THE IMPLICIT LEARNING OF STRUCTURE: ANALOGIC AND ABSTRACTIVE STRATEGIES IN ARTIFICIAL GRAMMAR LEARNING

By

JOHN R. VOKEY, B.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

IMPLICIT LEARNING

DOCTOR OF PHILOSOPHY (1983)
(Psychology)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: The Implicit Learning of Structure: Analogic and Abstractive Strategies in Artificial Grammar Learning

AUTHOR: John R. Vokey, B.A., (University of Winnipeg)

SUPERVISOR: Professor L. R. Brooks

NUMBER OF PAGES: X, 282, (37)

ABSTRACT

For years, psychologists and psycholinguists have described human learning and memory in terms of abstract knowledge structures such as grammars, orthographies, prototypes, scripts and rules. These abstractions often have been thought to be implicit, and acquired through some implicit process. Recently, some theorists have suggested that at least some of the behaviour that has been attributed to implicit, abstract knowledge may actually arise from people's memory for individual cases. In this thesis, this issue is investigated by addressing the recent claims of Arthur Reber and his associates for the rapid, implicit abstraction of artificial grammars. In a series of six experiments, subjects were trained with a sample of items generated from an artificial grammar, and then given a surprise classification test in which they were asked to sort new items with respect to grammatical status. In each experiment, it was demonstrated that what had previously been attributed to implicit abstraction of grammaticality was actually a function of the similarity between transfer items and specific training experiences. It was also shown that variations in how subjects are asked to learn the training items affects their sensitivity to the specific similarity of the transfer items rather than, as Reber and his associates had suggested, their sensitivity to the grammaticality of the items. These results are interpreted in terms of "breadth of transfer" – the degree to which subjects may generalize around their memory for specific experiences. Other results suggest that breadth of transfer is a function of the specific encoding of events, and that variables that affect the encoded similarity between events affect the likelihood of item-specific transfer. All of the results are interpreted in

terms of a model of structural learning that suggests that subjects' knowledge of complex domains is represented in memory as multiple traces, and that the close relationship found in these experiments between original item learning and subsequent tests of item recognition and categorical transfer is a function of the memorial distribution of these traces. Finally, the results of post-testing questionnaires that asked the subjects to attribute the bases of their transfer decisions indicated that the subjects' knowledge in these tasks is less implicit than previously suggested. In affirmation of the objective analyses of their behaviour, subjects attributed their transfer decisions primarily to the similarity between training and transfer items. The results are discussed in terms of a general framework that suggests that the knowledge underlying people's performance in complex domains consists of a mixture of episodic memory for individual instances and limited, explicit abstractions.

ACKNOWLEDGEMENTS

Memory for individual instances may play a large role in people's performance in many domains. This is certainly true of the work reported in this thesis. Without the many individual acts of kindness, the innumerable instances of support, and the specific exemplars of scholarship and of what a scientist should be that surrounded me during my time at McMaster University, this work might never have been completed. The memory of these will be with me and hopefully directing me always. I especially wish to express my gratitude to my thesis supervisor, Dr. Lee Brooks, who assisted and supported me not only in his advisory capacity, but also as a friend; and to the other members of my thesis committee, Dr. Ian Begg and Dr. William Carment, who made my graduate career both educational and enjoyable. Of the many friends whose support, discussion and comments were essential, special thanks go to Dr. Renaud Leblanc, Cem Kaner, Dr. Marvin Krank, Bruce Whittlesea, Don Shattuck, Dr. Larry Jacoby, and Don Maxwell, and to Dr. John Cote who, as my undergraduate advisor and close friend, set me on the road to graduate school. I also would like to thank Jan Kennet and Barbara Weingarten for their assistance in constructing some of the experimental materials and the thankless task of transcribing literally hundreds of pages of data. Finally, I would like to thank my wife, Christine, who not only assisted me in all phases of this project, but whose love and support were necessary for its completion. In but a small token of my appreciation, I dedicate this work to her.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Conventional wisdom has it that much human information processing is based upon implicit, abstract structures such as grammars, orthographies, scripts, frames, schemata, prototypes and rules. These abstract structures are generally considered to be the products of a lifetime of experience and to provide the basis for the expert skills so readily seen in such tasks as social interactions, games, the classification of natural objects, reading, speaking and writing grammatically. The structures underlying fully developed skills seem inevitably to be implicit. In contrast, performance directed by the explicit application of rules appears to be a slow and cautious or error-ridden process; it is only when the process becomes automatic and away from immediate awareness that skilled performance emerges. Moreover, although rarely discussed, it also appears that the apprehension of many of these abstract structures is, and may have to be largely implicit. Our experiences in education as well as investigations of explicit rule induction (e.g., Bruner, Goodnow & Austin, 1956) indicate that people have great difficulty explicitly learning rules that appear to be much less complex than those proposed, for example, by linguists.

Recently, researchers have found it possible to study this implicit apprehension and use of structure in the laboratory. Probably the most remarkable finding of this work is that with as little as 20 minutes of experience with only a few events of a complex, artificial stimulus domain, people perform in a manner consistent with the underlying structure or rules by which the stimuli were generated. This rapid acquisition is especially impressive since these rules appear to be beyond their ability to acquire explicitly (e.g., Posner & Keele, 1968, 1970; Reber, 1967, 1969, 1976). One way of accounting for these results, as well as for the apparent ubiquity of implicit abstractions in general, is to argue that people have considerable powers of abstraction that are not normally available for conscious, directed learning. This, in one form or another, is the accepted conclusion of most researchers in the field. In fact, in most areas of cognitive psychology, it is taken for granted that people can acquire these implicit, complex structures. The job usually is seen as one of assessing how these structures determine and coordinate behaviour rather than how they are acquired in the first place.

