Sycamore |
| | Spruce | | Willow |
| | Laurel | | Oak |
| | Yew | | Elm |
| | Pine | | Beech |
| Flowers | Garden | | Wild |
| | Marigold | | Daisy |
| | Carnation | | Bluebell |
| | Geranium | | Poppy |
| | Rose | | Cowslip |
| | Dahlia | | Buttercup |
| | Tulip | | Dandelion |
| Vegetables | Green | | Root |
| | Spinach | | Carrot |
| | Sprouts | | Swede |
| | Cabbage | | Radish |
| | Broccoli | | Parsnip |
| | Lettuce | | Potato |
| | Kale | | Turnip |

| Birds | (a) Edible | (b) Predatory |
|---|---|---|
| | Grouse | Falcon |
| | Quail | Owl |
| | Phesant | Condor |
| | Woodcock | Hawk |
| | Partridge | Buzzard |
| | Pigeon | Eagle |

| Sports | (a) Indoor | (b) Outdoor |
|---|---|---|
| | Badminton | Golf |
| | Fencing | Angling |
| | Boxing | Cricket |
| | Ping-pong | Tennis |
| | Squash | Soccer |
| | Judo | Netball |

APPENDIX THREE - MATERIALS FOR EXPERIMENT 3

Category : Birds

Typical words : Robin, canary, dove, sparrow, wren, starling, crow, thrush, seagull, eagle, pigeon.

Atypical words: Pelican, ostrich, stork, albatross, swan, turkey, vulture, flamingo, peacock, emu, penguin.

Category : Sports

Typical words : Tennis, rugby, swimming, hockey, badminton, boxing, golf, lacrosse, fencing, wrestling, volleyball.

Atypical words: Archery, bowls, hunting, sailing, hiking, riding, ping-pong, snooker, diving, judo, fishing.

Category : Vehicles

Typical words : Aeroplane, train, lorry, car, bus, tram, motorbike, tractor, van, bicycle, taxi.

Atypical words: Canoe, balloon, wheelchair, pram, cablecar, raft, lift, rickshaw, trailer, helicopter, cart.

Category : Vegetables

Typical words : Carrots, spinach, broccoli, peas, cauliflower, onions, cabbage, asparagus, sprouts, turnip, potatoes.

Atypical words: Artichoke, garlic, peppers, celery, beetroot, kale, watercress, yam, cucumber, endive, mushroom.

Category : Fruit

Typical words : Apple, strawberry, banana, orange, peach, grapes, cherries, pear, tangerine, plum, melon.

Atypical words: Tomato, olive, date, lemon, fig, cranberry, mango, coconut, guava, avocado, pumpkin.

Category : Clothing

Typical words : Shirt, dress, jacket, skirt, blouse, suit, pyjamas, slacks, socks, sweater, coat.

Atypical words: Girdle, apron, mittens, tuxedo, cape, tie, waistcoat, scarf, bolero, gloves, mac.

Category : Weapons

Pistol, dagger, spear, rifle, sword, bomb, missile, bayonet.

Category : Insects

Spider, beetle, locust, flea, butterfly, centipede, mosquito, lice.

APPENDIX FOUR - SUMMARY OF THE NORMATIVE DATA COLLECTED IN EXPERIMENT 4. (TYP = TYPICALITY; FAM = FAMILIARITY; A.F. = ASSOCIATIVE FREQUENCY; N.R. = NUMBER OF CATEGORY REJECTIONS; N.U.= NUMBER OF UNKNOWN RESPONSES)

Category : Birds

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Blackbird | 1.000 | 1.097 | 45( 9) | - | - |
| Sparrow | 1.047 | 1.032 | 48(14) | - | - |
| Robin | 1.093 | 1.129 | 45( 6) | - | - |
| Starling | 1.182 | 1.484 | 23( 2) | - | - |
| Thrush | 1.186 | 1.387 | 21( 2) | - | 1 |
| Pigeon | 1.250 | 1.097 | 26( 3) | - | - |
| Crow | 1.256 | 1.323 | 20( 2) | - | - |
| Seagull | 1.364 | 1.226 | 25( 2) | - | - |
| Swallow | 1.419 | 1.581 | 29( 7) | - | - |
| Wren | 1.465 | 1.613 | 11 - | - | - |
| Dove | 1.477 | 1.548 | 8( 1) | - | - |
| Cuckoo | 1.535 | 1.290 | 6 - | - | - |
| Hawk | 1.698 | 1.613 | 25( 1) | - | - |
| Woodpecker | 1.727 | 1.452 | 9 - | - | - |
| Swift | 1.732 | 1.903 | 6 - | - | 2 |
| Raven | 1.744 | 1.645 | 7 - | - | - |
| Nightingale | 1.773 | 1.903 | 9( 1) | - | - |
| Owl | 1.773 | 1.161 | 22 - | - | - |
| Eagle | 1.791 | 1.355 | 45( 3) | - | - |
| Lark | 1.795 | 2.452 | 7 - | - | - |
| Parrot | 1.837 | 1.290 | 20 - | - | - |
| Pheasant | 1.930 | 1.516 | 4 - | - | - |
| Canary | 1.953 | 1.258 | 14 - | - | - |
| Budgerigar | 1.977 | 1.419 | 19( 1) | - | - |
| Swan | 2.000 | 1.194 | 19( 3) | - | - |
| Chicken | 2.070 | 1.097 | 8 - | - | - |

| Word | TYP | FAM | A.F. | | N.R. | N.U. |
|---|---|---|---|---|---|---|
| Duck | 2.159 | 1.194 | 21 | - | - | - |
| Hen | 2.182 | 1.129 | 7( | 1) | - | - |
| Falcon | 2.182 | 2.161 | 5 | - | - | - |
| Albatross | 2.205 | 2.097 | 4 | - | - | - |
| Vulture | 2.295 | 1.742 | 10 | - | - | - |
| Peacock | 2.295 | 1.516 | 2 | - | - | - |
| Goose | 2.302 | 1.452 | 10 | - | - | - |
| Turkey | 2.302 | 1.258 | 3 | - | - | - |
| Warbler | 2.310 | 3.032 | 3 | - | 1 | 2 |
| Osprey | 2.326 | 2.290 | 6( | 1) | - | 1 |
| Heron | 2.326 | 2.000 | 10 | - | - | 1 |
| Grouse | 2.372 | 2.258 | - | - | - | - |
| Stork | 2.476 | 1.742 | 3 | - | - | 1 |
| Buzzard | 2.477 | 2.484 | 2 | - | - | - |
| Cockatoo | 2.548 | 2.194 | 2 | - | - | 2 |
| Flamingo | 2.651 | 1.806 | 8( | 1) | - | - |
| Tern | 2.714 | 3.677 | 1 | - | 2 | 5 |
| Pelican | 2.721 | 1.613 | 8 | - | - | - |
| Puffin | 2.905 | 1.968 | 4 | - | - | 1 |
| Woodcock | 2.905 | 3.161 | - | - | 1 | 2 |
| Quail | 2.977 | 2.968 | - | - | 1 | 1 |
| Condor | 3.023 | 3.516 | 1 | - | 2 | 5 |
| Ostrich | 3.047 | 1.742 | 12 | - | - | 1 |
| Toucan | 3.143 | 3.161 | 3 | - | 2 | 3 |
| Penguin | 3.227 | 1.323 | 9 | - | 1 | - |
| Emu | 3.512 | 1.839 | 4 | - | 2 | 1 |

Blue tit (34;1), chaffinch (12;0), finch (10;1), magpie (10;1), kingfisher (10;1), house martin (8;0), jay (8;0), kestrel (8;0), rook (8;0), greenfinch (5;0), jackdaw (5;0), crane (4;1), parakeet (4;0).

Category : Clothing

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Dress | 1.000 | 1.100 | 41( 7) | - | - |
| Skirt | 1.022 | 1.233 | 49( 3) | - | - |
| Trousers | 1.022 | 1.000 | 55( 2) | - | - |
| Shirt | 1.044 | 1.033 | 57( 6) | - | - |
| Jeans | 1.067 | 1.000 | 17( 4) | - | - |
| Jumper | 1.178 | 1.133 | 44( 6) | - | - |
| Jacket | 1.244 | 1.100 | 38( 3) | - | - |
| Suit | 1.267 | 1.167 | 4( 1) | - | - |
| Blouse | 1.289 | 1.133 | 36( 2) | - | - |
| Coat | 1.289 | 1.067 | 45( 2) | - | - |
| Cardigan | 1.422 | 1.500 | 24( 1) | - | - |
| Overcoat | 1.467 | 1.500 | 6 - | - | - |
| Socks | 1.600 | 1.067 | 60( 3) | - | - |
| Brassiere | 1.756 | 1.433 | - - | - | - |
| Slacks | 1.778 | 2.100 | 1 - | - | - |
| Anorak | 1.822 | 1.300 | 5 - | - | - |
| Pants | 1.822 | 1.233 | 21( 2) | - | - |
| Dungarees | 1.844 | 1.833 | 7 - | - | - |
| Tights | 1.955 | 1.367 | 26 - | - | 1 |
| Vest | 1.956 | 1.400 | 25 - | - | - |
| Shorts | 2.000 | 1.233 | 20 - | - | - |
| Stocking | 2.044 | 1.367 | 18 - | - | - |
| Parka | 2.070 | 1.900 | - - | - | 2 |
| Mackintosh | 2.136 | 2.100 | 11 - | - | 1 |
| Pyjamas | 2.205 | 1.300 | 2 - | - | 1 |
| Waistcoat | 2.333 | 1.600 | 15 - | - | - |
| Bikini | 2.444 | 1.433 | 2 - | - | - |
| Pinafore | 2.444 | 2.167 | 1 - | - | - |
| Smock | 2.523 | 2.767 | 1 - | - | 1 |
| Sari | 2.545 | 2.767 | - - | - | 2 |
| Scarf | 2.644 | 1.333 | 24 - | 1 | - |
| Overalls | 2.667 | 1.767 | 1 - | 1 | - |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Shawl | 2.733 | 1.800 | 2 – | – | – |
| Swimsuit | 2.756 | 1.367 | – – | 1 | – |
| Tie | 2.800 | 1.167 | 32 – | 1 | – |
| Tunic | 2.800 | 2.133 | – – | 1 | – |
| Gloves | 2.844 | 1.333 | 19 – | – | – |
| Hat | 2.844 | 1.467 | 37(14) | 1 | – |
| Bathrobe | 2.867 | 1.533 | 1 – | 1 | – |
| Romper suit | 2.933 | 2.833 | – – | – | – |
| Sandals | 3.067 | 1.367 | 5 – | 4 | – |
| Belt | 3.133 | 1.300 | 7 – | 5 | – |
| Mittens | 3.133 | 2.000 | 1 – | – | – |
| Cravat | 3.289 | 2.300 | 1 – | 1 | – |
| Slippers | 3.289 | 1.367 | 1 – | 4 | – |
| Beret | 3.333 | 1.733 | – – | 2 | – |
| Bow tie | 3.341 | 1.633 | – – | 1 | 1 |
| Corset | 3.356 | 2.300 | 2 – | – | – |
| Girdle | 3.356 | 2.267 | 1 – | – | – |
| Apron | 3.511 | 1.567 | – – | 5 | – |
| School cap | 3.600 | 1.767 | – – | 1 | – |
| Cassock | 3.614 | 3.667 | – – | 3 | 2 |
| Bolero | 3.644 | 3.967 | 1 – | 2 | 8 |
| Cricket cap | 3.978 | 2.067 | – – | 1 | – |
| Turban | 4.067 | 2.433 | – – | 3 | – |

Shoes (45;6), bra (22;2), t-shirt (22;0), boots (15;0), knickers (12;0), pullover (11;2), petticoat (11;0), sweater (6;1), leg warmers (6;0), underpants (6;0), underwear (6;0), raincoat (4;0), slip (4;0).

Category : Fish

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Cod | 1.040 | 1.129 | 53(14) | - | - |
| Trout | 1.120 | 1.290 | 43(11) | - | - |
| Salmon | 1.120 | 1.161 | 47( 5) | - | - |
| Herring | 1.140 | 1.613 | 32( 3) | - | 1 |
| Mackerel | 1.160 | 1.516 | 32( 1) | - | 1 |
| Plaice | 1.260 | 1.355 | 42( 2) | - | - |
| Haddock | 1.260 | 1.355 | 31( 2) | - | - |
| Sole | 1.640 | 1.645 | 13( 1) | - | - |
| Whiting | 1.688 | 2.452 | 8 - | 1 | 3 |
| Halibut | 1.688 | 2.032 | 15( 1) | - | 3 |
| Tuna | 1.700 | 1.484 | 14( 1) | - | - |
| Sardine | 1.720 | 1.355 | 18 - | - | - |
| Pike | 1.735 | 2.097 | 21( 5) | - | 1 |
| Bass | 1.771 | 2.258 | 5 - | 1 | 3 |
| Pilchard | 1.816 | 1.548 | 9 - | - | 2 |
| Carp | 1.920 | 2.290 | 10( 2) | - | 2 |
| Perch | 1.980 | 2.452 | 14 - | - | 2 |
| Whitefish | 2.020 | 3.000 | - - | - | 4 |
| Bream | 2.222 | 3.032 | 12( 2) | - | 8 |
| Stickleback | 2.306 | 2.097 | 12 - | - | 2 |
| Minnow | 2.320 | 2.387 | 5 - | - | - |
| Piranha | 2.340 | 1.677 | 6 - | - | - |
| Flounder | 2.400 | 3.452 | 1 - | - | 12 |
| Mullet | 2.413 | 3.129 | 3 - | - | 7 |
| Tench | 2.477 | 3.097 | 4( 1) | - | 10 |
| Swordfish | 2.560 | 1.677 | 10( 1) | - | - |
| Shark | 2.580 | 1.161 | 34( 3) | 2 | - |
| Chub | 2.591 | 3.129 | 2 - | 1 | 10 |
| Sturgeon | 2.694 | 3.290 | 2 - | 2 | 5 |
| Guppy | 2.696 | 3.323 | 6 - | - | 9 |
| Anchovy | 2.878 | 2.355 | 1 - | 5 | 3 |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Barracuda | 2.915 | 2.710 | 1 – | 1 | 5 |
| Ray | 3.020 | 2.548 | 1 – | 1 | 2 |
| Eel | 3.240 | 1.742 | 17( 1) | 3 | – |
| Turbot | 3.583 | 4.484 | – – | 4 | 28 |
| Lamprey | 3.591 | 4.258 | – – | 5 | 16 |
| Shad | 3.917 | 5.129 | – – | 4 | 29 |

Goldfish (26;5), roach (14;2), dogfish (14;1), catfish (12;0), angel-fish (11;0), skate (9;0), prawn (8;0), whale (7;1), shrimp (7;0), whitebait (6;1), hake (5;1), crab (5;0), dace (5;0), dolphin (5;0), kipper (5;0), rock-fish (5;0), coley (4;0), jellyfish (4;0), lobster (4;0), sprat (4;0).

Category : Flowers

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Rose | 1.040 | 1.032 | 64(22) | – | – |
| Daffodil | 1.100 | 1.194 | 48( 9) | – | – |
| Carnation | 1.120 | 1.129 | 20( 1) | – | – |
| Tulip | 1.140 | 1.194 | 39( 3) | – | – |
| Daisy | 1.180 | 1.194 | 46(12) | – | – |
| Buttercup | 1.240 | 1.161 | 32( 1) | – | – |
| Chrysanthemum | 1.260 | 1.516 | 26 – | – | – |
| Pansy | 1.306 | 1.645 | 20( 5) | – | 1 |
| Primrose | 1.340 | 1.645 | 11( 4) | – | – |
| Snowdrop | 1.340 | 1.419 | 13 – | – | – |
| Poppy | 1.380 | 1.387 | 11( 1) | 1 | – |
| Marigold | 1.400 | 1.645 | 12 – | – | – |
| Violet | 1.420 | 1.774 | 14 – | – | – |
| Bluebell | 1.460 | 1.581 | 15( 1) | – | – |

| Word | TYP | FAM | A.F. | | N.R. | N.U. |
|---|---|---|---|---|---|---|
| Crocus | 1.460 | 1.581 | 12 | - | - | - |
| Orchid | 1.531 | 1.452 | 11 | - | - | 1 |
| Geranium | 1.560 | 1.742 | 10 | - | - | 1 |
| Dahlia | 1.560 | 1.968 | 16( | 1) | - | - |
| Iris | 1.680 | 2.194 | 14( | 1) | - | - |
| Lily | 1.700 | 1.935 | 16 | - | - | - |
| Gladioli | 1.735 | 2.484 | 7( | 1) | - | 2 |
| Hyacinth | 1.776 | 2.129 | 9( | 9) | - | 2 |
| Narcissus | 1.854 | 2.774 | 6 | - | 1 | 2 |
| Petunia | 1.854 | 2.774 | 6 | - | - | 3 |
| Azalea | 1.959 | 3.065 | 3( | 1) | 1 | 4 |
| Rhododendron | 1.980 | 1.645 | 7 | - | 1 | 1 |
| Begonia | 2.063 | 2.710 | 6( | 1) | - | 3 |
| Freesia | 2.091 | 2.871 | 5 | - | - | 8 |
| Magnolia | 2.140 | 2.774 | - | - | 1 | - |
| Nasturtium | 2.239 | 3.484 | 6 | - | 1 | 10 |
| Lilac | 2.280 | 1.903 | 3 | - | 1 | - |
| Sweetpea | 2.300 | 1.581 | 5 | - | 2 | - |
| Anemone | 2.319 | 2.645 | 5( | 1) | 2 | 7 |
| Cowslip | 2.340 | 2.290 | 4 | - | - | - |
| Lavender | 2.340 | 1.903 | 4 | - | - | - |
| Dandelion | 2.360 | 1.226 | 17 | - | 3 | - |
| Peony | 2.364 | 3.935 | 2 | - | 1 | 18 |
| Waterlily | 2.500 | 1.710 | 1 | - | 2 | - |
| Aster | 2.500 | 3.581 | 2 | - | 1 | 15 |
| Gardenia | 2.522 | 3.968 | 1 | - | - | 7 |
| Jasmine | 2.532 | 2.871 | - | - | 2 | 3 |
| Camellia | 2.605 | 3.774 | 1 | - | - | 12 |
| Lotus | 2.813 | 2.645 | - | - | 2 | 3 |
| Jonquil | 3.139 | 5.419 | - | - | 2 | 35 |
| Zinnia | 3.229 | 5.419 | - | - | 2 | 37 |
| Phlox | 3.270 | 5.226 | - | - | 2 | 33 |

Sunflower (12;2), hydrangea (8;1), fuchsia (6;0), lily of the valley (6;0), cornflower (5;1), forget-me-not (5;0), honeysuckle (5;0), wallflower (5;0), foxglove (4;0).