One would suspect that such a widely accepted notion as implicit abstraction of structure would have a wealth of clearly supporting evidence. In particular, one would suspect that this assuredly must be true in the area of concept formation or object classification where the nominal task of the subjects is to learn to classify or sort events on the basis of some abstract principle or construct. Oddly, while there are innumerable demonstrations of <u>explicit</u> abstraction in a whole variety of different situations (e.g., Bruner <u>et al.</u>, 1956; Bourne, 1974), the purported demonstrations of <u>implicit</u> abstraction mentioned above have been relatively few in number and limited almost entirely to concepts in the form of prototypes. More seriously, the results of these investigations are equivocal, either with respect to what has been abstracted or whether implicit abstraction has occurred at all. There are a number of potential reasons for this lack of empirical support. The one to be entertained in this thesis is that the failure to produce unequivocal demonstrations of implicit abstraction occurs because implicit abstraction itself rarely, if ever, occurs.

An alternative to implicit abstraction has been suggested in a number of recent papers in the area of concept formation (Brooks, 1978; Hintzman & Ludlam, 1980; Medin & Schaeffer, 1978; Medin & Schwanenflugel, 1981; Medin & Smith, 1981). The basic notion here is that the implicit knowledge that a person brings to bear in many situations is not normally very abstract. Rather, it consists of a large collection of individual prior instances, at least some of which are similar to the present events in most respects. According to this view, the appearance that an implicit abstraction has been learned and is being used is often illusory. The rule-like consistency that people demonstrate is said to arise from the consistency of the separate events that they have experienced, rather than through the use of some single, common abstraction of the prior experiences.

The process of instance-based classification is one of close analogy (Brooks, 1978) and depends for its success upon the fact that events that are near twins of each other are likely to be members of the same category. Generally speaking, the more alike two events appear to be, the more probable it is that they share a common category label, and also share numerous other "irrelevant" properties that are not properties of the higher-order category to which they belong. It is the prevalence of these "irrelevant" properties that has often been cited as the major problem of analogical or metaphorical reasoning. But this is only a problem if the analogical process is used to infer the summary properties of some class or category of events. Analogy as a strategy for extracting class properties is fraught with error. As an identification strategy for individual events, however, it can be shown to be quite effective, particularly as the number of shared features between two events increases (see Brooks, 1978; Brooks, Note 1).

The framework of this thesis is to consider the possibility that human knowledge consists of a combination of explicit abstractions and collections of instances, without resorting to a notion of implicit abstraction. Key to this conception is the above-mentioned distinction between the information and processes that may be used in the identification of individuals as exemplars of a class and the information and processes that may be used for other purposes involving classes or categories. It is to this ability, the identification of exemplars as members of classes, that the research and theorizing in this thesis is directed. The argument is that analogy to individual instances, and not the implicit abstraction of the categorical structure, provides the best explanation for this ability.

The remainder of this chapter is devoted to a discussion of some of the more often cited reasons why analogy to individual instances has been considered to be an unlikely explanation for classification behaviour. These are arguments <u>de jure</u> which, as will be demonstrated, are based on questionable premises and assumptions about memory and processing, and what an individual instances model can and cannot do. The effects of these arguments on the <u>de facto</u> viability of an individual instances explanation are discussed in Chapter 2. These general arguments against the individual instances approach are discussed here as the argument from insufficiency, the argument from economy, and the argument from learning and memory.

<u>The Argument From Insufficiency</u>

Analogy to individual instances has not often been considered as a model of conceptual behaviour primarily because it is the antithesis of what is generally considered to be "conceptual" knowledge. Concepts are over-riding abstractions that, while <u>exemplified</u> by the individual instances, are not to be equated with the instances themselves. Thus, for example, knowing a number of different individuals, separately, as dogs, is not generally thought to be the same thing as "having" the concept "dog", and it is the learning of the concept with which, nominally speaking, concept formation research is concerned. This distinction between abstract classes and the individuals that exemplify any given class has long been heralded as one of the major advances of modern logic (e.g., Russell, 1946). Without this distinction, traditional logic found itself confronted with valid, but demonstrably false conclusions from true premises.[1.1*] However, that this distinction is necessary to avoid paradox and fallacy in reasoning about classes and the properties of individuals does not imply that it is necessary for the identification of individuals as members of classes. As mentioned, the processes and information that one uses to classify individuals may differ in both form and content from the logical representation (i.e., intensive, definitional) and use of the class or concept. Thus, even though memory for individual instances is insufficient for many conceptual operations, it could still be the basis for performance in other tasks such as identification.

The same distinction between the knowledge necessary for certain logical operations on concepts and that used in the identification of exemplars as members of a concept has recently been raised by Osherson and Smith (1981) and affirmed by Bourne (1982). They point out that a number of conceptual operations, such as generation of new concepts from the conceptual conjunction of the old, require knowledge in the form of strict defining characteristics to avoid paradox and logical fallacy. While the particular alternative that they were debating was conceptual knowledge represented in terms of typical characteristics for the purposes of identification, the point is the same as that made above with respect to individual instances. Consider Osherson and Smith's (1981) example of the generation of the concept "striped apple" from the conceptual conjunction of the concepts "striped" and "apple". If "striped" and "apple" are both represented in terms of their typical characteristics then the concept "striped apple" may be represented as the fuzzy intersection of its two constituents. This implies that a given striped apple is

* Notes may be found in the section labelled Notes following the References.

at best only as good a member of the concept "striped apple" as it is of the concept "striped" or the concept "apple". Yet it is clear that a striped apple is a better illustration of the concept "striped apple" than it is of either of the constituent concepts. Similar contradictions emerge for the construction of logically empty (i.e., if it is a member of one constituent concept it cannot be a member of the combination), logically universal (i.e., if it is a member of one constituent it must be a member of the combination) and disjunctive concepts (see Osherson & Smith, 1981, pp. 43 – 48). Osherson and Smith suggest that even though people's classification behaviour with natural categories appears to emanate from their knowledge of typical members rather than strict defining characteristics, one should not necessarily equate people's apparent representation of the conceptual information used for classification with the information about the concept used for other purposes.