Category : Food flavourings

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Garlic | 1.186 | 1.516 | 27 - | - | - |
| Salt | 1.233 | 1.000 | 51(13) | - | - |
| Pepper | 1.256 | 1.065 | 44( 3) | - | - |
| Sugar | 1.558 | 1.000 | 17( 4) | 1 | - |
| Ginger | 1.674 | 1.419 | 10 - | - | - |
| Mustard | 1.814 | 1.258 | 12 - | - | - |
| Vanilla | 1.814 | 1.226 | 25(13) | - | - |
| Cinnamon | 1.860 | 2.000 | 18( 3) | - | - |
| Allspice | 1.878 | 3.323 | 3 - | 1 | 9 |
| Sage | 1.884 | 2.000 | 19 - | - | - |
| Mint | 1.884 | 1.258 | 4( 1) | - | - |
| Nutmeg | 1.953 | 2.032 | 11 - | - | - |
| Curry | 2.000 | 1.387 | 21( 4) | 3 | - |
| Thyme | 2.047 | 2.290 | 27 - | - | - |
| Cloves | 2.070 | 2.129 | 8 - | - | - |
| Vinegar | 2.093 | 1.065 | 13( 1) | - | - |
| Rosemary | 2.233 | 2.484 | 17 - | - | - |
| Paprika | 2.238 | 2.968 | 16( 2) | 1 | 4 |
| Bayleaf | 2.302 | 2.355 | 4 - | - | - |
| Chives | 2.302 | 2.065 | 2 - | - | - |
| Basil | 2.333 | 2.774 | 7 - | - | 2 |
| Peppercorn | 2.415 | 2.161 | - - | 1 | 4 |
| Peppermint | 2.442 | 1.226 | 5( 3) | 2 | - |
| Cayenne | 2.452 | 3.452 | 2 - | 1 | 8 |

| Word | TYP | FAM | A.F. | | N.R. | N.U. |
|---|---|---|---|---|---|---|
| Saccharin | 2.535 | 1.806 | - | - | 5 | - |
| Oregano | 2.576 | 3.710 | 10 | - | 2 | 15 |
| Cocoa | 2.674 | 1.290 | 2 | - | 3 | - |
| Tarragon | 2.744 | 3.581 | 5 | - | 2 | 13 |
| Turmeric | 2.763 | 4.452 | 3 | - | 1 | 20 |
| Chocolate | 2.791 | 1.032 | 9( | 1) | 3 | - |
| Pickle | 2.837 | 1.258 | - | - | 3 | - |
| Dill | 2.846 | 3.097 | 5 | - | 2 | 7 |
| Cardamon | 2.966 | 5.387 | 1 | - | 1 | 35 |
| Cumin | 2.972 | 5.161 | 1 | - | 5 | 27 |
| Marjoram | 3.024 | 3.710 | 5 | - | 2 | 11 |
| Sesame | 3.171 | 3.387 | 1 | - | 3 | 7 |
| Mayonnaise | 3.233 | 1.290 | - | - | 4 | - |
| Borage | 3.607 | 5.484 | - | - | 2 | 35 |
| Chervil | 3.645 | 5.290 | - | - | 3 | 30 |
| Oil | 4.326 | 1.226 | 1 | - | 13 | - |

Spices (20;2), herbs (20;1), chilli (13;1), lemon (13;0), tomato puree (9;1), coffee (9;0), orange (8;2), almond (7;1), rum (6;2), sauces (6;1) tomato sauce (6;1), Oxo (5;1), beef stock (5;0), monosodium glutamate (4;2), black pepper (4;1), Bovril (4;0), cochineal (4;0), essences (4;0), peppers (4;0), strawberry (4;0), wine (4;0).

Category : Fruit

| Word | TYP | FAM | A.F. | | N.R. | N.U. |
|---|---|---|---|---|---|---|
| Apple | 1.023 | 1.063 | 69(30) | | - | - |
| Orange | 1.023 | 1.031 | 63(18) | | - | - |
| Pear | 1.163 | 1.188 | 61( 3) | | - | - |
| Banana | 1.233 | 1.125 | 53( 4) | | - | - |
| Grapefruit | 1.256 | 1.281 | 24 | - | - | - |
| Strawberry | 1.256 | 1.219 | 30 | - | - | - |
| Grape | 1.279 | 1.406 | 38 | - | - | - |
| Plum | 1.302 | 1.281 | 33( 1) | | - | - |
| Pineapple | 1.419 | 1.438 | 29( 1) | | - | - |
| Cherry | 1.419 | 1.469 | 19 | - | - | - |
| Peach | 1.419 | 1.469 | 39 | - | - | - |
| Lemon | 1.512 | 1.188 | 31 | - | - | - |
| Tangerine | 1.512 | 1.719 | 17 | - | - | - |
| Mandarin | 1.605 | 2.031 | 7 | - | - | - |
| Satsuma | 1.643 | 2.094 | 7 | - | - | 1 |
| Raspberry | 1.651 | 1.438 | 27 | - | - | - |
| Blackberry | 1.721 | 1.469 | 26 | - | - | - |
| Melon | 1.814 | 1.406 | 26( 3) | | - | - |
| Apricot | 1.814 | 1.656 | 10 | - | - | - |
| Blackcurrant | 1.881 | 1.594 | 13 | - | - | 1 |
| Gooseberry | 2.047 | 1.688 | 13 | - | - | - |
| Lime | 2.093 | 1.906 | 17 | - | - | - |
| Water-melon | 2.140 | 1.594 | 2 | - | - | - |
| Damson | 2.195 | 3.000 | 6 | - | - | 4 |
| Redcurrant | 2.429 | 2.438 | 4 | - | - | 3 |
| Nectarine | 2.615 | 3.125 | 16 | - | 2 | 7 |
| Avocado | 2.714 | 2.063 | 11 | - | - | 1 |
| Elderberry | 2.714 | 2.906 | 3 | - | 1 | 2 |
| Mango | 2.791 | 2.750 | 17( 4) | | - | 1 |
| Blueberry | 2.814 | 2.844 | 7 | - | - | - |
| Cranberry | 2.814 | 2.781 | - | - | - | - |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Pomegranate | 2.837 | 2.594 | 16 - | - | - |
| Fig | 2.837 | 2.031 | 7 - | - | - |
| Prune | 2.884 | 1.781 | 3( 1) | 2 | - |
| Date | 2.929 | 1.625 | 7 - | - | 1 |
| Raisin | 3.093 | 1.563 | 1 - | 2 | 1 |
| Greengage | 3.103 | 3.469 | 3 - | - | 8 |
| Guava | 3.485 | 4.469 | 2 - | 2 | 23 |
| Coconut | 3.581 | 1.688 | 3 - | 5 | - |
| Olive | 3.907 | 2.063 | 1 - | 6 | - |
| Pumpkin | 4.093 | 2.313 | - - | 7 | - |
| Almond | 4.721 | 1.719 | - - | 20 | - |
| Acorn | 5.023 | 2.000 | - - | 21 | 1 |

Passion-fruit (16;0), kiwi-fruit (6;0), loganberry (6;0), lichi (4;0), pawpaw (4;0).

<u>Category : Furniture</u>

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Chair | 1.000 | 1.065 | 66(45) | - | - |
| Armchair | 1.039 | 1.097 | 22( 3) | - | - |
| Table | 1.039 | 1.032 | 67(15) | - | - |
| Sofa | 1.098 | 1.323 | 22( 1) | - | - |
| Settee | 1.098 | 1.516 | 23 - | - | - |
| Bed | 1.176 | 1.032 | 50( 2) | - | - |
| Wardrobe | 1.216 | 1.258 | 40( 1) | - | - |
| Couch | 1.216 | 1.935 | 4 - | - | - |
| Suite | 1.471 | 1.903 | - - | - | - |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Dresser | 1.510 | 1.839 | 8 – | – | – |
| Desk | 1.529 | 1.323 | 31( 1) | – | – |
| Sideboard | 1.569 | 1.677 | 23 – | – | – |
| Cupboard | 1.647 | 1.258 | 26 – | 1 | – |
| Stool | 1.706 | 1.355 | 28( 1) | – | – |
| Cabinet | 1.765 | 1.774 | 17 – | – | – |
| Bookcase | 1.824 | 1.419 | 13 – | – | – |
| Bureau | 2.000 | 2.226 | 5 – | 1 | – |
| Cot | 2.118 | 2.065 | – – | – | – |
| Chest | 2.216 | 1.742 | 4 – | 1 | – |
| Bench | 2.235 | 1.484 | 3 – | – | – |
| Bunk | 2.392 | 1.677 | – – | 2 | – |
| Shelves | 2.627 | 1.355 | 10 – | 2 | – |
| Tallboy | 3.000 | 3.935 | 4 – | 2 | 9 |
| Sink unit | 3.588 | 1.645 | 4 – | 6 | – |
| Deckchair | 3.725 | 1.548 | 1 – | 5 | – |
| Wall mirror | 3.961 | 1.581 | – – | 10 | – |
| Bar | 4.039 | 1.452 | 1 – | 11 | – |
| Screen | 4.039 | 2.355 | – – | 5 | – |
| Bottle rack | 4.118 | 2.129 | – – | 7 | – |
| Pew | 4.294 | 2.452 | – – | 11 | – |
| Spice rack | 4.314 | 2.032 | – – | 9 | – |
| Trolley | 4.471 | 1.774 | 1 – | 15 | – |
| Waste paper basket | 4.471 | 1.290 | – – | 14 | – |
| Counter | 4.480 | 2.032 | – – | 13 | 1 |
| Hammock | 4.529 | 2.258 | – – | 15 | – |
| Painting | 4.804 | 1.677 | 1 – | 21 | – |
| Garden swing | 4.824 | 1.968 | – – | 19 | – |
| Park bench | 5.000 | 1.935 | – – | 25 | – |
| Ashtray | 5.137 | 1.452 | 1 – | 23 | – |
| Altar | 5.176 | 1.935 | – – | 27 | – |
| Library steps | 5.380 | 2.258 | – – | 31 | 1 |

Lamp (25;0), chest of drawers (24;0), dressing-table (20;0), carpet (17;0), coffee-table (13;0), television (9;0), bedside table (8;0), rug (7;0), pouffe (5;0), bath (4;0), cooker (4;0), curtains (4;0), drawers (4;0), fridge (4;0).

Category : Insects

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Fly | 1.116 | 1.031 | 44(15) | – | – |
| Ant | 1.116 | 1.031 | 49(22) | – | – |
| Beetle | 1.214 | 1.156 | 34( 3) | 1 | 1 |
| Cockroach | 1.349 | 1.875 | 28( 4) | – | – |
| Earwig | 1.349 | 1.750 | 10 – | – | 1 |
| Gnat | 1.372 | 2.063 | 8 – | – | – |
| Mosquito | 1.429 | 1.625 | 29 – | – | 1 |
| Wasp | 1.442 | 1.250 | 41( 3) | 1 | 1 |
| Flea | 1.465 | 1.375 | 14( 1) | – | – |
| Bee | 1.465 | 1.000 | 45( 3) | 1 | – |
| Cricket | 1.512 | 1.625 | 3 – | – | – |
| Ladybird | 1.595 | 1.375 | 17( 2) | 1 | 1 |
| Termite | 1.643 | 2.344 | 7 – | – | 2 |
| Dragonfly | 1.651 | 1.469 | 24 – | – | – |
| Locust | 1.651 | 1.469 | 7 – | – | – |
| Moth | 1.674 | 1.156 | 19 – | 1 | – |
| Mite | 1.707 | 2.969 | 3 – | – | 3 |
| Hornet | 1.814 | 2.250 | 9 – | 1 | 1 |
| Whitefly | 1.927 | 4.094 | – – | – | 12 |
| Caterpillar | 2.070 | 1.219 | 10 – | 2 | – |
| Tick | 2.071 | 2.844 | 5 – | 1 | 4 |
| Aphid | 2.079 | 3.000 | 3 – | 2 | 9 |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Butterfly | 2.093 | 1.063 | 30( 2) | 1 | - |
| Spider | 2.214 | 1.031 | 40( 9) | 8 | 1 |
| Louse | 2.302 | 2.344 | 7 - | 3 | 1 |
| Centipede | 2.674 | 1.781 | 10 - | 7 | - |
| Lacewings | 2.750 | 4.906 | 1 - | 1 | 19 |
| Silverfish | 3.108 | 4.313 | 1 - | 6 | 9 |
| Mantis | 3.154 | 3.344 | - - | 5 | 7 |
| Tarantula | 3.262 | 2.125 | 1 - | 8 | 1 |
| Cicada | 3.517 | 5.281 | - - | 6 | 34 |
| Scorpion | 3.651 | 1.500 | 3 - | 13 | - |
| Thrip | 3.714 | 5.438 | - - | 4 | 44 |
| Worm | 4.209 | 1.063 | 4 - | 20 | - |

Housefly (13;2), woodlouse (9;0), daddy-long-legs (8;0), grasshopper (8;0), bluebottle (7;0), greenfly (7;0), stick-insect (7;0), horsefly (5;0), midge (4;0), millipede (4;0).

Category : Sports

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Soccer | 1.000 | 1.129 | 11( 2) | - | - |
| Rugby | 1.000 | 1.097 | 47( 3) | - | - |
| Tennis | 1.022 | 1.032 | 53(15) | - | - |
| Badminton | 1.133 | 1.097 | 41( 3) | - | - |
| Basketball | 1.178 | 1.355 | 19 - | - | - |
| Hockey | 1.200 | 1.387 | 45( 5) | - | - |
| Squash | 1.267 | 1.226 | 42( 9) | - | - |
| Swimming | 1.400 | 1.129 | 49( 3) | - | - |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Baseball | 1.523 | 2.065 | 13 - | - | 1 |
| Running | 1.556 | 1.226 | 19( 1) | - | - |
| Golf | 1.733 | 1.419 | 17( 3) | - | - |
| Volleyball | 1.756 | 1.710 | 17 - | - | - |
| Ping-pong | 1.844 | 1.742 | - - | - | - |
| Boxing | 1.956 | 1.516 | 9( 1) | - | - |
| Sailing | 1.956 | 1.194 | 9 - | - | - |
| Javelin | 1.978 | 1.613 | 7 - | - | - |
| Discus | 2.000 | 1.677 | 5 - | - | - |
| Racing | 2.044 | 1.419 | - - | 1 | - |
| Lacrosse | 2.089 | 3.000 | 15 - | - | - |
| Skiing | 2.111 | 1.387 | 14( 1) | - | - |
| Gymnastics | 2.178 | 1.710 | 12( 1) | - | - |
| Rowing | 2.182 | 1.258 | 6 - | - | 1 |
| Polo | 2.356 | 2.226 | 4 - | - | - |
| Riding | 2.378 | 1.484 | 17( 1) | - | - |
| Fencing | 2.400 | 1.645 | 7 - | - | - |
| Handball | 2.409 | 2.774 | 2 - | - | 2 |
| Archery | 2.444 | 1.613 | 6 - | - | - |
| Canoeing | 2.467 | 1.226 | 13 - | - | - |
| Wrestling | 2.489 | 1.742 | 5 - | - | - |
| Judo | 2.545 | 1.677 | 6( 1) | - | 1 |
| Diving | 2.556 | 1.677 | 5 - | - | - |
| Bowls | 2.578 | 1.581 | 6 - | - | - |
| Snooker | 2.689 | 1.290 | 1 - | 3 | - |
| Skating | 2.689 | 1.419 | 3 - | - | - |
| Mountaineering | 2.711 | 1.484 | 3 - | 1 | - |
| Rifleshooting | 2.756 | 1.548 | 2 - | - | - |
| Karate | 2.867 | 1.645 | 5 - | - | - |
| Trampolining | 2.978 | 1.903 | - - | - | - |
| Billiards | 3.044 | 1.806 | 1 - | 5 | - |
| Fishing | 3.156 | 1.258 | 6 - | 1 | - |
| Pool | 3.244 | 1.548 | 1 - | 9 | - |
| Surfing | 3.267 | 1.581 | 1 - | - | - |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Croquet | 3.356 | 2.097 | 4 – | 2 | – |
| Hunting | 3.911 | 1.742 | 2 – | 7 | – |
| Potholing | 4.156 | 1.968 | – – | 6 | – |
| Hiking | 4.156 | 1.452 | 1 – | 10 | – |
| Ballet | 5.133 | 1.710 | – – | 26 | – |
| Dancing | 5.156 | 1.419 | 1 – | 24 | – |

Cricket (36;6), athletics (24;0), rounders (11;1), long jump (9;0), darts (8;0), high jump (8;0), hurdling (7;0), ice hockey (6;0), wind surfing (6;0), cross-country (5;0), hang-gliding (5;0), parachuting (5;0), climbing (4;0), cycling (4;0), ice skating (4;0), jogging (4;0), motor racing (4;0), scuba diving (4;0), water polo (4;0).

Category : Vegetables

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Carrot | 1.000 | 1.033 | 62(21) | – | – |
| Cabbage | 1.021 | 1.000 | 56(10) | – | – |
| Cauliflower | 1.104 | 1.133 | 41( 3) | – | – |
| Bean | 1.125 | 1.100 | 32 – | – | – |
| Pea | 1.146 | 1.033 | 50( 2) | 1 | – |
| Potato | 1.146 | 1.000 | 57(11) | – | – |
| Sprouts | 1.149 | 1.033 | 20 – | – | – |
| Onion | 1.375 | 1.067 | 38( 7) | – | – |
| Lettuce | 1.447 | 1.100 | 32( 1) | – | – |
| Swede | 1.543 | 1.967 | 29( 1) | – | 1 |

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Turnip | 1.604 | 1.300 | 31 - | - | - |
| Sweetcorn | 1.622 | 1.067 | 10 - | - | 2 |
| Broccoli | 1.638 | 1.700 | 20( 1) | - | 1 |
| Leek | 1.667 | 1.367 | 18( 1) | - | - |
| Spinach | 1.681 | 1.533 | 15( 2) | - | - |
| Parsnip | 1.702 | 1.667 | 18 - | - | 1 |
| Beetroot | 1.766 | 1.300 | 15( 1) | - | - |
| Cucumber | 1.936 | 1.267 | 14 - | - | - |
| Celery | 1.957 | 1.367 | 16 - | - | - |
| Asparagus | 1.958 | 2.067 | 11 - | - | - |
| Courgette | 1.977 | 1.867 | 12 - | - | 5 |
| Mushroom | 2.021 | 1.100 | 5 - | 4 | - |
| Radish | 2.125 | 1.500 | 7 - | - | - |
| Marrow | 2.170 | 1.933 | 17( 2) | - | 1 |
| Aubergine | 2.417 | 2.700 | 17( 2) | 2 | 2 |
| Watercress | 2.457 | 1.600 | 3 - | 2 | 2 |
| Lentils | 2.604 | 2.167 | 3 - | 2 | - |
| Artichoke | 2.604 | 2.567 | 6( 1) | - | - |
| Shallot | 2.689 | 3.700 | 1 - | - | 9 |
| Tomato | 2.771 | 1.000 | 33( 1) | 12 | - |
| Gherkin | 2.936 | 2.133 | - - | - | 1 |
| Pepper | 3.063 | 1.333 | 16 - | 5 | - |
| Kale | 3.156 | 4.233 | 4 - | 1 | 11 |
| Pumpkin | 3.292 | 2.333 | 2 - | 4 | - |
| Parsley | 3.404 | 1.233 | - - | 7 | - |
| Yam | 3.435 | 3.600 | 2 - | 3 | 5 |
| Chicory | 3.511 | 2.967 | - - | 6 | - |
| Fennel | 3.512 | 4.700 | - - | 4 | 19 |
| Garlic | 3.532 | 1.333 | 3 - | 9 | - |
| Chilli | 3.565 | 2.333 | 2 - | 5 | 1 |
| Endive | 3.575 | 4.833 | 1 - | 5 | 22 |

Runner bean (13;0), broad bean (8;0), corn (8;0), spring greens (6;0), French beans (5;0), green pepper (5;0), spring onion (5;0), kidney beans (4;0).

Category : Vehicles

| Word | TYP | FAM | A.F. | | N.R. | N.U. |
|---|---|---|---|---|---|---|
| Car | 1.000 | 1.000 | 66(58) | | - | - |
| Bus | 1.109 | 1.094 | 45( 2) | | - | - |
| Taxi | 1.174 | 1.219 | 5( 1) | | - | - |
| Van | 1.196 | 1.125 | 35( 1) | | - | - |
| Lorry | 1.370 | 1.125 | 50( 2) | | - | - |
| Motorbike | 1.522 | 1.031 | 34( 1) | | - | - |
| Train | 1.696 | 1.063 | 41 | - | - | - |
| Jeep | 1.696 | 1.781 | 2 | - | - | - |
| Scooter | 1.957 | 1.625 | 7 | - | - | 1 |
| Tube-train | 1.978 | 1.219 | - | - | - | - |
| Ambulance | 2.098 | 1.219 | 3 | - | - | 1 |
| Bicycle | 2.109 | 1.031 | 45 | - | 1 | - |
| Tram | 2.435 | 2.125 | 18 | - | - | - |
| Fire-engine | 2.478 | 1.375 | 4 | - | 1 | - |
| Aeroplane | 2.630 | 1.094 | 23 | - | 2 | - |
| Milk-float | 2.778 | 1.594 | 4 | - | 2 | 1 |
| Dustcart | 2.804 | 1.969 | - | - | 2 | - |
| Carriage | 2.848 | 1.656 | 4 | - | - | - |
| Ferry | 2.957 | 1.531 | 2 | - | 1 | - |
| Hovercraft | 2.978 | 1.531 | 9 | - | 1 | - |
| Tractor | 3.022 | 1.625 | 10 | - | 3 | - |
| Boat | 3.043 | 1.125 | 17 | - | 4 | - |

| Word | TYP | FAM | A.F. | | N.R. | N.U. |
|---|---|---|---|---|---|---|
| Cart | 3.087 | 1.750 | 10 | – | 4 | – |
| Helicopter | 3.130 | 1.438 | 7 | – | 1 | – |
| Tricycle | 3.196 | 1.969 | 7 | – | 2 | – |
| Ship | 3.239 | 1.188 | 19 | – | 3 | – |
| Car-ferry | 3.283 | 1.625 | – | – | 2 | – |
| Bulldozer | 3.391 | 1.969 | 1 | – | 2 | – |
| Hydrofoil | 3.432 | 2.469 | 1 | – | 2 | 4 |
| Steamroller | 3.478 | 1.938 | 2 | – | 2 | – |
| Tank | 3.478 | 1.563 | 12 | – | 3 | – |
| Wheelchair | 3.543 | 1.375 | – | – | 4 | – |
| Ocean liner | 3.578 | 1.688 | – | – | 4 | 1 |
| Cablecar | 3.696 | 1.938 | – | – | 1 | – |
| Rickshaw | 3.773 | 2.906 | 3 | – | 4 | 5 |
| Canoe | 3.826 | 1.344 | 3 | – | 5 | – |
| Pram | 3.889 | 1.656 | 3 | – | 6 | 1 |
| Spaceship | 3.891 | 1.719 | 3 | – | 6 | – |
| Airship | 3.913 | 2.125 | 2 | – | 2 | – |
| Sleigh | 3.935 | 2.063 | – | – | 5 | – |
| Submarine | 4.022 | 1.719 | – | – | 8 | – |
| Shuttle | 4.023 | 2.563 | – | – | 4 | 3 |
| Glider | 4.109 | 1.813 | 1 | – | 7 | – |
| Sled | 4.217 | 2.781 | – | – | 6 | 3 |
| Trolley | 4.217 | 2.063 | – | – | 8 | – |
| Balloon | 4.239 | 1.500 | 1 | – | 10 | – |
| Toboggan | 4.311 | 2.344 | 1 | – | 8 | 2 |
| Dodgem | 4.489 | 1.938 | – | – | 9 | 1 |
| Lift | 4.500 | 1.313 | – | – | 16 | – |
| Hang-glider | 4.565 | 1.656 | – | – | 12 | – |
| Raft | 4.674 | 1.875 | 2 | – | 13 | – |
| Skates | 4.848 | 1.656 | – | – | 15 | – |
| Skateboard | 4.891 | 1.594 | 3 | – | 17 | – |
| Escalator | 5.283 | 1.344 | – | – | 28 | – |

Truck (20;0), coach (17;0), plane (13;1), moped (10;0), motorcycle (9;0), horse (7;0), push-bike (7;0), tandem (7;0), wagon (6;0), barge (5;0), buggy (4;0), Ford/ Fiesta/Mini (4;0), forklift truck (4;0), juggernaut (4;0), rocket (4;0).

Category : Weapons

| Word | TYP | FAM | A.F. | N.R. | N.U. |
|---|---|---|---|---|---|
| Machine-gun | 1.045 | 1.100 | 9 - | - | - |
| Revolver | 1.068 | 1.233 | 6 - | - | - |
| Gun | 1.068 | 1.000 | 59(37) | - | - |
| Rifle | 1.068 | 1.133 | 34( 4) | - | - |
| Pistol | 1.091 | 1.100 | 22( 2) | - | - |
| Shotgun | 1.273 | 1.200 | 3 - | - | - |
| Bomb | 1.341 | 1.200 | 37( 3) | - | - |
| Sword | 1.364 | 1.133 | 38( 5) | - | - |
| Grenade | 1.386 | 1.600 | 15 - | - | 1 |
| Spear | 1.535 | 1.400 | 24( 2) | - | 1 |
| Flick-knife | 1.568 | 1.800 | 1 - | - | - |
| Bayonet | 1.659 | 1.933 | 3 - | - | - |
| Arrow | 1.705 | 1.267 | 15( 1) | - | - |
| Missile | 1.773 | 1.300 | 7 - | - | - |
| Torpedo | 1.818 | 1.567 | 5 - | - | - |
| Cannon | 1.864 | 1.433 | 19 - | - | - |
| Knife | 1.864 | 1.067 | 53( 7) | - | - |
| Crossbow | 2.023 | 1.633 | 7 - | - | - |
| Explosive | 2.068 | 1.333 | 2 - | - | - |
| Sabre | 2.093 | 2.567 | 3 - | - | 3 |

| Word | TYP | FAM | A.F. | | N.R. | N.U. |
|---|---|---|---|---|---|---|
| Landmine | 2.114 | 1.900 | – | – | – | – |
| Bazooka | 2.119 | 2.567 | 3 | – | – | 6 |
| Club | 2.159 | 1.867 | 16 | – | – | – |
| Axe | 2.205 | 1.100 | 8 | – | – | – |
| Mortar | 2.205 | 2.133 | 5 | – | 2 | 1 |
| Harpoon | 2.591 | 1.967 | 1 | – | 1 | – |
| Dynamite | 2.614 | 1.300 | – | – | 2 | – |
| Hatchet | 2.628 | 2.167 | 3 | – | 1 | 1 |
| Lance | 2.682 | 2.267 | 6 | – | 2 | – |
| Machete | 2.744 | 2.967 | 2 | – | 2 | 5 |
| Catapult | 2.773 | 1.833 | 5 | – | 1 | 1 |
| Cut-throat razor | 2.864 | 1.833 | – | – | 2 | – |
| Whip | 3.045 | 1.400 | 3 | – | – | – |
| Crowbar | 3.114 | 1.833 | 1 | – | – | – |
| Rocket | 3.159 | 1.400 | 8 | – | 4 | – |
| Sling shot | 3.341 | 2.767 | 5 | – | 3 | 1 |
| Chain | 3.523 | 1.433 | 3 | – | 1 | – |
| Laser | 3.886 | 2.067 | – | – | 3 | 1 |
| Dart | 4.182 | 1.367 | 1 | – | 5 | – |
| Hammer | 4.205 | 1.300 | 5 | – | 6 | – |

Dagger (20;1), bow and arrows (18;0), truncheon (8;0), nuclear bomb (7;1), stone (7;0), fist (6;0), razor blade (5;0), cutlass (5;0), rope (5;0), atom bomb (4;1), airgun (4;0), brick (4;0), longbow (4;0).

## APPENDIX FIVE - MATERIALS FOR EXPERIMENT 5

### Category : Weapons

Typical words : Revolver, spear, crossbow, bayonet, explosive, grenade, arrow, sabre, flick-knife, cannon, missile, shotgun.

Atypical words: Hatchet, whip, dart, catapult, club, hammer, dynamite, crowbar, rocket, axe, chain, laser.

### Category : Birds

Typical words : Starling, crow, dove, raven, parrot, swallow, thrush, wren, nightingale, hawk, cuckoo, seagull.

Atypical words: Penguin, ostrich, flamingo, stork, peacock, hen, puffin, goose, vulture, turkey, pelican, duck.

### Category: Clothing

Typical words : Dungarees, tights, cardigan, jumper, overcoat, blouse, slacks, anorak, bra, skirt, jacket, vest.

Atypical words: Slippers, pyjamas, bow-tie, scarf, belt, bathrobe, sandals, overalls, swimsuit, gloves, hat, apron.

### Category : Fruit

Typical words : Peach, blackcurrant, mandarin, apricot, tangerine, raspberry, cherry, lime, plum, satsuma, blackberry, pineapple.

Atypical words: Avocado, acorn, watermelon, pumpkin, date, prune, fig, olive, almond, raisin, gooseberry, coconut.

Category : Furniture

Typical words : Sofa, dresser, bureau, sideboard, bookcase, wardrobe, cot, desk, settee, cabinet, couch, stool.

Atypical words: Shelves, garden-swing, ashtray, wall-mirror, bar, sink-unit, waste paper basket, counter, deck-chair, painting, trolley, bunk.

Category : Vehicles

Typical words : Taxi, hovercraft, jeep, ambulance, lorry, milk-float, van, carriage, tram, fire-engine, scooter, tube-train.

Atypical words: Balloon, dodgem, hang-glider, ship, wheelchair, pram, submarine, lift, skateboard, helicopter, tank, canoe.

Category : Sports

Typical words : Basketball, squash, ping-pong, volleyball, javelin, racing, hockey, baseball, golf, boxing, discus, sailing.

Atypical words: Hiking, bowls, fishing, skating, shooting, billiards, karate, surfing, hunting, diving, mountaineering, canoeing.

Category : Vegetables

Typical words : Sprouts, courgette, turnip, leek, swede, lettuce, parsnip, asparagus, cauliflower, bean, spinach, broccoli.

Atypical words: Watercress, chilli, garlic, celery, parsley, mushroom, tomato, pumpkin, gherkin, cucumber, pepper, radish.

## APPENDIX SIX - MATERIALS FOR EXPERIMENT 6

### Category : Countries

France, Ireland, Japan, Mexico, Canada, Germany, India, Peru, Java, Zambia, Australia, Columbia, Guatemala, Egypt, Lybia, China.

### Category : Fish

Typical words : Haddock, sardine, whiting, halibut, tuna, bass, pilchard, carp.

Atypical words: Shark, eel, swordfish, stickleback, bream, guppy, piranha, perch.

### Category : Vegetables

Typical words : Sprouts, sweetcorn, broccoli, leek, cucumber, parsnip, beetroot, spinach.

Atypical words: Celery, asparagus, courgette, marrow, aubergine, peppers, tomato, radish.

### Category : Vehicles

Typical words : Taxi, ambulance, scooter, tram, fire-engine, aeroplane, carriage, hovercraft.

Atypical words: Tractor, boat, cart, helicopter, tricycle, ship, tank, pram.

### Category : Birds

Typical words : Starling, wren, dove, raven, woodpecker, nightingale, crow, cuckoo.

Atypical words: Duck, heron, ostrich, flamingo, pelican, vulture, penguin, goose.

Category : Insects

Typical words : Earwig, flea, ladybird, termite, gnat, mosquito, cockroach, dragonfly.
Atypical words: Moth, hornet, caterpillar, butterfly, spider, centipede, louse, tick.

Category : Sports

Typical words : Soccer, basketball, golf, baseball, javelin, running, discus, sailing.
Atypical words: Lacrosse, skiing, gymnastics, riding, canoeing, fencing, bowls, judo.

## APPENDIX SEVEN - MATERIALS FOR THE "DIFFERENT" LISTS IN EXPERIMENT 7

Category : Insects

Typical words : Flea, earwig, cockroach, gnat, mosquito, termite.

Atypical words: Scorpion, centipede, tarantula, butterfly, caterpillar, louse.

Category: Fish

Typical words : Halibut, herring, mackerel, sole, pike, tuna.

Atypical words: Shark, swordfish, piranha, eel, stickleback, anchovy.

*Memory & Cognition*
*1983, 11 (6), 616-620*

# The auditory recency advantage in longer term free recall is not enhanced by recalling prerecency items first

JOHN M. GARDINER and MARGARET M. GARDINER
*The City University, London, England*

and

VERNON H. GREGG
*Birkbeck College, University of London, London, England*

*Memory & Cognition*
*1983, 11 (6), 616-620*

# The auditory recency advantage in longer term free recall is not enhanced by recalling prerecency items first

JOHN M. GARDINER and MARGARET M. GARDINER
*The City University, London, England*

and

VERNON H. GREGG
*Birkbeck College, University of London, London, England*

**Gardiner and Gregg (1979) showed that in a free-recall paradigm in which each list word is embedded in a continuous stream of subject-vocalized distractor activity, recency recall was greater when the words were presented auditorily rather than visually. The experiment described here showed that this auditory advantage persisted even when list and distractor items were both spoken at a controlled pace by the experimenter, and that it was little influenced by instructions to give priority in recall either to the beginning or to the end of the list. These results strengthen the conclusion that this effect cannot be accommodated by any echoic memory theory and, because the effect was not enhanced when prerecency items were recalled first, demonstrate an additional difference between it and the somewhat similar auditory advantage found in immediate recall.**

The recency effect in immediate recall is greater if the items in a list are spoken rather than written, and it has been widely accepted that this auditory advantage arises from echoic memory (e.g., Broadbent, Vines, & Broadbent, 1978; Crowder, 1976; Crowder & Morton, 1969; O. C. Watkins & M. J. Watkins, 1980). Echoic memory interpretations depend critically on evidence that auditory, but not visual, recency is vulnerable to interference from subsequent auditory input, and, of course, such modality-specific interference effects are well documented. However, evidence that seems to be incompatible with any echoic memory theory is now accumulating from a number of sources (for a recent review, see Gardiner, 1983). For example, there is now evidence that auditory and visual input may be functionally equivalent when the visual input entails lipreading (e.g., Campbell & Dodd, 1980; Gardiner, Gathercole, & Gregg, 1983; Spoehr & Corin, 1978). There is also evidence, in a study by Gardiner & Gregg (1979), of an auditory advantage in the longer term free-recall paradigm of Tzeng (1973; also see Bjork & Whitten, 1974), that is, when a spoken distractor task occurs before and after the presentation of every *single* word in the list. The present study was empirically and methodologically oriented, and it was concerned with this last finding.