An important conclusion follows from the distinction between the identification routines and the "core" knowledge of a concept. Even in those situations in which it is clear that people explicitly know the rule, it still need not be assumed that reference to the rule is the basis for many of their judgements about the individual exemplars (Bourne, 1982; Medin & Schaeffer, 1978; Armstrong, Gleitman & Gleitman, in press). The fact that I know and can verbalize the definition of triangularity in Euclidean space, for example, does not mean that this information is activated in all or even many of the interactions that I may have with triangles. It is quite possible that reference is made to this rule only when I am asked to justify some decision I have made about a given triangle or triangles in general. In many other situations, including labelling, drawing and imagining triangles, I may be drawing upon far more specific information to accomplish the task. The same is true for any abstract representation of a category. That classification behaviour is consistent with a rule, prototype, or feature frequency list as the basis of the behaviour does not necessarily imply that some abstract structure is the basis for classification. Nor does the fact that strong evidence for some abstract representation is found in some non-classification task (e.g., naming of the rule, production of the previously unseen prototype), discredit the possibility that memory for individual instances is the basis of the classification behaviour. Once it is acknowledged that different conceptual tasks may involve different processes and information, the arguments against individual instances as the basis of classification by reason of logical paradox for some tasks or an inability to account adequately for all facets of conceptual behaviour may be seen as irrelevant even when the different conceptual tasks involve what is nominally the "same" concept.

The position taken in this thesis allows plenty of room for the operation of prototypes, grammars, rules, scripts, frames and abstract representations of many forms. There is too much evidence for such structures to deny their existence or their importance. However, we should be wary of the too ready acceptance of these structures as implicit and gained through some implicit process. As is discussed in a review of the literature in Chapter 2, the evidence for implicit abstractions is equivocal. It may be the case that people do obtain and use abstractions such as prototypes, schemata, frames, etc. in these tasks. But if so, the position of this thesis is that such knowledge is probably explicit, and was gained through some explicit process. Where the behaviour has the appearance of implicit abstraction, this appearance is illusory, arising from the distribution of the individual events that people use to accomplish the task. For much of the research discussed in Chapter 2 there is no question that abstraction has occurred. However, it is probably best represented as explicit abstraction on the part on the experimenters than implicit abstraction on the part of the experimental subjects.

The Argument From Economy

The reigning doctrine of implicit abstraction is supported by an emphasis on what may be referred to as the "economy of abstraction" argument. This argument is predicated on the notions that the human information processor is limited in capacity and, as a consequence, is constantly striving to minimize both the amount of processing to be done and the amount of information that must be retained. Abstracting and retaining only the essential invariances of some class of events reduces the memory burden when the learner is confronted with new members of the class (e.g., Oldfield, 1954; Fried & Holyoak, 1982). Stated this way, the argument is almost certainly true. For any given class, it is clearly more economical to remember a concise description of the class than to populate memory with every class member encountered. The problem with the argument is that the statement of economy extends no further than the particular class in question. When one considers instead the vast number of classes to which any single event may belong, representing every possible class in an abstract form rather than retaining the individual exemplars very quickly loses any semblance of being "economical".

The problem is clearly evident when one considers it in terms of the level of abstraction in a hierarchical category structure. Higher-order abstractions are of no use in making subordinate distinctions. Having some high-level abstraction of "dogness" that allows for the species classification of my pet as a dog does not allow for the subordinate classifications from an experience with him as a mongrel, a friendly beast, an occupant of my house, or as one of a number of events that I fondly refer to as "Barth". For each of these successively subordinate distinctions, a different abstraction would be required. In addition, since people clearly do recognize individual events qua individuals, the implicit abstraction hypothesis would appear to burden memory not only with a plethora of abstractions, but also with separate memories for many of the individual events that we experience. This may be true, but it is hardly economical.

In contrast, one argument for entertaining the notion that people generally rely on knowledge at least as specific as an instance hinges on the fact that lower levels of specificity do allow for classification at superordinate levels. Encountering some furry beast, for example, that appears to be a virtual twin of my dog allows me to conclude not only that it is a dog, but also probably a mongrel, and, with less certainty, a friendly one at that. By analogy to specific experiences with my own dog, I gain the information necessary to label the new animal as a dog and possibly as friendly, if such is required, and important information about how to respond to its presence. In fact, if the twinning was close enough and the context appropriate, I would probably conclude that it was my dog and react accordingly.

There is another aspect of the hierarchical structure of categories that seriously questions the economy of abstract representations. In addition to often being confronted with questions at a level subordinate to the original distinction, the regularities inherent in these lower levels may often help in answering questions at the original level. That is, arguing that it is always more economical to classify events using the highest level of abstraction appropriate for the immediate question (e.g., a rule for "dogness" to the question "Is it a dog?") is to suggest that for most domains the class events are so uniformally distributed that local predictors would be of little benefit for the immediate question. Aside from a few artificial stimulus domains created for investigations of explicit rule induction (e.g., Bruner et al., 1956), most investigations of object classification learning have used sets of stimuli with extensive a priori clustering of the stimuli, primarily in an attempt to model the clustering evident in natural categories (Rosch, 1978). Many prototype theories, for example, emphasize the usefulness of the central clustering of many categories, but they, as well as rule models, ignore the local domains of similarity that occur in the overall general clustering of events in a particular category for the purposes of applying a general label. For example, animals in the category "dog" may cluster about some central tendency or prototypical dog. However, if this prototype is the basis of classifying animals into the category "dog" then it would be ignoring the subordinate clusters of different breeds of dogs, and ignoring, at a lower level yet, the cluster of experiences relating to contextually correlated experiences with a particular dog. Any advantages accruing to this redundancy of information and similar encoding contexts are lost to the higher levels of abstraction. Thus, arguing for the use of abstractions at the highest available level suggests that we are rarely in situations in which information about lower-level distinctions (e.g., the breed, name or temperament of a particular dog) is also required or potentially useful. This may be reasonable for those experimental situations in which a single category label has stood for all of the information available about any individual, but it appears to be a poor model of real-world experiences and situations (see Brooks, Note 1).

The arguments about the economy of different representations of categorical structure are, at any rate, somewhat of a pseudo-issue. Few people would argue that all knowledge is exclusively either abstractions or individual instances. The issue rather is one of whether for a given task one or the other is predominant. There are many situations in which there is no question that the basis of performance is best represented as abstractive. This is most clearly seen in tasks requiring explicit rule induction, and in many of our most spectacular scientific and engineering successes. But, as Brooks (Note 1) argues, the brilliance and intelligence associated with abstraction should not blind us to the fact of its limited utility in many, if not most, areas of our more mundane performances, and its possible hindrance to successful or even adequate performance in many other tasks and situations. One ~~of the~~ major arguments of this thesis, in line with the notions of Brooks (1978; Note 1), is that abstraction is far less ubiquitous than often has been assumed. For many tasks ~~and~~ situations it is clearly not <u>necessary</u> that our knowledge about the domain be in an abstract form. The argument here is that it also may not be "economical" since our world is only rarely so well determined or organized that abstractly representing our experiences prepares us adequately for the future demands of our knowledge. One of the major advantages of retaining relatively complete representations of the individual events that we experience is the flexibility that they allow for subsequent tasks.

<u>The Argument From Learning and Memory</u>

It is common knowledge that learning individuals <u>qua</u> individuals is a difficult process, requiring many exposures. Moreover, common knowledge informs us that our memory for individual events is fragile at best, unless elaborate procedures are employed. Thus, memory for individual instances as the basis of much of our structural knowledge would appear to be an unlikely possibility. However, research on people's recognition memory for pictures experienced only once suggests that our memory for singly encountered events may not be as bad as common knowledge suggests. Shepard (1967), Standing, Conezio, and Haber (1970) and Standing (1973) demonstrated that people can discriminate previously experienced from novel pictures even if training consisted of as many as 10,000 different pictures. Moreover, recent work by Kolers (1979), Jacoby and Dallas (1981), and others (e.g., Jacoby & Witherspoon, 1982) suggests that the influence on performance of a single, prior exposure may have been underestimated. Kolers demonstrated that both the orientation and typeface of script given a single reading influenced the probability of its subsequent recognition as well as the speed with which it was re-read. Moreover, this influence of a single prior exposure on re-reading time extended to events subsequently encountered a year later. Jacoby and Dallas (1981) found that a single, prior presentation of words under a number of different encoding conditions strongly affected their subjects' ability to detect the words in a tachistoscopic word identification task, although, as indexed by a second group of subjects given the same tasks and words, some of the encoding conditions did not aid subsequent recognition performance. These results suggest that a single, prior episode can have large effects on subsequent experiences with the same nominal event, and that our memory for individual experiences, at least in so far as performance is concerned, may not be as fragile as common wisdom would suggest. Thus, it is also possible that such may be true for the different, but highly similar events in a classification task.

The argument from learning and memory typically invokes a second, and potentially more damaging, component. As is detailed in Chapter 2, much of the variation in our conceptual performance appears to occur in the absence of any corresponding variation in people's memory for individual, prior experiences. As is discussed in subsequent chapters, it is this lack of any apparent relationship between memory for individual events and conceptual performance that, as much as anything else, is cited as a major reason for dismissing the individual instances model. If people are able to classify novel exemplars correctly, for example, while demonstrating little or no memory for the original training exemplars, it does appear unlikely that it is the memory of these individual training exemplars that is responsible for the classification performance. Yet, despite the apparent unlikelihood, this is exactly what I am suggesting. In fact, as will be discussed in more detail later, the individual instances approach that I will propose holds that very good memory (as commonly assessed, e.g., recognition) for individual exemplars should be associated with reduced categorical transfer, while relatively poor memory (again, as commonly assessed) for specific exemplars should, up to some point, be associated with increased categorical performance. It is this paradoxical relationship between the memory for specific events and their usefulness for categorical transfer that, I believe, is the key to an understanding of individual instances as the basis of implicit, structural knowledge. For example, the independence between recognition and performance that Jacoby and Dallas (1981; see also Jacoby & Witherspoon, 1982) found for many of their experimental conditions, suggests that the influence of specific, prior events can occur without the subject's being able to specify the source of the influence. This independence also may be operative in many classification tasks in which the subject's knowledge is also apparently implicit. Thus, the lack of conscious contact with specific, prior experiences during a classification task (ie., the typical measure of "memory" for specific instances in concept research) does not rule out the possibility that specific, prior episodes in memory are responsible for categorical transfer.

The argument for instance-based classification has its parallels in other areas of cognitive psychology, particularly the study of recognition memory. Both the encoding-specificity (Tulving & Thomson, 1973, for the seminal statement; Craik, 1981, for a useful review) and the multiple-trace (Madigan, 1969; Hintzman & Block, 1971; Hintzman & Stern, 1978; Hintzman, Nozawa, & Irmscher, 1982; Kolers, 1979) approaches to memory are arguing for an increased role of event-specific or episodic information as a way of accounting for context and repetition effects. The experiments and theorizing entertained in this thesis borrow heavily from these notions. In particular, the instance model proposed herein to account for the experimental results takes the instance models of Brooks (1978) and Medin and Schaeffer (1978) further in the direction of specificity by proposing that many aspects of categorical transfer result not just from the distribution of individual exemplars in memory, but also from the distribution of the specific experiences with each exemplar.