This research was supported by a grant from the Social Science Research Council. Requests for reprints may be sent to John M. Gardiner, Psychology Division, The City University, Northampton Square, London EC1V 0HB, England.

The finding had not been anticipated either on theoretical or on empirical grounds, because, in accord with echoic memory interpretations, the same distractor task had been shown to remove the auditory advantage when the task occurred only after all the words in the list had been presented (Gardiner, Thompson, & Maskarinec, 1974; also see Broadbent et al., 1978, and Martin & Jones, 1979). In several tests of its generality, however, Gardiner and Gregg (1979) failed to discover any boundary conditions of the longer term auditory advantage, and they obtained no evidence that that effect differed in any other respects from the more familiar immediate-recall effect. In immediate recall, there are only two variables for which some influence upon the auditory advantage seems firmly established. One of these variables is phonological similarity among the list items, which is known to sharply reduce, if not eliminate, the immediate-recall effect (see, e.g., M. J. Watkins, O. C. Watkins, & Crowder, 1974); the possible influence of this variable on the longer term effect is the subject of another, forthcoming study (Gregg & Gardiner, in press). The other variable is the instructions with regard to the order of recall.

There are three immediate-recall studies that have examined the effect of order-of-recall instructions on the auditory advantage. Craik (1969), using modified free-recall instructions, and Madigan (1971), using forward- or backward-serial-order recall instructions, found that when recall starts from the beginning of the list, recency recall is reduced but the auditory advantage is greatly enhanced. And Nilsson, Wright, and Murdock (1979) found a similar pattern of results with instructions to recall words near the end of the list first, in either a forward or a backward order, although, not surprisingly, their effect was much less pronounced. The major purpose of the present study was to investigate the effect of instructions to start recall from the beginning, rather than from the end, of the list on the auditory advantage in the distractor paradigm and so provide a further convergent test with respect to the immediate- and longer term recall effects. The question at hand was whether the auditory advantage in longer term recall is similarly enhanced when prerecency items are recalled first.

A further aim of the present study was to replicate the longer term effect under more rigorously controlled conditions. In Gardiner and Gregg's (1979) study, list items had been spoken at the time of presentation by the experimenter or presented for subject vocalization manually via decks of cards. Distractor items had also been presented on cards for a self-paced counting task: counting aloud backward by threes from a given number. And order of recall was controlled only insofar as subjects were advised that they would find it helpful to try to recall the last words in each list first. So, in the following experiments, not only were the subjects instructed to recall items from either the later or the earlier part of the list first (cf. Craik, 1969), but also both distractor and list items were prerecorded. Otherwise, the procedure was comparable to that used in the later experiments by Gardiner and Gregg (1979), including the use of a list length of only six words and a shortened, 10-sec period of interspersed distraction.

## METHOD

### Subjects

Thirty-two undergraduate students at The City University were tested individually. They were paid for their services.

### Design

The experimental design was completely within-subjects, with list modality (auditory or visual), recall instructions ("end" or "beginning"), and serial position as the principal independent variables.

All subjects were presented with a total of 20 experimental lists, with six words in each. Ten lists were presented auditorily and 10 visually. Subjects were tested in two separate ½-h sessions, with a ½-h break between sessions, and list modality was blocked by sessions, such that half the subjects had auditory lists in the first session and half had visual ones. Within each session, half the lists were followed by "end" instructions, and half by "beginning" instructions. The ordering of these instructions with respect to lists was determined randomly, but with the additional constraint that no more than three consecutive lists were followed by the same instruction. Before and after the presentation of each list word, subjects had to copy down a sequence of five three-digit numbers. The numbers were spoken in the same voice, in the same manner, and at the same rate as the words. List words, distractor materials, and the ordering of instructions were all balanced or rotated across other experimental conditions.

### Materials and Procedure

By sampling randomly without replacement, four different sets of 20 lists of six words were constructed from a pool of 480 common, two-syllable nouns. Sufficient batches of five three-digit numbers were taken from random-number tables to put together with these word sets such that each word was preceded and followed by one batch of numbers. Two recordings of each of the four sets of word lists and the associated numbers were made. Two videorecorders were used to record "auditory" and "visual" tapes simultaneously. This was done in the following way. One of the two recorders had the video channel permanently switched off and so this recorded the auditory lists (the experimenter read aloud the list words as well as the distractor digits). The second recorder was also programmed to record the distractor digits, but at each word presentation the sound channel was switched off. At the same time, a camera was automatically switched on to record through the video channel a printed representation of the appropriate word: for this, each word was simply displayed on a card in front of the camera. A Commodore PET microcomputer was programmed to control switching from sound to video channels and to control the timing of distractor and word sequences by displaying warning lights for the experimenter while the recording was in progress. Great care was taken, in making the recording, to ensure that the recorded list and distractor items were all spoken in a regular and even manner, with pitch, stress, and tone as similar as possible. For the subsequent benefit of the subjects, bleeps were also recorded at certain points during the sequence.

For any one list, the sequence of events was as follows. First, a bleep was recorded, as a warning signal for the subjects, and there was then a 2-sec unfilled interval before the first of the five three-digit numbers was recorded. These numbers were recorded regularly at the rate of one triad every 2-sec period that elapsed, and, during the same interval after the fifth triad, another bleep was recorded, this one to signal the onset of a word presentation. At the same instant, the visual presentation of a word was recorded on the videotape, with the duration of the word lasting a full 2-sec interval. The auditory presentation of the same word was taped on the other recorder during the same 2-sec period, but it was spoken so as to occur more or less during the middle of the period, that is, in rhythm with each three-digit number. Within the next 2-sec period, the first of the next batch of three-digit numbers was recorded, and so on to the end of the list. At the end of a list, within the 2-sec period directly following that in which the last numbers were presented, a general recall instruction was recorded. For auditory lists, this signal consisted of the experimenter's saying "recall"; for visual lists, a diagonal line across the screen was recorded. Two additional tapes were recorded, each one having a single practice list for each modality of presentation. All subjects received the same practice lists at the beginning of the appropriate testing session.

All the subjects first read detailed, typewritten instruction sheets, which were then supplemented orally by the experimenter. They were informed that the experiment involved two separate but equally important tasks, a copying task and a memory task. They were given recall booklets with one page for each list, each page containing seven columns of five lines to be used to write down the numbers in the copying task. Each page also contained a space labeled "words recalled." The importance of accurately copying down the numbers was stressed, and the subjects' performance of this task was carefully and manifestly monitored by the experimenter. The subjects were also told to focus on each word presentation only during its occurrence and to try not to think about the words in any way while they were copying down the numbers.

Lastly, the subjects were of course told that, in addition to the general, recorded recall signal, they would, at the end of each list, be given particular recall instructions by the experimenter; the experimenter would at that time say either "end" or "beginning." They were told that these instructions meant, respectively, that they should start by writing down any words they could remember from the second or from the first part of the list, and only then to go on to write down any other words that they could still remember. They were told that, in both cases, they could write the words in any order they liked, so long as they gave priority in recall as instructed. Recall of each list typically took about half a minute or so.

## RESULTS

The recall data, summarized in Figure 1, appear to show a pronounced auditory advantage in recall of recency items and, moreover, an advantage that does not seem much affected by recall instructions, despite the marked effect of those instructions upon the level of recall of recency items generally. This description of the data was well supported statistically by the results of an overall analysis of variance of individual subject recall scores. The serial position effect was significant [$F(5,150) = 22.09$, $MSe = 1.17$, $p < .001$]. Recall was superior with auditory presentation [$F(1,30) = 7.04$, $MSe = 1.50$, $p < .01$] and also with "end" instructions [$F(1,30) = 5.03$, $MSe = 1.13$, $p < .05$]. Significant interactions involving serial position indicated that the auditory advantage was localized over the later serial positions [$F(5,140) = 5.40$, $MSe = 1.10$, $p < .001$], and that so was the effect of instructions [$F(5,140) = 4.13$, $MSe = 0.95$, $p < .001$]. There was also a significant interaction between instructions and modality [$F(1,30) = 4.93$, $MSe = 0.61$, $p < .05$]: The superiority in recall with "end" instructions was confined largely to auditory lists. Although the three-way interaction was not significant, Figure 1 shows that that interaction was due to differences in prerecency, not in recency, recall. The magnitude of the modality difference in recency recall was essentially unchanged.

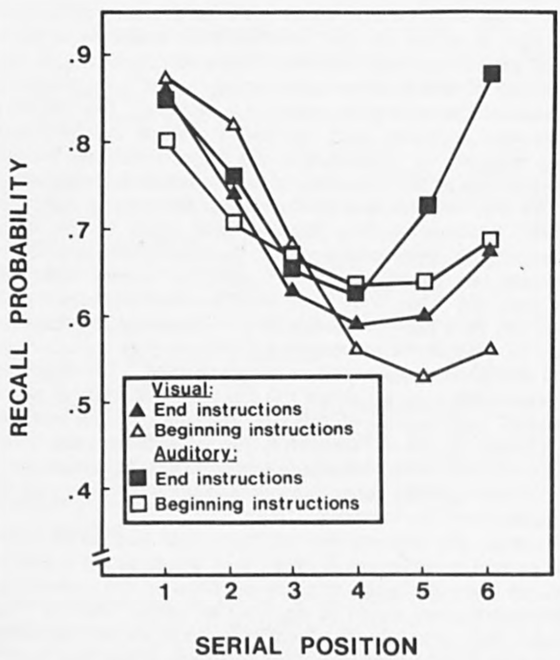



**Figure 1. Probability of recall as a function of presentation modality and recall instructions.**

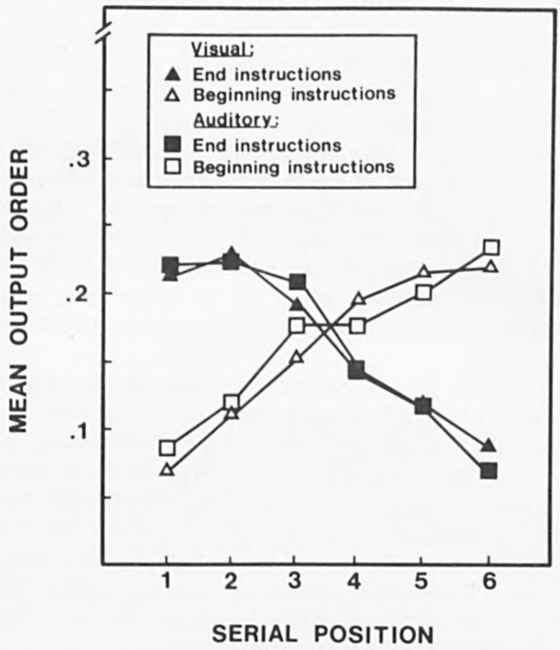



**Figure 2. Mean output order as a function of presentation modality and recall instructions.**

Because the existence of differences in the overall level of recall may itself lead to differences in output-order measures, order-of-recall data were analyzed using a normalization procedure, according to which the mean output position of items recalled from each serial position was proportionalized on the mean total in each condition. These data are summarized in Figure 2. It is apparent that subjects complied well with the recall instructions. Moreover, there seems to be little indication that the pattern of recall order was affected by modality. These conclusions were well supported statistically by the results of an overall analysis of variance which, in order to use more reasonable estimates of output-order distributions, was carried out on macrosubject scores obtained by collapsing data over every four successively tested subjects. The interaction between instructions and serial position was highly significant [$F(5,35) = 228.67$, $MSe = 0.0005$, $p < .001$]. The serial-position effect was also significant [$F(5,35) = 7.43$, $MSe = 0.0005$, $p < .001$]. A similar analysis of the raw output-order data gave comparable results.

## DISCUSSION

It is clear that the auditory advantage observed in previous similar experiments (Gardiner & Gregg, 1979) is readily obtained under the much more rigorous conditions of the present experiment. It is of some importance to have shown: that the effect does not depend critically on subject vocalization of a self-paced distractor task; that it occurs when both list and distractor items are spoken in a regular and even manner by the experimenter; and that it does not depend in any simple way on recall order. The experiment also showed that the effect was quite definitely not enhanced by recalling prerecency items first. This outcome contrasts strongly with that from immediate-recall studies (Craik, 1969; Madigan, 1971; see too Nilsson et al., 1979).

Instructions to give priority in recall to the beginning of the list did, however reduce recency recall generally, as had been found in previous studies—excepting an experiment in this very distractor paradigm described by Whitten (1978). Whitten, however, had presented 18-word lists and in word-doubles rather than one at a time. And although his analysis revealed no effect of instructions, his data do show that recall of the last word-double in a list was about 20% lower with "beginning" than with "end" instructions—a reduction of a similar order of magnitude to that observed in our data. There was also an interaction between instructions and modality, but this apparently came about because in the visual, but not in the auditory, modality, our data replicate an observation made by Dalezman (1976). In a *delayed*-recall test, he too found that similar recall instructions increased recall of first-recalled items but correspondingly reduced recall of later-recalled items, thus leaving overall recall unchanged. This was not so in the auditory modality, where our data show that recall of prerecency items was not reduced by recalling recency items first. This cannot be related to modality differences in the output-order data, and it seems possible that greater susceptibility of visual items to output interference may have contributed to this slightly complex pattern of results (Madigan, 1971; Nilsson et al., 1979).

Perhaps the most interesting finding is that the auditory advantage in longer term recall has now been shown not only to be not removed by subsequent auditory input, but also to be not enhanced by starting recall from the beginning of the list. Moreover, we present evidence in a forthcoming study that, at least under certain conditions, the effect is not eliminated or even reduced by phonological similarity among list items (Gregg & Gardiner, in press). With respect to none of these variables does the effect seem to correspond functionally with the auditory advantage in immediate recall. By the same token, we have found little evidence in this paradigm of any dissociation between auditory and visual recency. This is consistent with the possibility that whatever gives rise to recency recall here may also be partly responsible for the auditory recency advantage. In particular, we had previously speculated that the auditory advantage in longer term recall might simply reflect greater temporal distinctiveness in the auditory mode (Gardiner & Gregg, 1979). This conjecture accords well with the view that recency itself is due to a backward-scanning retrieval strategy which utilizes ordinal retrieval cues (e.g., Bjork & Whitten, 1974; Glenberg, Bradley, Stevenson, Kraus, Tkachuk, Gretz, Fish, & Turpin, 1980; Glenberg & Kraus, 1981). And we note that, although their theory does not explain them, our findings in this paradigm generally lend quite good support to a two-process theory of the sort proposed by Glenberg et al. (1980) to account for long-term serial position effects.

Evidence of a somewhat different auditory advantage in longer term recall does not, of course, necessarily mean that this effect might not be integrated with other modality differences in recency recall in some quite general theory. Considered more broadly, a temporal-distinctiveness hypothesis fits nicely with some other evidence—including evidence that the auditory mode is indeed more specialized for temporal processing (Metcalfe, Glavanov, & Murdock, 1981)—and Gardiner (1983) has argued that it may provide one such possibility; other similarly broad hypotheses have been proposed recently by Campbell and Dodd (1980) and by Shand and Klima (1981) in connection, respectively, with recency advantages obtained with lipreading and with sign language in immediate recall. If a theory that accounts for both the immediate and the longer term auditory advantage is to be viable, then some satisfactory resolution of apparent differences between those effects has to be achieved. This remains feasible. In the present case, for instance, it is possible that the finding of an enhanced auditory advantage in immediate recall when recall starts from the beginning of the list might be due largely to the presence of a ceiling effect when recall starts from the end of the list, and hence be of little theoretical significance (for further discussion, see Gardiner, 1983).

Be that as it may, the present, empirically motivated study strengthens the conclusion that the auditory advantage in longer term recall cannot be explained in terms of an echoic memory interpretation.[1] Moreover, it demonstrates an additional difference in the nature of the effect in this paradigm compared with that observed in immediate recall. But it leaves the question of the relation between immediate and longer term effects quite open. Although the evidence distinguishing between the two effects suggests that they may require some degree of theoretical separation, it does not compel that conclusion.

## REFERENCES


Bjork, R. A., & Whitten, W. B. Recency-sensitive retrieval processes in long-term free-recall. *Cognitive Psychology*, 1974, 6, 173-189.

Broadbent, D. E., Vines, R., & Broadbent, M. H. P. Recency effects in memory, as a function of modality of intervening events. *Psychological Research*, 1978, 40, 5-13.

Campbell, R., & Dodd, B. Hearing by eye. *Quarterly Journal of Experimental Psychology*, 1980, 32, 85-99.

Craik, F. I. M. Modality effects in short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 658-664.

Crowder, R. G. *Principles of learning and memory.* Hillsdale, N.J: Erlbaum, 1976.

Crowder, R. G., & Morton, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, 5, 365-373.

Dalezman, J. J. Effects of output order in immediate, delayed, and final recall performance. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 597-608.

Gardiner, J. M. On recency and echoic memory. *Philosophical Transactions of the Royal Society London, B*, 1983, **302**, 267-282.

Gardiner, J. M., Gathercole, S. E., & Gregg, V. H. Further evidence of interference between lipreading and auditory recency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1983, **9**, 328-333.

Gardiner, J. M., & Gregg, V. H. When auditory memory is not overwritten. *Journal of Verbal Learning and Verbal Behavior*, 1979, **18**, 705-719.