There are a number of additional advantages to the specificity, or individual instances, perspective. Primary among these is that it does not attribute to the learner an ability of unconscious abstraction that vastly exceeds his or her conscious ability. Consequently, it also provides an explanation for the development of expertise. Increasing the number and range of specific events in memory increases the probability that there will be at least one prior experience available to guide the processing of, and response to most new events within the domain of expertise. More generally, it provides a mechanism for dealing with the complexity of the world around us. By analogy to specific events, it is not necessary that we have a complete understanding of most domains to function efficiently within them. Moreover, it does not require consistency across a broad range of events, allowing for the appropriate modification of our behaviour in response to local idiosyncracies or predictors that are only rarely included in some of our most spectacular successes of explicit abstraction. Finally, a number of investigators have suggested that people's estimates of the relative frequency of events arise from the retrieval of individual episodes rather than some summary statistic associated with the events in question (e.g., Hintzman & Stern, 1978; Tversky & Kahneman, 1973). If it is the case that we retain and use individual events for this purpose, it seems all the more likely that event-specific memory plays a large role in many other decision tasks, including classification and recognition.

An Overview of the Thesis

The subsequent chapters of this thesis are an attempt to challenge directly the widely held notion of implicit abstraction in concept formation and classification. In Chapter 2, the concept learning literature is reviewed and it is concluded that the evidence for implicit abstraction in these tasks is at best equivocal. Included in Chapter 2 is a discussion of the recent emphasis on the learning of "ill-defined" categories.

Chapter 3 of the thesis presents a discussion of the basic experimental paradigm used in all of the experiments to be reported. As is discussed there, a major problem with much of the research in the classification literature is a confounding of the similarity between specific events and their categorical status. The discussion in Chapter 3 presents a number of solutions to this problem, as well a number of arguments with respect to the assessment of categorical transfer.

The following six chapters present a series of experiments designed to investigate implicit learning from the perspective of individual instances.[1,2] These experiments were constructed most directly to challenge the particular demonstrations of implicit abstraction of structure offered by Arthur Reber (1967, 1969, 1976) and his associates (Reber & Lewis, 1977; Reber & Allen, 1978; Allen & Reber, 1980; Reber, Kassin, Lewis, & Cantor, 1980), although the last three experiments were designed with a more general appeal. The work of Reber and his associates is summarized at the end of Chapter 2.

The final chapter of the thesis presents a summary of the experiments and a discussion of the relationship between concept formation and episodic memory research. Included is a brief discussion of the implications of this work for the fields of education and artificial intelligence.

Chapter 2

RULES, PROTOTYPES, AND INSTANCES: A LITERATURE REVIEW

Early research on concept formation was concerned with the explicit abstraction of rules and regularities. Recent research has emphasized the implicit learning of the "ill-defined" or non-rule structure of many natural categories, resulting in an equation of what is learned with the structure of the domain in question. However, as I argue in the first section of this chapter, the apparent differences engendered by the different category structures might better be conceived as a property of the manner in which people deal with a category rather than a property of the category itself. One advantage of this view is that it allows for the investigation of the implicit learning of structure using well-defined, rather than "ill-defined" category structures, such as the artificial grammars used by Reber and Allen (1978) and Brooks (1978), yielding greater experimental control over the experimental materials. Subsequent sections of the chapter review the basic findings relating to the learning of category structure, and discuss the various models that have been proposed. The chapter ends with an extended discussion of the learning of artificial grammars, and an individual instances explanation of implicit learning of structure is presented.

The Structure of Categories

About 1972, a minor revolution occurred in the study of concept formation and classification. Prior to this, investigations of concept learning used stimulus structures in which the concept or category was defined by a single, typically simple rule (e.g., Bruner et al., 1956). In contrast, recent research has emphasized the importance of investigating the learning of ill-defined categories (Neisser, 1967). This research has been directed at the learning of categories whose category membership is graded and in which no single rule or list of defining features correctly predicts classification of all of the members of the category. Categories of this sort are characterized by "best" members, typical and atypical exemplars, and a general "fuzziness" to their definitions. In effect, the category is defined in terms of its best members or central tendency (mean or mode). A given instance is considered to be a member of one category rather than another if it bears more in common with the best members of one category than with those of some alternate category. Rosch (1973, 1975a, 1975b, 1978) and her associates (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Rosch, Simpson, & Miller, 1976) have argued persuasively that this ill-defined or prototype structure is common to most natural categories, particularly those that they refer to as "basic level categories" in which the categories are said to be maximally separable.[2.1] Examples of categories of this type are provided by such concepts as birds, the colour red, and hand-written letter Es (Homa, Sterling & Treppel, 1981). In fact, by this definition, the various manifestations of the same nominal person, for example, may be considered to be exemplars of the ill-defined category provided by the label for that person. In each case, membership in the category for any given instance is graded, and some instances on the fringe of a category (e.g., penguins in the case of birds, the letter E in a hastily scrawled message, or the appearance of a friend who has just removed his ten-year beard and has dramatically altered his style of dress) may be difficult to recognize as members of their appropriate category. Recent concept formation and classification research on categories of this type has attempted to ascertain what it is that people learn about the category, how this information is represented in memory, and how it is used to classify exemplars into their appropriate categories.

These concerns contrast with those of traditional concept learning research, in which at least one aspect of what is learned, the rule, is obvious. In traditional concept learning, the research has been directed at the explicit induction of relatively simple rules, and has focussed on such concerns as the hypothesis testing and strategy choices of the subjects (Bourne, 1974). Research in this tradition typically has used stimuli consisting of a small number of easily separable, orthogonal dimensions with uniform distributions of a small (and equal) number of discrete values along each dimension (Bourne, 1982). Moreover, in direct contrast to research with ill-defined concepts, the primary task confronting subjects with these logical concepts is usually the explicit identification of the features that are relevant to a specified form of rule (i.e., attribute identification). In these cases, the subjects are told the form of the rule for classification (e.g., conjunctive, disjunctive) and, in some cases, even which stimulus dimensions are involved (e.g., Bourne, 1982). According to Rosch (1978), these various aspects of the structure of the concepts and the tasks presented to the subjects in the traditional paradigms render them poor models of concept formation and classification for natural categories. Recent research within the traditional paradigm, however, has emphasized the importance of feature frequency information within these tasks, and has led some theorists to suggest that, contrary to Rosch's (1978) contention, at least some aspects of <u>what is learned</u> with rule-defined concepts and ill-defined concepts may not be all that different (e.g., Bourne, 1982; Bourne, Ekstrand, Lovallo, Kellogg, Hiew, & Yaroush, 1976; Kellogg, 1980).

The conclusion that there may be little functional difference between the two paradigms in what subjects learn is consistent with the work of a number of researchers whose experimental paradigms straddle those associated with traditional and ill-defined concept research. These experiments are characterized by having well-defined concepts, or rules determining category membership, but the subjects' task is to learn to classify events and not explicitly search for rules or defining features (e.g., Hayes-Roth & Hayes-Roth, 1977; Reber, 1967, 1969, 1976). The rules, however, are typically complex (e.g., an artificial grammar), and are not readily amenable to explicit solution with a limited number of experiences with the exemplars. In fact, if the rules are complex, the exemplars may cluster in the category in a fashion similar to the ill-defined or prototype structure described by Rosch (1978). And even if such is not the case, there almost always will be local domains of clustering within the category (see Brooks, Note 1). Possibly because of this, the classification and recognition performance of subjects in these experiments is similar to that of subjects learning ill-defined concepts. Classification accuracy is typically less than perfect but better than chance, and subjects respond to exemplars of the category in a graded fashion despite the fact that, in some sense, all exemplars are equally "good" members of the category. In fact, Hayes-Roth and Hayes-Roth (1977) found that even subjects who were informed of the rules for classification showed effects of category structure (i.e., the distribution of exemplars presented during training) that were similar to those of subjects not so informed, despite the fact that since they had the rules they could have ignored the category structure when classifying and recognizing the stimuli. Similarly, Armstrong _et al._ (in press) demonstrated that many of the "typicality" effects in identification accuracy, latency to identify, and priming, thought by Rosch and others to be evidence of the use of abstract prototypes, were found to occur as readily with "well-defined" concepts (e.g., "odd number") in which the subjects had already acknowledged the non-protypical nature of the categories as they were for "ill-defined" concepts (e.g., "vehicle") that the subjects agreed admitted graded membership. Moreover, Brooks (1978) and Reber (1976) have demonstrated that subjects instructed to induce the rules explicitly from experiences with exemplars from a complex stimulus domain (and who, consequently, fail miserably at the task) perform less well on a subsequent classification test than do subjects who are simply asked to memorize ~~the~~ stimuli. Apparently, unlike the the subjects not instructed to look for rules, in their unsuccessful search for rules, the rule-instructed subjects were unable to obtain and utilize much of the structural information of the stimuli (Brooks, 1978; Reber, 1976). Thus, the important distinction may not be between rule-defined and ill-defined category structure, but rather between rule-defined and ill-defined treatment of the category by the subjects for a given task. If the rules underlying category membership are, as judged by the subjects, exceedingly complex, the processes and information that subjects use for classification (rather than rule explication) may be quite independent of whether the category, as defined by the experimenter, is ill-defined or well-defined.

One way of looking at ill-defined categories is to consider them to be categories for which the person using them has not as yet determined the complex rules defining category membership. In fact, "ill-definedness" might better be characterized as a property of the information about the category that a person uses in his or her dealings with a category than as a property of the category itself. That is, a category is "ill-defined" if a person interacts with it for a given task in such a way that what is being directly reflected is the distributional structure of the category rather than the underlying rules. The following example should make the point clear. As Rosch and her associates have stressed, a category such as "dogs" is ill-defined. Some dogs are better exemplars of the category than are others, and no list of defining features readily available to most observers appears to be capable of correctly labelling all dogs while correctly rejecting all other animals. Yet to a zoologist, the definition of a dog can be framed in a quite straightforward, rule-defined manner, although it may require dissection, blood tests and other operations to assess the defining characteristics for any given animal. Thus, at least at one level, the concept dog is not ill-defined. The same applies to birds, colours, and numerous other categories that have been referred to as ill-defined. The important point, however, is that while a category such as "dogs" may not be ill-defined to a zoologist in his or her role as a zoologist, it is probably the case that he or she, like the rest of the population of observers, treats the category as ill-defined in most interactions with it. That is, for many decisions and tasks involving the category and its members the information that is being reflected in people's behaviour arises from the distributional structure of the category rather than the rule or defining features known only to the practising zoologist. This becomes particularly evident when it is recognized that even for the many concepts that have defining features of which most observers are aware, the observers' treatment of the category will be ill-defined because oftentimes the information or features necessary to apply the rule or feature list are unavailable, unobservable or simply too costly to obtain for the immediate task. This seems to be all the more likely for those situations in which the observer has neither the time nor inclination to obtain a rule in the first place. Thus, by concentrating on people's <u>treatment</u> of a category rather than the structure of the category itself, a rapprochement may be obtained between research with categories with known definitional rules and research with categories for which the rules have not (or cannot) be specified by either the subject or the experimenter. Moreover, this characterization of "ill-definedness" stresses the distinction mentioned in Chapter 1 between "the concept" and the processes and information about the concept that a person may use in many conceptual tasks (Osherson & Smith, 1981; Bourne, 1982;

Armstrong et al., in press).