Gardiner, J. M., Thompson, C. P., & Maskarinec, A. S. Negative recency in initial free recall. *Journal of Experimental Psychology*, 1974, **103**, 71-78.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., Fish, J. H., & Turpin, B. A. M. A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, **6**, 355-369.

Glenberg, A. M., & Kraus, T. A. Long-term recency is not found on a recognition test. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **7**, 475-479.

Gregg, V. H., & Gardiner, J. M. Phonological similarity and enhanced auditory recency in longer-term free recall. *Quarterly Journal of Experimental Psychology, A*, in press.

Madigan, S. A. Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology*, 1971, **87**, 294-296.

Martin, M., & Jones, G. V. Modality dependence of loss of recency in free recall. *Psychological Research*, 1979, **40**, 273-289.

Metcalfe, J., Glavanov, D., & Murdock, M. Spatial and temporal processing in the auditory and visual modalities. *Memory & Cognition*, 1981, **9**, 351-359.

Nilsson, L. G., Wright, E., & Murdock, B. B., Jr. Order of recall, output interference, and the modality effect. *Psychological Research*, 1979, **41**, 63-78.

Shand, M. A., & Klima, E. S. Nonauditory suffix effects in congenitally deaf signers of American Sign Language. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **7**, 464-474.

Spoehr, K. T., & Corin, W. J. The stimulus suffix effect as a memory coding phenomenon. *Memory & Cognition*, 1978, **6**, 583-589.

Tzeng, O. J. L. Positive recency in delayed free recall. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 436-439.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 430-447.

Watkins, O. C., & Watkins, M. J. The modality effect and echoic persistence. *Journal of Experimental Psychology: General*, 1980, **109**, 251-278.

Whitten, W. B. Output interference and long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, **4**, 685-692.

## NOTE

1. This was disputed by an anonymous referee who suggested that echoic memory might somehow "sneak through" in this paradigm and be transformed into a more permanent trace during the gap between word and number presentations. In fact we had considered this possibility earlier, in thinking about our original findings (Gardiner & Gregg, 1979), and we had discounted it for several reasons. It is not only post hoc: it seems implausible. Why does echoic memory not sneak through in the similar gaps between list and distractor items in paradigms where distractors occur just at the end of the list? How might having distractors after every word facilitate the recoding of echoic information?

(Manuscript received December 3, 1982;
revision accepted for publication May 27, 1983.)

American Journal of Psychology, in press



# Concerning some more Evidence of an Auditory Advantage in Prerecency as well as Recency Recall

John M. Gardiner
The City University, London

Vernon H. Gregg
Birkbeck College, London

Margaret M. Gardiner
The City University, London

Running head: AUDITORY ADVANTAGE IN RECALL

Abstract

Comparisons of recall of spoken and written lists reveal an advantage in recall to spoken items. This auditory advantage is typically, but not invariably, restricted to recency items and it is typically, but not invariably, removed when distractor items are spoken after the list. The effect has generally been attributed to a short-lived echoic memory store. An experiment is described in which, contrary to this view, an auditory advantage occurred when each list item was directly followed (and preceded) by a sequence of spoken distractors. This auditory advantage occurred for prerecency as well as recency items, and it occurred under free recall, serial recall and backward recall conditions.

The normally good recall of the last few items in a list depends partly on the mode in which items are presented. This recency effect is usually enhanced, for example, when the items are spoken, rather than written. An auditory recency advantage occurs in both free and serial recall, and the effect has similar properties in each paradigm. In each case, the effect is typically reduced or eliminated by subsequent, spoken distractors (e.g., J.M. Gardiner, Thompson & Maskarinec, 1974; O.C. Watkins & M.J. Watkins, 1980); it is vulnerable to phonological similarity amongst the list items (M.J. Watkins, O.C. Watkins & Crowder, 1974), and it is independent of their syllabic length (M.J. Watkins, 1972; M.J. Watkins & O.C. Watkins, 1973). Moreover, there is little evidence to indicate any fundamental distinction between the free and serial recall effects. Hence it is reasonable to attribute them to the same underlying memory system.

The general view has been that the auditory recency advantage originates in echoic memory. The theory is that echoic information persists somewhat longer than corresponding information in the visual mode, and at least long enough to supplement recall of the last few items in a list. More specific echoic memory theories of this type have differed on whether echoic memory is conceived as an auditory sensory store or in more operational terms; on whether echoic information is recoded into a short- or a long-term memory system; and on whether it might persist more indefinitely and so supplement recall directly. Following Crowder and Morton's (1969) highly influential theory, however, the predominant view has been that of an auditory sensory memory that is subject to rapid decay and to erasure by subsequent, similar auditory input. (For more discussion see, e.g., Crowder, 1978; 1983; J.M. Gardiner, 1983; O.C. Watkins & M.J. Watkins, 1980.)

The very predominance of this view has led to the auditory recency advantage being emphasized rather separately from other evidence showing that an auditory advantage is not in fact always pinned to the recency effect. There is sometimes an auditory advantage in prerecency as well as recency recall. An auditory advantage extending through the list seems invariably to have been obtained with recall (and recognition) probe procedures and when spoken and written items are presented in a mixed list (Brems, 1983; Murdock, 1967;

1968; Murdock & Walker, 1969; Nilsson, 1979; Nilsson, Ohlsson & Rönneberg, 1980; M.J. Watkins, 1983). Also, an auditory advantage extending through the list has sometimes--but not invariably--been obtained when subjects were given a concurrent, interitem monitoring task (Routh, 1970; 1971; 1976). There is evidence too of a through-list effect when the list words constitute a sentence (Johannson & Nilsson, 1979).

The purpose of this article is to describe a single, large study, the results of which unexpectedly revealed another auditory advantage extending through prerecency as well as recency recall, and to offer some speculations about the significance of this effect. In the procedure used, each list item is directly preceded and followed by a sequence of spoken distractor items. Not only does a recency effect occur in this continual distractor paradigm (Bjork & Whitten, 1974; Tzeng, 1973), but so does an auditory recency advantage (J.M. Gardiner & Gregg, 1979). The auditory recency advantage in this paradigm has now been obtained over a fairly wide variety of experimental conditions, including several in which list and distractor items were spoken by the same voice (see J.M. Gardiner, M.M. Gardiner & Gregg, 1983; J.M. Gardiner & Gregg, 1979; Gregg & J.M. Gardiner, 1984; see too, Glenberg, 1984). Because no echoic memory theory predicts an auditory recency advantage when list items are directly followed by same-voice distractor items, it seems reasonable to conclude that the effect in this distractor paradigm can not be attributed to echoic memory.

The motivation for the present experiment stemmed from the desirability of obtaining further evidence on the possible influence of order of recall on the auditory recency advantage in this distractor paradigm, especially under "same-voice" conditions. J.M. Gardiner et al. (1983) found that giving subjects instructions to recall first items from the beginning as opposed to the end of the list reduced recency generally but did not apparently influence the auditory recency advantage. Glenberg (1984, Exp. 3) also showed that the effect occurred irrespective of whether subjects were cued for forward serial recall, backward serial recall, or free recall. Only in the former study, however, were list and distractor items spoken in the same voice. And in both studies order of recall was prescribed <u>after</u> each list had been presented. The present experiment was designed to investigate the possibility that the effect might also be obtained when list and distractor items are spoken in the same voice and instructions regarding recall order are given prior to list presentation.

There were three order-of-recall conditons: free recall, (forward) serial recall, and backward serial recall (cf. Glenberg, 1984; Madigan, 1971). Following J.M. Gardiner et al.'s (1983) procedure, both list and distractor items were prerecorded on videotapes in the experimenter's voice in such a way that the distractors were identical for each list mode. In contrast with J.M. Gardiner et al., who simply compared auditory with visual presentation, in the present study list items were visually displayed for each list mode and the experimenter's voice was recorded in conjunction with the visual display for the auditory condition. This was therefore more properly an "auditory-visual" condition, and an experimenter-analogue of the more commonly used subject-vocalization procedure. Previous experiments in the distractor paradigm had shown that an auditory recency advantage occurs with subject-vocalization (J.M. Gardiner & Gregg, 1979; Gregg & J.M. Gardiner, 1984) and the general finding, in more conventional paradigms, is that subject and experimenter vocalization are equivalent, at least with respect to the auditory recency advantage (e.g., Gathercole, J.M. Gardiner & Gregg, 1982). Our auditory-visual condition was somewhat more rigorous than subject-vocalization, for the sole difference with respect to presentation modality was whether the subjects heard as wellas saw the to-be-recalled items. Presentation modality was designated a between-subjects variable. The subjects attended three separate sessions, each session one week apart. In the first session, all the subjects were given free recall tests. In the second and third sessions, half had serial (forward) then backward recall, half had backward then serial recall tests. An auditory recency advantage was expected in free recall, in backward recall, and quite possibly in serial recall too.

Method

Subjects

The subjects were 48 undergraduate students at The City University, London, none of whom had attended a psychology course or participated in a psychology experiment before, and they were paid for their participation. The subjects were assigned to one of two equal groups alternately by order of their arrival in the laboratory.

Design

The experimental design was a 2 x 3 x 8 factorial with modality (visual or auditory-visual) as a between-subjects variable, and order (free, serial or backward recall) and serial position (1-8) as within-subjects variables. All subjects were presented with 30 lists of eight words, 10 in each of three test sessions at weekly intervals. For one group of 24 subjects, all words were presented in the auditory-visual condition, for the other group of 24 subjects, all words were presented in the visual condition. In all other respects subjects in both groups were treated identically. The words and digits used were rotated and balanced across all other experimental conditions. The videotapes used to present the materials were recorded in parallel such that identical tapes were used for subjects in each group, except that in the auditory-visual group in addition to seeing the words the subjects heard the experimenter say them. Before and after every single word in all experimental conditions the subjects heard the experimenter say five three-digit numbers and their task was to copy them down accurately. Order was a between-sessions variable. In the first of the three sessions all subjects were tested under free recall conditions. This was to avoid any bias in free recall performance that might result were free recall to follow one or other of the ordered recall tests. In the second and third sessions, half the subjects in each group were tested under serial then backward recall conditions, and half under backward then serial recall conditions.

Materials and Procedure

Three different sets of 240 words were randomly selected from a pool of 600 common words, mostly two-syllable concrete nouns. Each set was divided arbitrarily into 30 different lists of eight words, which were then divided arbitrarily into three subsets of 10 lists each. Each subset of 10 lists was blocked with respect to modality and order variables such that it was used equally often in the first, second and third test sessions--and therefore under free, serial and backward recall conditions--and equally often in visual and auditory-visual presentation conditions.

Word order within lists and list order within sets was constant. For each set of 240 words a further 24 words were selected, divided into three lists of eight words, and arbitrarily designated as practice lists for each order condition in each group. Practice lists were not rotated across other experimental conditions. In conjunction with each set of 240 words, three-digit numbers were selected from random number tables in sufficient quantity so that there were five such numbers before and after every single word in every list. Similar, additional numbers were selected for the practice lists.

These materials were recorded for presentation in a manner quite similar to that described by J.M. Gardiner et al. (1983). Visual and auditory-visual versions of the lists were recorded simultaneously. The presentation rate for all items was 2s and within a list the sequence of events was as follows. Exactly 2s after the onset of a warning signal (an auditory bleep), the first of five three-digit numbers was recorded and the other three-digit numbers followed at a regular 2s-rate. Numbers were spoken in a steady and even manner, and as single digits within each triad; they were recorded auditorily not visually. Directly after the 2s interval in which the fifth number occurred, a list word was recorded visually from the VDU of a Commodore-PET microcomputer and remained on the screen for the full 2s. In this 2s period the word was also spoken, again in a steady and even manner, and in rhythm with the spoken numbers. For the first and last 100ms of this 2s period, an auditory bleep was recorded to indicate subsequently to subjects the onset and offset of the presentation. The spoken word was recorded on only one of the two videotapes, the auditory-visual one; the audio channel on the other videotape was switched off for this 2s period. In the 2s period directly following that in which a word was recorded, the first of the next batch of five three-digit numbers was recorded--and so forth to the end of the list. After the fifth number following the last word in a list, a horizontal line signalling recall was recorded from the VDU and remained on the screen for 2s; in addition, during this period the word "recall" was spoken and recorded for the auditory-visual group.

During the presentation of each list, the subject worked on one page of a response booklet which had a form along the top for the numbers and another below it for the words. The subject's task with the numbers was to accurately write down each triad in the separate space provided directly upon hearing it. For serial and backward recall, separate, appropriately numbered lines were provided for each word. That is, from top to bottom, the direction in which responses were written, the lines were numbered 1-8 for serial recall and 8-1 for backward recall. The subjects had one practice list in the appropriate condition at the beginning of each of the test sessions. The general instructions given corresponded with those used in previous similar experiments. In particular, subjects were told that speed and accuracy in the copying task were essential and that their performance in that task would be monitored--as indeed it was. They were also told not to associate the words in the list, to think of each word separately and only when it was being shown, and not to rehearse words while copying the numbers. The subjects in the auditory-visual group were told to look up at the screen each time a word was spoken. For free recall, the subjects were told to write the words in any order they liked. For serial and backward recall, they were told not only that they must recall the words in each list in the prescribed order but that no backtracking in recall was allowed; that the same word should not be written more than once, and that they should put a dash to indicate nonrecall. Compliance with instructions and task requirements was carefully monitored during the course of the experiment and on occasion, where necessary, instructions were repeated at an appropriate point in the procedure.

Results

Performance in the distractor task was essentially perfect under all experimental conditions and subjects very rarely needed further instructions from the experimenter during the course of the experiment. The mean recall probabilities are summarized in Figure 1. The figure shows an advantage to the auditory-visual condition over prerecency as well as recency positions. The figure also indicates that this effect is little influenced by the order of recall, whether free, serial or backward, although order of recall does seem to affect the serial position curves generally, especially over recency positions: relative to free recall, the recall of recency items seems enhanced by backward recall and reduced by serial recall.

---

Insert Figure 1 about here

---

The foregoing description of the data is well supported statistically. An ANOVA carried out on the number of words recalled by each subject under each experimental condition revealed a significant effect of modality, $F(1,46) = 11.15$, $MSe = .267$, $p < .005$, a significant effect of order, $F(2,92) = 18.56$, $MSe = .056$, $p < .001$ and a significant effect of serial position, $F(7,322) = 116.71$, $MSe = .030$, $p < .001$. The Order x Serial Position interaction was significant, $F(14,644) = 20.09$, $MSe = .021$, $p < .001$. The Modality x Serial Position interaction was not significant, $F < 1$; and neither was the Modality x Order interaction, $F(2,92) = 1.63$, $MSe = .056$, or the three-way interaction, $F < 1$.

The results of other, supplementary ANOVAS are in general quite consistent with those results and revealed little else of note. For example, separate ANOVAS on the recall scores for each session revealed that within free recall, serial recall and backward recall scores there are significant modality and serial position effects but no modality x serial position interaction--$F < 1$ in each case. The results of another ANOVA, on the combined serial and backward recall scores, showed, however, no effect of order, $F < 1$, confirming that, as Figure 1 indicates, overall recall probability was not affected by recalling in serial or backward order. Leniently-scored data for serial and backward recall (that is, where credit was given even when an item from a list was recalled in the wrong position) were similar to those summarized in Figure 1 and so too were the results of the comparable ANOVA carried out on them. There was no significant effect due to the ordering of serial and backward recall conditions in the second and third sessions, $F < 1$.

## Discussion

The results of this study showed that varying the order of recall produced systematic changes in the level of recency recall. This replicates some observations made in this distractor paradigm by J.M. Gardiner et al. (1983). In both studies recency recall was reduced when recall started from the beginning of the list and in both studies recall order had a similar influence irrespective of presentation modality. However, unlike the previous study, the auditory advantage in the present study was quite unexpectedly obtained over prerecency as well as recency recall.

With the wisdom of hindsight, the finding of an auditory advantage through prerecency as well as recency recall might not be completely unexpected. There are earlier intimations of such an effect in this paradigm. For example, Experiments 3, 4 and to a lesser extent Experiment 5, in J.M. Gardiner and Gregg's (1979) study show clear indications of an auditory advantage earlier in the list, as does at least the backward recall condition of Experiment 3 in Glenberg's (1984) study. Two previous studies by Routh (1970, 1976), mentioned in the Introduction, also show such an effect and those studies are formally quite similar to ours, in that another task, albeit a silent one, was interpolated between the to-be-recalled items. And yet in other similar experiments, the auditory advantage remains confined to recency (J.M. Gardiner & Gregg, 1979, Exps. 1 & 2; Glenberg, 1984, Exps. 1, 2, 4 & 5; Gregg & J.M. Gardiner, 1984; Routh, 1971).

We have failed to detect any systematic covariable in these and other relevant studies that might lead to some insight into this state of affairs. The possibility that perhaps seems most likely, for example, is that double-presentation in the auditory-visual condition gives rise to a through-list effect, and indeed Routh's (1970) study, in which this effect was obtained, used the subject-vocalization procedure. But then the effect was also found with simple auditory versus visual presentation in Routh's (1976) study. Moreover the evidence when subject and experimenter vocalization have been directly compared shows that they are equivalent with respect to the auditory recency advantage, and that subject-vocalization may sometimes lead to *poorer* recall of prerecency items than visual presentation (e.g., Gathercole et al., 1982). We assumed that our auditory-visual procedure is equivalent to subject-vocalization and therefore also to auditory presentation alone, but perhaps this assumption may not be entirely warranted. Subject-vocalization may demand attentional and processing resources not demanded by hearing another person's voice. Were this to influence recall, it is prerecency recall that seems particularly likely to be affected. Be that as it may, all that can confidently be concluded at present is that a through-list auditory advantage can not readily be attributed to double-presentation.

Also, it should be emphasized that from the standpoint of testing echoic memory theory, it is an auditory-visual versus visual comparison that seems ideal, for the theory attributes the auditory recency advantage to the availability of an additional source of information with auditory input.

In the Introduction we pointed out that an auditory advantage that extends through prerecency as well as recency recall seems to occur invariably both with a probe testing procedure and with mixed-list presentation (e.g., Brems, 1983; Murdock, 1968; Murdock & Walker, 1969; M.J. Watkins, 1983). An obvious implication is that when a similar effect unpredictably turns up in other recall procedures, it may do so for precisely the same underlying reasons. That being so, an auditory advantage that extends through the list might be theoretically more fundamental than one which is pinned to recency (J.M. Gardiner, 1983). Performance in free and serial recall tasks is highly susceptible to a variety of subject strategies at encoding and retrieval and also to output interference. It is conceivable that these factors normally obscure a potential auditory advantage, except over recency positions. That the advantage extends through the list with a probe testing procedure is especially important to this speculation, for that procedure minimizes the effects of such strategies and, in addition, it minimizes output interference.

Although the existence of an auditory advantage in the face of same-voice distractors that directly follow the list items is counter to echoic memory theory, it is, of course, possible to entertain the idea that, for some reason, in this distractor paradigm subjects are able to recode echoic information despite the distractors (see J.M. Gardiner & Gregg, 1979). But we can think of no good reason why subjects might be able to do this. What is it about this procedure that might enable subjects to do some nifty recoding that they cannot do in a serial recall suffix experiment, or in other recall tasks where list items are followed by spoken distractors (see, e.g., Broadbent, Vines, & Broadbent, 1978; J.M. Gardiner et al., 1974; Gathercole, Gregg, & J.M. Gardiner, 1983)? Why should the presence of additional distractors throughout list presentation facilitate the recoding of echoic information? The idea is not only post hoc; it seems implausible. For it to be taken seriously, there would first have to be some a priori grounds specifying precisely why and under what conditions recoding of echoic information will sometimes be possible despite subsequent, same-voice distractors. The temporal interval between list and distractor items provides an obvious candidate (Crowder, 1978; Watkins & Todres, 1980), but it does not seem a hopeful one: This interval was of the order of about 1s in the present experiment and in previous similar experiments (e.g., J.M. Gardiner et al., 1983), an interval that is well within the range of those characterizing standard serial and free recall experiments in which distractors occur just after list presentation.

An alternative interpretation of the auditory advantage which we have suggested is that it may reflect enhanced temporal discriminability, or greater temporal distinctiveness, in the auditory mode (J.M. Gardiner, 1983; J.M. Gardiner & Gregg, 1979). This conjecture is not without empirical support and it can accommodate a variety of other findings, including the fact that in the standard suffix paradigm the suffix effect is attenuated when suffix and list items are spoken in a different voice, and including the more recently discovered suffix and modality effects with silent lipread and mouthed presentation (see J.M. Gardiner, 1983, for more discussion). Another, not incompatible, possibility is to think of the auditory modality as an attribute that is somehow more salient than a corresponding visual attribute but which gives rise to a retrieval cue that is, like any other retrieval cue, subject to cue-overload (Glenberg, 1984).

Finally, the experimental results described here provide yet another demonstration of commonality of modality effects across different recall procedures. In the present case this is important not just because it supports the argument for a common explanation, but also because in immediate tests the auditory recency advantage seems enhanced when serial recall is compared with free or backward recall (e.g., Madigan, 1971). However, this enhancement may not be theoretically important. For example, the rather small magnitude of the modality difference in immediate free or backward recall may merely reflect the operation of a ceiling effect, a possibility which is supported by evidence showing that in certain delayed tests, free and backward recall show a large auditory recency advantage (J.M. Gardiner, Gathercole

& Gregg, 1983; see too, O.C. Watkins & M.J. Watkins, 1980). And in the continuous distractor paradigm, J.M. Gardiner et al (1983) found that although recency recall was generally reduced by recalling prerecency items first, the auditory recency advantage was not enhanced. Those results parallel the results of the experiment described here, albeit with an auditory advantage manifest only at recency.

References

Bjork, R.A., & Whitten, W.B. (1974). Recency-sensitive retrieval processes in long-term free recall. Cognitive Psychology, 6, 173-189.

Brems, D.J. (1983). On the relation between short- and long-term modality effects. Unpublished master's thesis, Rice University, Houston.

Broadbent, D.E., Vines, R., & Broadbent, M.H.P. (1978). Recency effects in memory as a function of modality of intervening events. Psychological Research, 40, 5-13.

Crowder, R.G. (1978). Mechanisms of auditory backward masking in the stimulus suffix effect. Psychological Review, 85, 502-524.

Crowder, R.G. (1983). The purity of auditory memory. Philosophical Transactions of the Royal Society London B, 302, 251-265.

Crowder, R.G., & Morton, J. (1969). Precategorical acoustic storage (PAS). Perception & Psychophysics, 5, 365-373.

Gardiner, J.M. (1983). On recency and echoic memory. Philosophical Transactions of the Royal Society London B, 302, 267-282.

Gardiner, J.M., Gardiner, M.M., & Gregg, V.H. (1983). The auditory recency advantage in longer-term free recall is not enhanced by recalling prerecency items first. Memory & Cognition, 11, 616-620.

Gardiner, J.M., Gathercole, S.E., & Gregg, V.H. (1983). Further evidence of interference between lipreading and auditory recency. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 328-333.

Gardiner, J.M. & Gregg, V.H. (1979). When auditory memory is not overwritten. Journal of Verbal Learning and Verbal Behavior, 18, 705-719.

Gardiner, J.M., Thompson, C.P., & Maskarinec, A.S. (1974). Negative recency in initial recall. Journal of Experimental Psychology, 103, 71-78.

Gathercole, S.E., Gardiner, J.M., & Gregg, V.H. (1982) Modality and phonological similarity effects in serial recall: Does one's own voice play a role? Memory & Cognition, 10, 176-180.

Gathercole, S.E., Gregg, V.H., & Gardiner, J.M. (1983). Influences of delayed distraction on the modality effect in free recall. British Journal of Psychology, 74, 223-232.

Gregg, V.H. & Gardiner, J.M. (1984). Phonological similarity and enhanced auditory recency in longer-term free recall. Quarterly Journal of Experimental Psychology, 36A, 13-27.

Glenberg, A.M. (1984). A retrieval account of the long-term modality effect. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 000-000.

Johansson, B.S. & Nilsson, L.-G. (1979). Boundary conditions for the modality effect: Type of material and time of test. Scandinavian Journal of Psychology, 20, 27-35.

Madigan, S.A. (1971). Modality and recall order interactions in short-term memory for serial order. Journal of Experimental Psychology, 87, 294-296.

Murdock, B.B., Jr. (1967). Auditory and visual stores in short-term memory. Acta Psychologica, 27, 316-324.

Murdock, B.B., Jr. (1968). Modality effects in short-term memory: Storage or retrieval. Journal of Experimental Psychology, 77, 79-86.

Murdock, B.B., Jr. & Walker, K.D. (1969). Modality effects in free recall. Journal of Verbal Learning and Verbal Behavior, 8, 665-676.

Nilsson, L.-G. (1979). A functional interpretation of the modality effect. Scandinavian Journal of Psychology, 20, 259-265.

Nilsson, L.-G., Ohlsson, K., & Ronnberg, J. (1980). Processing and storage explanations of the modality effect. Acta Psychologica, 44, 41-50.

Routh, D.A. (1970). 'Trace strength', modality, and the serial position curve in immediate memory. Psychonomic Science, 18, 355-357.

Routh, D.A. (1971). Independence of the modality effect and amount of silent rehearsal in immediate serial recall. Journal of Verbal Learning and Verbal Behavior, 10, 213-218.

Routh, D.A. (1976). An 'across-the-board' modality effect in immediate serial recall. Quarterly Journal of Experimental Psychology, 28, 285-304.

Tzeng, O.J.L. (1973). Positive recency in a delayed free recall. Journal of Verbal Learning and Verbal Behavior, 12, 436-439.

Watkins, M.J. (1972). Locus of the modality effect in free recall. Journal of Verbal Learning and Verbal Behavior, 11, 644-648.

Watkins, M.J. (1983). Short-term memory and the gentle law of auditory ascendancy. Unpublished manuscript.

Watkins, M.J. & Todres, A.K. (1980). Suffix effects manifest and concealed: Further evidence for a 20-second echo. Journal of Verbal Learning and Verbal Behavior, 19, 46-53.

Watkins, M.J. & Watkins, O.C. (1973). The postcategorical status of the modality effect in serial recall. Journal of Experimental Psychology, 99, 226-230.

Watkins, M.J., Watkins, O.C. & Crowder, R.G. (1974). The modality effect in free and serial recall as a function of phonological similarity. Journal of Verbal Learning and Verbal Behavior, 13, 430-447.

Watkins, O.C. & Watkins, M.J. (1980). The modality effect and echoic persistence. Journal of Experimental Psychology: General, 109, 251-278.

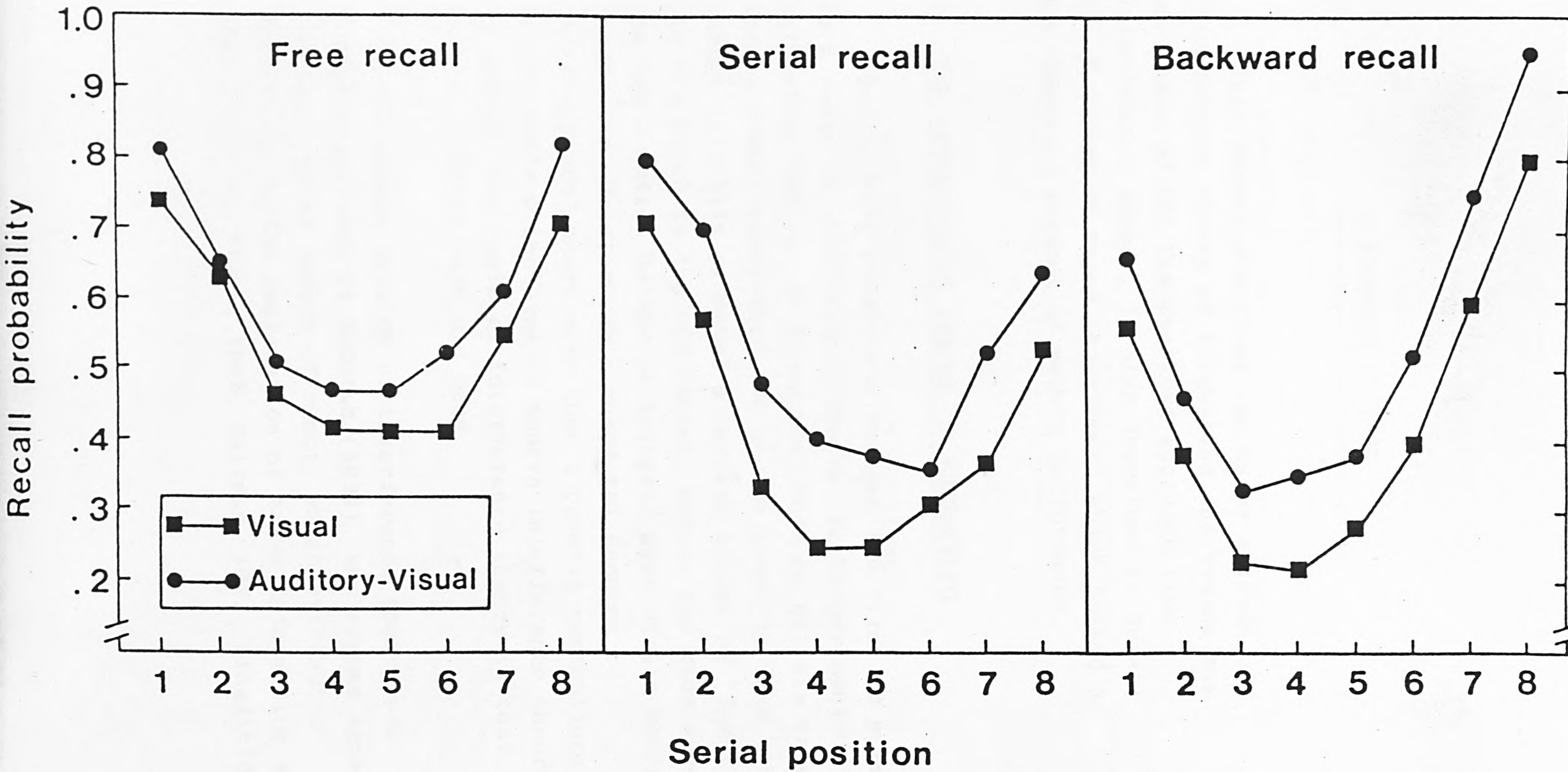
Author Notes

The research described in this article was supported by a grant from the British Social Science Research Council. Margaret M. Gardiner is now at the Human Factors Technology Centre, ITT Europe, Harlow, England.

Figure Caption

Figure 1

Probability of recall as a function of mode of presentation and order-of-recall instruction.


Free recall
Serial recall
Backward recall
Recall probability
1.0
.9
.8
.7
.6
.5
.4
.3
.2
1 2 3 4 5 6 7 8
Serial position
Visual
Auditory-Visual

# A HISTORICAL NOTE

This report describes the major strands of interference theory of forgetting and traces the emergence of the Cue-Overload approach from interference theory. It also describes in detail the findings of major experiments which tested a Cue-Overload account of memory performance.

## I - THE INTERFERENCE THEORY OF FORGETTING

Munsterberg (1889) was amongst the first to point to a possible interference theory, in "experiments" into motor habits. He found that when he put his watch into a pocket other than the one he normally used, this change initially promoted a certain amount of "fumbling" for the watch in the old pocket, before the knowledge of its new location became an integral part of his motor repertoire. Munsterberg's anecdotal account of "interference" bears more than a passing resemblance to the basic principles of modern interference theory of memory. The basis of interference theory is that new habits conflict with old ones.

The modern version of interference theory was formally outlined by McGeoch (1932), who argued against the trace decay theory of memory which attributed forgetting to the degradation of trace information with time (see too, Brown, 1958; Reitman, 1971; Wingfield &

Barnes, 1972). McGeoch commented that time per se accounts for nothing in nature and suggested instead that interference from things learned after acquisition of to-be-remembered items is one of the major causes of forgetting. Experimental evidence for an interference theory of forgetting is extensive. It comes historically from two main sources: proactive inhibition studies and retroactive inhibition studies (e.g., Melton & Irwin, 1940; Postman, Stark & Fraser, 1968; Barnes & Underwood, 1959; Postman & Underwood, 1977; Greenberg & Underwood, 1950). According to interference theory, when something is learned and then its retention is tested, forgetting can be traced both to the learning which took place prior to the original learning (proactive inhibition) or to learning which followed the original learning (retroactive inhibition). It has been argued -- Underwood (1957) -- that in fact perhaps proactive inhibition is the most significant cause of forgetting, and more significant than retroactive inhibition, though both contribute to final recall performance.

One can identify three different strands, or versions, of interference theory: the response competition hypothesis, the unlearning with spontaneous recovery hypothesis, and the list differentiation hypothesis. While all of them attribute forgetting to interference operating on the original trace, they differ in the factors they claim to exist behind such interference. These three strands of interference theory will now be considered in more detail.

a. The response competition hypothesis

Most of the data in support of the response competition hypothesis comes from studies using the retroactive inhibition paradigm, where subjects had to learn two lists of paired-associates which shared the same stimulus terms, but different response terms. The normal forgetting observed when recall of the first list was tested was then attributed to the fact that incompatible responses (first and second list response terms) were attached to the same stimulus term. It is also a postulate of the hypothesis that both first- and second-list response terms were available at the time of recall, but that response competition led to the blocking of responses and to the production of interlist intrusions at recall (see, e.g., McGeoch, 1942). This version of the hypothesis also suggests that the greater the number of second-list intrusions produced during first-list recall, the greater response competition and hence the poorer recall. However, Melton & Irwin (1940) showed that the relation between interlist intrusions and forgetting is not quite this simple. Specifically, they found that whereas response competition (as measured by intrusion rate) increased up to a point and then decreased as a function of the amount of interpolated learning, forgetting continued to increase with increasing amounts of interpolated learning. Melton & Irwin thus suggested that there must be two factors involved in forgetting in the retroactive inhibition paradigm, and that response competition is only one of these. The second factor was initially labelled "Factor X", but then became better known as the "unlearning" of first-list associations due to interpolated learning. Once unlearned, a response is assumed to be no longer available, and so cannot compete for output with any other item. The concept of unlearning will be discussed further in the next section. Before then, however, it is also worth pointing out that there is a second version of the response competition hypothesis, one which conceives of "non-specific" response competition (e.g., Newton & Wickens, 1956). The major difference between the specific version and the non-specific version of the hypothesis centres on the source of the interference. Whereas according to McGeoch competition comes from the fact that two specific response terms are linked to the same stimulus term, the Newton & Wickens version considers instead the set subjects acquire to give second-list responses to the stimulus item. Thus competition at retrieval is of a more generalised kind, less tied to the particular items.