The Learning of Categorical Structure

According to the preceding discussion, treating a category in an ill-defined manner for some task involves using in some fashion or another the distributional structure of the category. There have been numerous demonstrations that people given experience with exemplars of a category are sensitive to the distributional structure of the set of exemplars experienced and, through this, the distributional structure of the category as a whole. The paradigm most often used in the investigation of people's learning of categorical structure is that of Posner and Keele (1968, 1970). The stimuli are random-dot patterns.[2.2] Nine dots are randomly dispersed on a page, producing a generative pattern called the prototype. Distortions of the prototype are then produced by moving each dot according to some statistical rule. If the average distance that the dots have been moved to produce a given pattern is small, the pattern is referred to as a low-distortion pattern because it will tend to resemble the generative prototype. Large average movements of the dots produce high-distortion patterns that have little resemblance to the generative prototype. A prototype pattern and its set of distortion patterns is called a concept. Subjects' mean judged similarity of the distortions to their generative prototype is reasonably well fit by a $\log_2$ transformation of the average distance (city-block) the dots have been moved for a given distortion (Posner, Goldsmith & Welton, 1967). Thus, it has been common to express the degree of distortion of a pattern from its generative prototype in information theory terms of bits/dot, although more recent investigations have reverted to a mean Euclidean distance per dot measure (e.g., Homa, 1978).

In Posner and Keele's paradigm, subjects are given training with a sample of distortions from each of a number of different prototypes, and their task is to learn to label, with feedback, the correct category of each pattern. In most investigations of this type, the prototypes are not presented during training. After training, the subjects typically have been given a classification task, without feedback, of the original training stimuli, the previously unseen prototypes, and a set of new distortions of each prototype. The usual results, which have been replicated numerous times, are that subjects demonstrate high classification accuracy for the training exemplars and the prototypes, and less, but still above chance, accuracy for the new distortions, with the accuracy for any given new distortion varying inversely as a function of its distance from the prototype (e.g., Posner & Keele, 1968, 1970; Strange, Keeney, Kessel, & Jenkins, 1973; Peterson, Meagher, Chait, & Gillie, 1973).

Since the original demonstration by Posner and Keele (1968), researchers have attempted to assess what the subjects have learned in these tasks by manipulating a number of parameters of the original paradigm. Probably the most persuasive of these for the implicit abstraction position has been the effects of inserting a delay between training and the classification test. A number of investigators have found that delaying the classification test for periods of a few days or weeks reduces the classification accuracy for the training exemplars, but has little or no effect on the classification accuracy for the categorical prototypes or novel exemplars (e.g., Posner & Keele, 1970; Homa, Cross, Cornell, Goldman, & Scwartz, 1973; Strange et al., 1970). Results such as these have been interpreted to suggest that memory for the training exemplars is more susceptible to forgetting than is some abstract representation of the exemplars. Hence, while immediate classification is thought to be a function of both memory for the training exemplars and some abstracted prototype, classification following a delay is said to be primarily a function of the abstract representation that the subjects have formed from their experience with the training exemplars (e.g., Robbins, Barresi, Compton, Furst, Russo, & Smith, 1978; Posner & Keele, 1970).

Other parametric investigations have demonstrated that training subjects with high- or mixed-distortion stimuli rather than low-distortion stimuli of a given prototype results in superior classification accuracy (Homa & Vosburgh, 1976; Posner & Keele, 1968; but see Peterson et al., 1973). As well, increasing the number of exemplars of a given category experienced during training, the "category size" effect (Homa et al., 1981), increases the subsequent classification accuracy for new items from the same category in an immediate (Homa, 1978; Homa et al., 1973; Homa & Hibbs, 1978; Homa & Vosburgh, 1976) and delayed (Homa et al., 1973; Homa & Hibbs, 1978; Homa & Vosburgh, 1976) classification test. In addition, increasing the number of categories to be discriminated during training and test also increases classification accuracy (Homa & Chambliss, 1975).

Similar findings using different stimuli and category structures have also been reported. Reed (1972), using categories composed of schematic faces varying along four dimensions (e.g., nose length, eye width), found that classification accuracy of novel faces varied as a function of similarity between any given exemplar and its category prototype. Using structurally analogous, biographical descriptions produced essentially the same results (Reed & Friedman, 1973). Goldman and Homa (1977), using schematic face stimuli similar to those of Reed (1972), replicated the positive effect on classification accuracy of increasing the number of training exemplars. Rosch, Simpson, and Miller (1976) investigated the effect of different stimulus structures on subjects' performance in a number of different tasks. They used three different stimulus structures, each requiring different similarity metrics for the calculation of "typicality" or distance from the category prototype.[2,3] These were: (1) random-dot patterns, in which similarity to the prototype is determined as discussed earlier; (2) stick figures, that are structurally equivalent to the schematic faces used by Reed (1972) and in which distance to the prototype is measured by the sum of differences along each of four quantifiable dimensions (e.g., length of legs, diameter of head); and (3) letter/digit strings in which distance to the prototype is quantified in terms of a family resemblance score (i.e., the sum of frequency-weighted letters/digits in common with other members of the category). The results were essentially equivalent for the three types of stimuli. Errors in classification, reaction time to classify the stimuli, ratings of typicality, and production of stimuli from memory were all systematically related to the typicality of the stimuli. Less typical stimuli resulted in more errors in classification, slower times to classify the stimuli, lower typicality ratings, and were reproduced less often and much later in the stimulus production task. Rosch et al. (1976) attribute these results to the structure of the stimulus sets, but explicity do not suggest any process whereby the subjects apprehend the structure or how the acquired information is used by the subjects within the experimental tasks.

In all of the above experiments, subjects were given discrimination training with exemplars from a number of different categories, and were tested by a classification task for what they had learned. Some researchers, however, have attempted to investigate the learning of categorical structure with what may be referred to as a recognition paradigm. Here, the training phase consists of having the subjects study the exemplars from only a single category in preparation for a test of item recognition. The dependent variable in these studies is the rate of "old" responses to various types of test stimuli. In general, the results are similar to those found with discrimination training and classification. Subjects are more likely to misrecognize as "old" the prototype of the category, and false-positive responses to new exemplars vary as an inverse function of the distance between the item and the prototype (Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974, 1977; Bransford & Franks, 1971; Franks & Bransford, 1971; Reitman & Bower, 1973; Lasky & Kallio, 1978). These results suggest that, at least for the materials used in these tasks (e.g., identikit faces, patterns of geometric forms, biographical descriptions, and compound sentences), classification and recognition use of similar types of information.