The generalised response competition idea was further elaborated by Postman, Stark & Fraser (1968), who proposed that the entire set of first-list responses, in a retroactive inhibition experiment, is suppressed, or inhibited, as a consequence of response competition during the learning of the second list. Suppression is supposed to be carried out by a "selector mechanism" (see Underwood & Schulz, 1960) which tries to ensure that the subjects produce only the correct responses required at each stage of the retroactive inhibition procedure. It is further assumed that the response selector mechanism has a certain amount of inertia, so that after second-list learning subjects still have a tendency to keep responding with second-list items to the prompts shared by both lists. It is this inability to quickly shift back to first-list responses after learning the interpolated list which is designated "interference" in the Postman, Stark & Fraser (1968) theory, and it represents a "response-set" interference. Hence this theory places the locus of interference at the retrieval stage (when subjects are attempting first-list recall) and of suppression at the learning stage (of second-list items). These two complementary processes, under the control of the selector mechanism, are seen to account for recall in the retroactive inhibition paradigm. The theory has received a fair amount of experimental support (e.g., Postman & Stark, 1969; Postman, Stark & Henschel, 1969; Postman & Warren, 1972), and has fared quite well in studies which compared it with alternative explanations of forgetting, though it has been noted (e.g., Postman & Underwood, 1973) that some of the tests of its validity have been prompted by a poor understanding of its exact postulates. Its appeal may also lie in the fact that it does not specifically deny the possibility that items may still be available, even after interference. In other words, suppression need not entail a loss of the items from memory.

b. The unlearning with spontaneous recovery hypothesis

This second strand of interference theory argues that the process of learning subsequent material, in a retroactive inhibition paradigm, leads to the unlearning, or loss of availability, of trace information relating to the original material.

The notion that forgetting might be due to an unlearning process was originally proposed by Melton & Irwin (1940), in the study mentioned in the previous section, as a two-factor theory of recall based on unlearning and response competition. The original unlearning formulation had strong affinities with S-R language theories or, indeed, any explanations of learning and forgetting in terms of processes similar to those thought to govern conditioning. Unlearning is often interpreted as a memory-analogue of the process of extinction (see, e.g., McGovern, 1964; Postman & Underwood, 1973).

Barnes & Underwood (1959) were among the first to offer some evidence for unlearning under conditions where response competition effects were minimised. The paradigm they used for their research -- known as the MMFR paradigm*-- has since been extensively used to test unlearning interpretations. Barnes & Underwood (1959) demonstrated that the recall of first-list responses, using the MMFR paradigm, depended crucially on the degree to which second-list items had been learned. However, unlearning was not complete in their study, in other words subjects could still recall words from the first list even under conditions where the second list had been studied over 20 trials. (In fact, in retroactive inhibition designs, unlearning very rarely exceeds 50%, Postman & Underwood, 1973; see also Petrich, 1975). Their results were interpreted as showing that in an A-B, A-C MMFR paradigm the nonreinforced production of first-list responses during second-list learning had led to the extinction, or unlearning, of the first list (A-B) associations. The notion of nonreinforcement of first-list

----------------

* - Modified modified free recall.

responses during second-list learning is central to the unlearning argument, in that theorists who have proposed unlearning accounts always seem to consider the elicitation of inappropriate responses (intrusions) a necessary antecedent for unlearning. Intrusions can be of two types: covert (non-generated) or overt (generated). Research has shown that, in fact, both types of intrusion may have the same effects on recall (e.g., Thune & Underwood, 1943; Keppel & Rauch, 1966; see also Underwood, 1945; 1949).

The unlearning hypothesis also suggests that responses that have been unlearned can, nevertheless, recover in time: a process akin to spontaneous recovery in S-R theories. This assumption has been well-supported by the available evidence (e.g. Kamman & Melton, 1967; Postman, Stark & Fraser, 1968; Postman, Stark & Henschel, 1969; Forrester, 1970; Martin & Mackay, 1970; Shulman & Martin, 1970).

The unlearning hypothesis notion that production of inappropriate responses leads to unlearning of initial material through a process similar to extinction has received some support from studies which have manipulated the likelihood of such intrusions being produced. These studies are subsumed under the general title of tests of an "Elicitation hypothesis" (see, e.g., Postman & Underwood, 1973), and have shown that with increases in similarity amongst the items in the lists learned, the greater the number of interlist intrusions and the greater their effect on recall (e.g., Postman, Keppel & Stark, 1965; Friedman & Reynolds, 1967; see also Birnbaum, 1968).

This facet of the unlearning hypothesis received further attention from studies which dealt specifically with the idea that loss of information about list <u>membership</u>, resulting in greater numbers of potential intrusions, might be the process responsible for forgetting in paradigms involving multi-list learning. The next section will consider these studies in more detail.

It has been suggested (e.g., Tulving & Madigan, 1970) that the original view of unlearning as the extinction of learned associations has now been abandoned in favour of an interpretation of unlearning that follows more closely the mechanisms described for the response-set <u>suppression</u> part of the response competition hypothesis. This reinterpretation of unlearning would seem to follow easily from Postman, Stark & Fraser's (1968) and Postman & Stark's (1969) description of the response competition notion, where suppression and unlearning seemed to be used interchangeably (see too Postman & Underwood, 1973), and it helps to bring together two different strands of interference theory, a fact that can only be to its advantage.

### c. <u>The list differentiation hypothesis</u>

The idea that forgetting in multi-list paradigms could be due to mounting confusion about which items belong in which lists was initially proposed in studies by Underwood and colleagues (e.g., Underwood & Ekstrand, 1966; 1967; Underwood & Freund, 1968). These showed that providing strong temporal cues to differentiate between lists reduced the amount of interference observed in proactive inhibition paradigms.

Subsequent research concentrated on determining which factors might contribute to loss of discriminability, above and beyond the loss of information about the temporal separation of the lists learned (see, e.g., Winograd, 1968; Hintzman & Waters, 1969; 1970; McCrystal, 1970). These studies showed that the relative frequency of items within a list helps list discrimination, in the sense that subjects find it easier to discriminate between two lists when one of them contains many repeated presentations of the same items, than when every item is unique. Further, the earlier in the lists those repetitions occur, the easier it is to discriminate between lists. Similarly, repeating whole-list presentations, in multi-trial list learning, also improves discriminability.

Explanations of forgetting simply in terms of loss of list differentiation are, however, not entirely consistent with the available data. It has been shown that heavy retroactive inhibition losses may still occur in the absence of large deficits in the ability to discriminate between lists (Barnes & Underwood, 1959, using the MMFR procedure). Postman & Underwood (1973; see also Martin, 1971) have also argued that it may not be useful to try to use the list differentiation idea in situations other than those which led to its development. It may not be wise to try to explain interference effects of the kind observed in retroactive and proactive inhibition paradigms simply in terms of list differentiation.

The foregoing discussion outlined three different versions of the interference theory of forgetting. It may be noted that only the first two of these versions -- the response competition hypothesis and the unlearning with spontaneous recovery hypothesis -- did not necessarily entail actual loss of trace information. The third version would require such an assumption, since loss of information about the temporal separation of lists, for example, must necessarily reflect loss of information from the trace. This distinction is important in view of later developments in theoretical approach, proposed amongst others by Tulving and colleagues. These suggested that, rather than reflecting interference with stored information, forgetting reflects instead the failure to reinstate an appropriate retrieval environment. The next section of this report now deals with these later theories, subsumed under the collective title of cue-dependent forgetting ideas.

## II - CUE-DEPENDENT FORGETTING THEORIES

Tulving & Madigan (1970) first drew the distinction between trace- and cue-dependent forgetting, in an article in which they strongly criticize the basic assumptions of interference theory. Their view of this theory, at the time they were writing, included mainly the response-set suppression, or generalized response competition views described earlier, with the unlearning view subsumed under the response-set suppression hypothesis. Tulving & Madigan's argument centred on the simple finding that certain types of forgetting, or apparent forgetting, cannot easily be attributed to unlearning, displacement or decay of a memory trace, nor to competition, but must instead be interpreted in terms of the presence or absence of appropriate retrieval cues. In a later paper, Tulving (1974) expanded this view of forgetting as a loss of retrieval, rather than trace information, and described three sources of evidence for his new theory: single trial free recall experiments, multi-list retroactive inhibition experiments and subject-generated recognition task experiments. We will now consider each of these sources of evidence in turn.

a. Single-trial free recall

Consider the situation where subjects learn one long list of words, assorted members of a variety of categories presented in no specific order, and are then asked to recall them. Their failure to recall all the items in the list is interpreted by trace-dependent forgetting theories (e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965; Glanzer, 1972) as loss of information from a short-term store, before it had had a chance to be transferred to a more permanent and spacious long-term store. However, providing the subjects, after this recall test, with the names of the categories represented in the list leads to a dramatic increase in the number of words recalled. This was the finding of an experiment by Tulving & Pearlstone (1966), and it was additionally used to demonstrate the distinction between the information which is available in memory and that which is accessible to recall. Clearly, any theory that attributes forgetting to loss of trace information would have trouble accounting, unaided, for the data obtained with cued recall. However, the results of this experiment are readily interpreted in terms of cue-dependent forgetting. What the cues did was to provide additional retrieval information to gain access to the available contents of memory.

b. Retroactive interference

Tulving & Psotka (1971) demonstrated that the effects of retroactive inhibition in a multi-list retroactive inhibition paradigm could be eliminated by providing at recall additional recall cues, in the shape of the names of the categories from which items had been drawn. Interference theory would have trouble interpreting the loss of such an effect: Postman (e.g., Postman & Keppel, 1967; Postman & Underwood, 1973; see also Keppel, 1968) attributes the losses in recall due to retroactive inhibition to the unlearning of both specific (inter-item) and general (context-item) associations in the lists. The concept of unlearning in turn implies the loss of some specific associations. Postman & Stark (1969) explain forgetting in this paradigm as the result of response competition among the response sets for the specific lists and to the inertia of a response selector mechanism which is supposed to allow the subjects to switch from one list context to another, as required by the task. Though this theory does not necessarily imply that the traces stored are lost as a consequence of response competition, it assumes that recall is impossible because of it, rather than because the subject lacks appropriate retrieval information. Faced with the elimination of the retroactive inhibition effect with cueing, the unlearning version of interference theory would have to explain why the presentation of specific retrieval cues can restore recall, if forgetting is a consequence of unlearning, and the response-set suppression hypothesis has to explain how retrieval cues reduce response competition or overcome the inertia of the selector mechanism. Neither can do so without postulating additional processes.

However, it could be argued that both the effects of response competition and the operation of the selector mechanism involved in the latter theory could depend on the information contained in the retrieval environment, such that cues will reduce one and facilitate the other. Accepting this view, however, reduces the differences between response-set suppression and cue-dependent forgetting to such an extent that the former becomes a special case of the latter, in postulating that forgetting is a cue-dependent phenomenon.

The pattern of results obtained by Tulving & Psotka (1971) is readily interpretable in terms of cue-dependent theory. This would argue that the retroactive inhibition effects obtained under uncued conditions reflect changes in the retrieval information available at the time of recall rather than loss of information from the trace. These changes are brought about by the activities interpolated between learning the words for recall and their test (e.g. interpolated learning and recall of other lists). The category cues provided at a later stage in the experiment simply help reinstate the recall environment to its original state, and enable the subjects to gain access to the information available in memory.

c. Failure of recognition

Many theorists consider the problem of retrieving information in a recognition test to be simpler than in a recall test. One such widely accepted theory argues that recognition involves a simple decision as to whether the item was encountered in a list; recall is assumed to involve the generation of a number of candidates, followed by some decision as to their appropriateness (e.g., Bahrick, 1970; Kintsch, 1970; McCormack, 1972). A dual-process theory such as this assumes that failing to recognise a word must implicitly reflect its absence from the memory store -- the absence of its memory trace. Such a theory would never conceive of a situation where recall could be higher than recognition: if the trace is not available (cannot be recognised) it should not be amenable to generation and identification either.

Tulving & Thomson (1973) provide evidence of just such a situation. Subjects were asked to memorise the second term of a series of 24 pairs of words and, after presentation, were given a series of 24 strong associates of the to-be-remembered items and asked to free-associate to them. This generation task produced about 70% of the to-be-remembered words. Subjects were then asked to select from among the words generated every word they recognised as a target item. About 24% of the generated items were identified. Finally, the subjects were given the 24 paired-associates from the first list and asked to write down as many of their paired target words as possible. 63% of the target words were produced. The experiment clearly shows that there are situations where recall can greatly exceed recognition, for the same pool of target words.

Though these results are inconsistent with most theories of recognition memory, since they usually see failure to recognise as an overt demonstration of a degraded trace, they are nevertheless highly consistent with a cue-dependent forgetting account. According to this approach, recognition failure, like recall failure, reflects an absence of the appropriate retrieval information. The theory simply assumes that subject-generated copies of target words are not as good as paired-associates when it comes to reinstating the appropriate retrieval environment.

The cue-dependent forgetting approach offers some advantages over its rival trace-dependent theories (see, e.g., Tulving, 1974). As the above three examples illustrate, it offers a more parsimonious account of diverse instances of forgetting, since one very simple set of basic assumptions can accommodate a variety of findings for which the trace approach has to postulate several different theories. It is also easier to verify experimentally whether cue information has been lost (by providing cues and observing recall) than it is to verify whether a trace has been lost. Cue-dependent forgetting also makes it a meaningful task to search for situations in which recall is higher than recognition, and it makes it possible to understand how this phenomenon comes about. The same cannot be said of trace-dependent theories.

It could be argued, however, that cue-dependent forgetting can offer all these advantages only at the cost of a certain implicit circularity: the appropriate retrieval environment is said to have been reinstated when recall goes up; recall goes up when the appropriate retrieval environment is present. However, its value in bringing together a variety of different paradigms and in pointing to new directions of investigation should not be ignored. The search for more data to test the theory will, hopefully, also help to reduce its circularity. The cue-dependent approach also cannot be easily refuted, since it can "wriggle out" of potentially embarrassing situations -- for instance if cueing does not improve recall -- by simply downgrading or denying the retrieval-usefulness of additional cues provided. As such, it is best seen as a general orienting principle, to guide theoretical thinking, than as a fully-fledged formal theoretical statement. In this general, framework for research format, cue-dependent forgetting does not, for example, deny that traces may change over time. It simply argues that in order to understand forgetting it is not necessary to postulate trace decay or degradation, since loss of retrieval information can much more fully account for the available data. Similarly, it does not claim that, given the right cues, any memory can potentially be retrieved, no matter how old or weak, and readily acknowledges that it is quite possible that certain kinds of retrieval information, or cognitive environment, may never be reinstated after learning.

The cue-dependent forgetting approach outlined above received more detailed theoretical treatment in a series of papers by Watkins & Watkins (e.g., 1975; 1976; Watkins, 1979; see also Mueller & Watkins, 1977; Todres & Watkins, 1981) under the guise of Cue-Overload theory or Cue-Overload Principle. Watkins & Watkins sought to answer some of the questions raised by the Tulving (1974) exposition, in particular those concerned with the nature of the interaction between stored information and retrieval cues. The next section will now consider this approach in more detail.

## III - THE CUE-OVERLOAD THEORY OF MEMORY

Like Tulving's general orienting principle described above, Cue-Overload theory is also not proposed as a fully-fledged theory, though its specification is, at times, much more stringent than Tulving's. In its simplest form, Cue-Overload theory proposes that recall is mediated by cues and that these cues become overloaded as they come to refer to more and more items. Overload on a cue, in turn, determines its ability to retrieve the items to which it relates. The theory does not specify exactly how a cue is used to retrieve the items it subsumes, nor does it actually have anything to say about the manner in which items come to be associated with a cue. It would seem clear, however, that this aspect of it could be elaborated on by a theory like encoding specificity. A later section in this report will consider this question in more detail. Before that, we shall describe the original Watkins & Watkins experiments which led to the formulation of Cue-Overload theory.

### a. Experimental background to Cue-Overload theory

Watkins & Watkins (1975) report two studies which investigated a Cue-Overload interpretation of the phenomenon of proactive inhibition build-up, and in which the approach is contrasted with an encoding view of proactive inhibition build-up (e.g., Fozard & Waugh, 1969; Petrusic & Dillon, 1972; Dillon, 1973). They used a technique previously used by Turvey, Mosher & Katz (1971) and by Carey (1973), in which subjects are given an additional final test of all the items presented in the course of the experiment. One difference was, however, that in the Watkins & Watkins experiments the test was a recall one, whereas in the two earlier studies recognition tests had been given. The rationale behind the experiment was that, if recall deficits in the proactive inhibition paradigm are due to progressively poorer encoding of items into memory, this should also be reflected in the final test, in that the serial position of the various items used in the experiment should be associated with systematic differences in recall level. If, however, the effect is simply one attributable to the number of items subsumed under particular cues, then in a suitably designed experiment it should be possible to demonstrate an independence between final recall performance and list presentation order.

Experiment 1 in the Watkins and Watkins study presented subjects with a long series of Brown-Peterson trials, with the items in each run of 3 lists drawn from the same category. Unlike the original Brown-Peterson technique, recall for the items presented was tested very occasionally during presentation -- only often enough to demonstrate within-category proactive inhibition build-up. After all lists had been presented, a final recall test was given. In this, subjects were cued with the names of the categories presented in the experiment. The results of interest concerned both the initial recall data, for those categories tested in the course of the presentation, and the final recall data, for all categories. The results of the experiment showed that on initial testing proactive inhibition was obtained, in that recall declined from the first to the third list within each category. Final recall data showed no decline in recall accross list position, for both tested and untested categories. The results were therefore in good accord with the Cue-Overload position.

Experiment 2 used a similar procedure to Experiment 1, but in addition also varied the number of items from each category that were presented to the subject. The prediction was that, if recall depended on cue-overload brought about by the number of items subsumed by a category cue, then increasing the number of those items should also increase overload and decrease recall. The results of this experiment replicated the essence of the results of the previous study and, in addition, showed that final recall probability was an inverse function of the number of items from a category that had been presented, but not of their position in the presentation order. The data from both experiments were interpreted as presenting a better explanation of proactive inhibition build-up in terms of Cue-Overload than differential encoding.

Watkins & Watkins (1976) then turned to the release from retroactive inhibition paradigm, and suggested a stripped-down version of the traditional method of retroaction, which they labelled the Method of Interpolated Attributes. This procedure was seen to be particularly apt for testing a Cue-Overload interpretation, since it guaranteed encoding equivalence across all the lists used to test recall. Recall impairments could thus be confidently attributed to the retrieval stage. Watkins & Watkins report the results of 12 experiments which used the Method of Interpolated Attributes to test for release from retroactive inhibition across a variety of materials (cf. similar studies by Wickens, 1970). Their results show that for rhyme classes, taxonomic categories, letters vs. digits, place of articulation and words vs. numbers, the Method of Interpolated Attributes could demonstrate release from retroactive inhibition. With the remaining 7 attributes used in the study, release was not obtained. Watkins & Watkins suggest that, since they provide more information about which attributes are used as retrieval cues in a retroactive inhibition paradigm, the findings of their study may eventually help to develop a stricter format for Cue-Overload theory.

A study by Mueller & Watkins (1977) shifted paradigms yet again, and looked instead at the part-set cueing paradigm of Slamecka (1968) and of Brown (1968). Mueller and Watkins provide an interpretation of the part-set cueing effect in terms of the greater load on the retrieval cue when additional words are provided at recall. This interpretation of the effect is discussed in more detail in Chapters 1 and 4 of this thesis (pp. 21-22 and 144-146). Mueller and Watkins had two major aims in mind with the 4 experiments they did: first, they wanted to test the prediction, derived from Cue-Overload, that the part-set cueing effect, if it was due to cue-overload, ought to apply to virtually any effective mode of categorisation (i.e. to any potential cue); and, second, that with categorised lists inhibition through part-set cueing requires that the list words given at recall should be instances of the category under test (i.e., potentially associated with the cue) and not merely members of the study list. Experiment 1 tested this latter prediction, by providing subjects with words, at recall, that were either related or unrelated to the category being tested, and comparing recall in this condition with recall in a control condition, where no lists words were presented. Their results demonstrated that only category members consistent with the category being tested actually inhibited recall when provided as cues for recall. In other words, no inhibition with part-set cueing was obtained for unrelated items. Experiments 2-4 then looked at the effects of part-set cueing across three conceptual categories: rhyme sets, intuitive sets formed with the same sorting technique used by Mandler (1967; Mandler & Pearlstone, 1966) and arbitrary sets formed through a variation of the A-B, A-C paradigm. All three sets of materials produced part-set cueing inhibitory effects which were interpreted as supporting a Cue-Overload interpretation. It was noted, too, that Experiment 4 was very similar to an experiment conducted by Postman, Stark & Fraser (1968), which led to the development of the response-set suppression idea described in the early part of this report. Mueller & Watkins (1977) argue that the response-set suppression idea cannot fully account for the part-set cueing effect obtained, since the experiment had been so designed that the function of the "selector mechanism" suggested by Postman et al. would have been largely precluded. They suggest that the Cue-Overload position is, if anything, closer to the traditional response competition idea, a parallel that was also noted in this report in connection with the description of Tulving's cue-dependent forgetting theory.

Todres & Watkins (1981) extended the Mueller & Watkins (1977) study by demonstrating that part-set cueing effects could be obtained when memory was tested using a recognition rather than a recall test. The study includes 4 experiments. Experiment 1 investigated the effects of intra- and extra-list cueing on recognition, using lists made up of members of different taxonomic categories. The results show that part-list cueing does slightly impair recognition, particularly when cues are members of the categories used but were not themselves included in the list. Experiment 2 replicated the extralist cueing effect obtained in Experiment 1. The third experiment then tried to obtain further evidence on the effects of intralist cueing, which had not been reliable in the first experiment, and replicated the initial study without the extralist condition. No recognition inhibition was obtained with intralist cueing, a finding that was attributed to the total reinstatement of the study list context with this type of cue. In order to test this hypothesis, Todres & Watkins then changed their original experimental lists from a blocked-by-category format to a random one, and repeated Experiment 3. The results demonstrated a significant effect of intralist cueing on recognition with these new list formats.

In their concluding remarks, Todres & Watkins point out that their suggested explanation for the failure to obtain a part-list cueing effect with intralist cues, in Experiment 1 and 3 (that it was due to the total reinstatement of the learning context for the lists used) is not new. In effect, it is contained in essentially the same form in Tulving & Thomson's (1973) encoding specificity principle. Previous studies had also shown that recognition performance is better if word order is preserved at learning and test (Jacoby, 1972; Light & Schurr, 1973), and that it is also improved if paired associates used at learning are re-presented at test (Thomson, 1972; Tulving & Thomson, 1971). This point will be taken up again later in this report.

Finally, an experiment by Parkin (1980) sheds some more light on the relationship between encoding conditions and Cue-Overload. Parkin used a levels-of-processing paradigm, in which he varied the type of processing carried out on the materials interpolated between memory list and recall in the Watkins & Watkins (1976) Method of Interpolated Attributes procedure. Specifically, the interpolated items could either be processed at a semantic level (determining the subcategory membership of each item) or at a non-semantic level (reporting the number of syllables contained in each word). The rationale behind the experiment was that if semantic and nonsemantic tasks do represent qualitatively different forms of processing, then only the semantic task should lead to cue-overload for interpolated items belonging to the same category as target items. This is because, presumably, taxonomic category membership information is located at a semantic level. The results of the experiment supported the original hypothesis. Categorically similar interpolated material only led to retroactive inhibition (and hence only overloaded the category cue) when it was processed semantically. Parkin concluded that the results of this experiment support a locus for Cue-Overload at the <u>encoding</u> stage (rather than at the retrieval stage, as argued by Watkins & Watkins). He bases this conclusion on the finding that there were very few intrusions in the condition showing the greatest recall decrement, so that an encoding deficit would seem a more likely explanation. His suggestion is that as more items come to be associated with the same cue, their encoding becomes less distinct, they lose differentiation and become more difficult to discriminate. However, it would be equally plausible to suggest that loss of discrimination should lead to more, not less intrusions (by virtue of subjects adopting a guessing strategy). To argue for an encoding explanation on the basis of lack of intrusions is also to ignore the fact that overt intrusions, as mentioned earlier in this report, are not the only intrusions produced: covert intrusions have an at least equal effect on retroactive inhibition build-up.

b. Cue-Overload and the Encoding Specificity Principle

It was mentioned in connection with the description of the Cue-Overload theory, and of Todres & Watkins (1981) experiments, that it would seem logical that the Tulving & Thomson (1973) Encoding Specificity Principle should complement the Cue-Overload approach, to the extent that the former specifies under which conditions a retrieval cue is likely to be associated with certain items. The association of encoding specificity and cue-overload has received little formal acknowledgement in the literature, but considering the wealth of evidence supporting encoding specificity (see Tulving, 1983), it is reasonable to assume that such an association was always implied and would be generally acceptable to cue-dependent forgetting theorists. It is also of interest to note that even theorists normally associated with encoding views of memory now accept that encoding and retrieval are necessarily interdependent in the manner encapsulated by the encoding specificity principle (see, e.g., Fisher & Craik, 1977; Craik, 1983).

## CONCLUSION

This historical note covered in some detail the main versions of interference theory, and traced the emergence of cue-dependent forgetting theories as a reaction to the original interference theory notions. However, it should be noted that to view both sets of theories in this way is not to deny that some of their postulates may show some parallels, so that in some instances, some might consider the differences between these theories to be more a matter of semantics than of essence. Where appropriate, these points of similarity were noted. It was also pointed out that cue-dependent forgetting theories appear to have some advantages over interference accounts. An additional advantage is that they also seem to fit better into the currently predominant information processing approach to memory theory.

# REFERENCES

Atkinson, R.C. & Shiffrin, R.M. (1968) Human memory: A proposed system and its control processes. In: K.W. Spence & J.T. Spence (Eds.) The Psychology of Learning and Motivation, vol. 2. New York: Academic Press.

Bahrick, H.P. (1970) A two-phase model for prompted recall. Psychological Review, 77, 215-222.

Barnes, J.M. & Underwood, B.J. (1959) "Fate" of first-list associations in transfer theory. Journal of Experimental Psychology, 58, 97-105.

Birnbaum, I.M. (1968) Response-class similarity and first-list recall with mixed and unmixed transfer designs. Journal of Experimental Psychology, 77, 542-546.

Brown, J. (1958) Some tests of the decay theory of immediate memory. Quarterly Journal of Experimental Psychology, 10, 12-21.

Brown, J. (1968) Reciprocal facilitation and impairment in free recall. Psychonomic Science, 10, 41-42.

Carey, S.T. (1973) Delayed recognition testing, incidental learning and proactive inhibition release. Journal of Experimental Psychology, 100, 361-367.

Craik, F.I.M. (1983) On the transfer of information from temporary to permanent memory. Philosophical Transactions of the Royal Society London, B302, 341-359.

Dillon, R.F. (1973) Locus of proactive interference effects in short-term memory. Journal of Experimental Psychology, 99, 75-81.

Fisher, R.P. & Craik, F.I.M. (1977) The interaction between encoding and retrieval processes in cued recall. Journal of Experimental Psychology: Human Learning and Memory, 3, 701-711.

Forrester, W.E. (1970) Retroactive inhibition and spontaneous recovery in the A-B D-C paradigm. Journal of Verbal Learning and Verbal Behavior, 9, 525-528.

Fozard, J.L. & Waugh, N.C. (1969) Proactive inhibition of prompted items. Psychonomic Science, 17, 67-68.

Friedman, M.J. & Reynolds, J.H. (1965) Retroactive inhibition as a function of response-class similarity. Journal of Experimental Psychology, 74, 351-355.

Glanzer, M. (1972) Storage mechanisms in recall. In: G.H. Bower (Ed.) The Psychology of Learning and Motivation, vol. 5, New York: Academic Press.

Greenberg, R. & Underwood, B.J. (1950) Retention as a function of stage of practice. Journal of Experimental Psychology, 40, 452-457.

Hintzman, D.L. & Waters, R.M. (1969) Interlist and retention intervals in list discrimination. Psychonomic Science, 17, 357-358.

Hintzman, D.L. & Waters, R.M. (1970) Recency and frequency as factors in list discrimination. Journal of Verbal Learning and Verbal Behavior, 9, 218-221.

Jacoby, L.L. (1972) Effects of organisation on recognition memory, Journal of Experimental Psychology, 92, 325-331.

Kamman, R. & Melton, A.W. (1967) Absolute recovery of first-list responses from unlearning during 26 minutes filled with an easy or difficult information processing task. Proceedings of the 75th Annual Convention, American Psychological Association, 2, 63-64.

Keppel, G. (1968) Retroactive and proactive inhibition. In: T.R. Dixon & D.L. Horton (Eds.) Verbal Behavior and General Behavior Theory. Englewood Cliffs: Prentice-Hall.

Keppel, G. & Rauch, D.S. (1966) Unlearning as a function of second-list error intrusions. Journal of Verbal Learning and Verbal Behavior, 5, 50-58.

Kintsch, W. (1970) Models for free recall and recognition. In: D.A. Norman (Ed.) Models of Human Memory, New York: Academic Press.

Light, L.L. & Schurr, S.C. (1973) Context effects in recognition memory: Item order and unitization. Journal of Experimental Psychology, 100, 135-140.

Mandler, G. (1967) Organization and Memory. In: K.W. Spence & J.T. Spence (Eds.) The Psychology of Learning and Motivation, vol. 1, New York: Academic Press.

Mandler, G. & Pearlstone, Z. (1966) Free and constrained concept learning and subsequent recall. Journal of Verbal Learning and Verbal Behavior, 5, 126-131.

Martin, E. (1971) Verbal learning theory and independent retrieval phenomena. Psychological Review, 78, 314-332.

Martin, E. & Mackay, S.A. (1970) A test of the list differentiation hypothesis. American Journal of Psychology, 83, 311-321.

McCormack, P.D. (1972) Recognition memory: How complex a retrieval system? Canadian Journal of Psychology, 26, 19-41.

McCrystal, T.J. (1970) List differentiation as a function of time and test order. Journal of Experimental Psychology, 83, 220-223.

McGeoch, J.A. (1932) Forgetting and the law of disuse. Psychological Review, 39, 352-370.

McGeoch, J.A. (1942) The Psychology of Human Learning, New York: Longmans.

McGovern, J.B. (1964) Extinction of associations in four transfer paradigms. Psychological Monographs, 78, (16, No. 593).

Melton, A.W. & Irwin, J.M. (1940) The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. American Journal of Psychology, 53, 173-203.

Mueller, C.W. & Watkins, M.J. (1977) Inhibition from part-set cueing: A cue-overload interpretation. Journal of Verbal Learning and Verbal Behavior, 16, 699-709.

Munsterberg, H. (1889) -- as cited in J.A. McGeoch (1942) The Psychology of Human Learning. New York: Longmans.

Newton, J.M. & Wickens, D.D. (1956) Retroactive inhibition as a function of the temporal position of interpolated learning. Journal of Experimental Psychology, 51, 149-154.

Parkin, A.J. (1980) Levels of processing and the cue overload principle. Quarterly Journal of Experimental Psychology, 32, 427-434.

Petrich, J.A. (1975) Storage and retrieval processes in unlearning. Memory and Cognition, 3, 63-74.

Petrusic, W.M. & Dillon, R.F. (1972) Proactive interference in short-term recognition and recall memory. Journal of Experimental Psychology, 95, 412-418.

Postman, L. & Keppel, G. (1967) Retroactive inhibition in free recall. Journal of Experimental Psychology, 74, 203-211.

Postman, L., Keppel, G. & Stark, K. (1965) Unlearning as a function of the relationship between successive response classes. Journal of Experimental Psychology, 69, 111-118.

Postman, L. & Stark, K. (1969) Role of response availability in transfer and interference. Journal of Experimental Psychology, 79, 168-177.

Postman, L., Stark, K. & Fraser, J. (1968) Temporal changes in interference, Journal of Verbal Learning and Verbal Behavior, 7, 672-694.

Postman, L., Stark, K. & Henschel, A. (1969) Conditions of recovery after unlearning. Journal of Experimental Psychology, 81 (1, Pt. 2)

Postman, L. & Underwood, B.J. (1973) Critical issues in interference theory. Memory and Cognition, 1, 19-40.

Postman, L. & Warren, L. (1972) Temporal changes in interference under different paradigms of transfer. Journal of Verbal Learning and Verbal Behavior, 11, 120-128.

Reitman, J.S. (1971) Mechanisms of forgetting in short-term memory. Cognitive Psychology, 2, 185-195.

Schulman, H.G. & Martin, E. (1970) Effects of response-set similarity on unlearning and spontaneous recovery. Journal of Experimental Psychology, 86, 230-235.

Thomson, D.M. (1972) Context effects in recognition memory. Journal of Verbal Learning and Verbal Behavior, 11, 497-511.

Thune, L.E. & Underwood, B.J. (1943) Retroactive inhibition as a function of degree of interpolated learning. Journal of Experimental Psychology, 32, 185-200.

Todres, A.K. & Watkins, M.J. (1981) A part-set cueing effect in recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 7, 91-99.

Tulving, E. (1974) Cue-dependent forgetting. American Scientist, 62, 74-82.

Tulving, E. (1983) Elements of episodic memory. New York: Oxford University Press.

Tulving, E. & Madigan, S.A. (1970) Memory and verbal learning. Annual Review of Psychology, 21, 437-484.

Tulving, E. & Pearlstone, Z. (1966) Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 5, 381-391.

Tulving, E. & Psotka, J. (1971) Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. Journal of Experimental Psychology, 87, 1-8.

Tulving, E. & Thomson, D.M. (1971) Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 87, 116-124.

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

Turvey, M.T., Mosher, D.L. & Katz, L. (1971) Subsequent recognition of items subjected to proactive interference in short-term memory. Psychonomic Science, 25, 365-367.

Underwood, B.J. (1945) The effect of successive interpolations on retroactive and proactive inhibition. Psychological Monographs, 59 (3. No. 273).

Underwood, B.J. (1949) Proactive inhibition as a function of time and degree of prior learning. Journal of Experimental Psychology, 39, 24-34.

Underwood, B.J. (1957) Interference and forgetting. Psychological Review, 64, 49-60.

Underwood, B.J. & Ekstrand, B.R. (1966) An analysis of some shortcomings in the interference theory of forgetting. Psychological Review, 73, 540-549.

Underwood, B.J. & Ekstrand, B.R. (1967) Studies of distributed practice XXIV: Differentiation and proactive inhibition. Journal of Experimental Psychology, 74, 574-580.

Underwood, B.J. & Freund, J.S. (1968) Effect of temporal separation of two tasks on proactive inhibition. Journal of Experimental Psychology, 78, 50-54.

Underwood, B.J. & Schulz, R.W. (1960) Meaningfulness and verbal learning. Philadelphia: Lippincott.

Watkins, M.J. (1979) Engrams as cuegrams and forgetting as cue-overload: A cueing approach to the structure of memory. In: C.R. Puff (Ed.) Memory organization and structure. New York: Academic Press.

Watkins, M.J. & Watkins, O.C. (1976) Cue-Overload theory and the Method of Interpolated Attributes. Bulletin of the Psychonomic Society, 7, 289-291.

Watkins, O.C. & Watkins, M.J. (1975) Build-up of proactive inhibition as a cue-overload effect. Journal of Experimental Psychology: Human Learning and Performance, 104, 442-452.

Waugh, N.C. & Norman, D.A. (1965) Primary memory. Psychological Review, 72, 89-104.

Wickens, D.D. (1970) Encoding categories of words: An empirical approach to meaning. Psychological Review 77, 1-15.

Wingfield, A. & Byrnes, D.L. (1972) Decay of information in short-term memory. Science, 176, 690-692.

Winograd, E. (1968) List differentiation as a function of frequency and retention interval. Journal of Experimental Psychology Monograph, 76 (2, Pt. 2).