Models of Structural Learning

The models that have been proposed to account for the apparent learning of categorical structure fall into three general classes. The first two, prototype and attribute-frequency models, hold that subjects abstract some central tendency of the category from their experience with the exemplars. Generally speaking, the difference between these two models may be seen as a difference between a parametric (i.e., mean or average) representation in the case of prototype models, and a non-parametric (i.e., mode) representation in the case of attribute-frequency models (Neumann, 1974, 1977; Fried & Holyoak, 1982). The third class of models, individual instance models, holds that both the average and modal aspects of the structure of the category are indirectly retained by means of the distribution of instances in memory.

The prototype of a category has been characterized in two different ways. Only one of them is parametric in the sense used above. The other, proposed by Tversky (1977) in his development of similarity as a feature-matching process, characterizes a prototype of a category (and there may be more than one prototype for a given category) as that instance in the complete set of exemplars of the category that maximizes the number of features in common with all other exemplars, while minimizing the number of features unique to itself. If, for example, each exemplar of the category is presented equally often and all similarity matches are symmetric, the prototype of the category is the exemplar with the highest mean similarity (as defined by Tversky, 1977) to all other exemplars of the category. Although Tversky does not explicitly develop the notion, the prototype that a subject developed from his or her experience with some set of exemplars of a category would be, according to this model, that instance that bore the above-described relations to the set of experienced exemplars rather than to the category as a whole. For all of the different models, the proposed representation of the category in memory is that derived from the experienced exemplars. Only to the extent that the these experiences are representative of the category in question will the various structures resemble the complete distribution of categorical exemplars. Since the materials used in most experiments are chosen to be representative of the various categories, equating the subjects' representation of the instances with that of the category as a whole is not unreasonable. However, it should be kept in mind when it is claimed, for example, that a subject has "learned the category prototype" that this is a short-hand statement for the claim that the subject has learned an abstract representation of the category that, because of the representative distribution of instances experienced, is equivalent to the full-distribution prototype of the category.

The parametric version of prototypes was briefly discussed above in the context of the Rosch et al. (1976) study, and has been the version most often assumed in experimental tests of the various models. This model characterizes the prototype of a category as being composed of the average value along each dimension of variation of the exemplars of the category (Reed, 1972). A more general characterization of this version of prototype is to consider the prototype to occupy the central position in an n-dimensional space such that it is least distant from all other exemplars of the category (Homa & Rhoades, 1981; Homa et al., 1981;

Neumann, 1977). This characterization allows for the representation of stimuli such as schematic faces in much the same way as random-dot patterns. Stimuli with non-continuous or even unknown dimensions (e.g., letter strings) also may be represented in this way. Each feature may be treated as a binary (presence - absence) dimension. In each case, the prototype is the multi-dimensional average of the exemplars (Neumann, 1977). With the prototype and, hence, the category, defined in this way, membership in the category for any given stimulus is a function of its distance from the prototype. How this distance or similarity between stimuli should be calculated is discussed shortly.

Attribute-frequency models, like prototype models, come in two forms. The simplest is that in which the subject is said to accumulate the frequency of each individual feature of the exemplars (e.g., Chumbley, Sala & Bourne, 1978; Kellogg, 1981). By the alternative model the subject accumulates not only the frequency of the individual features, but also the frequency of all possible combinations of features (e.g., Reitman & Bower, 1973; Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974). Regardless of whether the counted attributes are strictly single features, or single features and combinations of features, subjects' assessment of membership in the category for any given exemplar is a function of the accumulated frequencies of the attributes of the exemplar. The prototype of a category, according to this view, is best conceived as a multi-dimensional mode (e.g., Bourne, 1982; Goldman & Homa, 1977; Neumann, 1974, 1977). The prototype is the stimulus that contains the modal values along each dimension of variation of the categorical exemplars.

Instance models have received scant attention in the concept learning literature, although most theorists would acknowledge that memory for individual instances plays some role in classification and recognition following categorical training with the exemplars of a category. At the very least, most people would agree that some classification responses for previously experienced events arise from the memory for those events. What instance theorists are suggesting, however, is an extension of the notion. For example, if a nominally novel event was virtually indistinguishable from a previously experienced event, subjects could classify the new event by a process similar to misrecognition. More generally, classification is said to be a function of generalization around what is known about specific exemplars of the category. Thus, even if the subject can correctly judge that some novel exemplar is indeed novel, it is still possible for the subject to catergorize the exemplar on the basis of its similarity to individual instances in memory.

Three versions of instance-based classification models have been proposed. The two most often investigated have been the "nearest neighbour" and the "average distance" models (Reed, 1972). The nearest neighbour models propose that, following training, subjects classify any given instance into the category of the most similar exemplar in memory. The model can be expanded to include the _n_-closest neighbours where classification is determined by majority vote (Reed, 1972). Note that for the nearest neighbour model, which exemplars in memory determine classification for any given instance are a function of the instance to be classified. In contrast, according to the average distance model, subjects classify any given instance into the category whose exemplars in memory are, on average, most similar to the instance in question. For each classification decision, all of the exemplars in memory for each category are assessed. The third model, the "context-cue" model recently has been proposed by Medin and Schaeffer (1978). In many respects the context-cue model is another version of an average distance model, although as detailed below, the operative principle is more in line with the nearest neighbour model. It differs, however, in the metric proposed to assess the similarity between events. Since the assessment of similarity between events is of central concern for instance models and prototype models alike, a detailed description of the context-cue model will be deferred to the general discussion of the assessment of similarity presented below.

Tests of the Models

All of the models described, and variations of each general type, can account for the basic findings discussed above. Prototype models, however, have to invoke additional processes to handle the effects of category variability, category size, and number of categories discriminated during training. To account for these results, some prototype theorists have suggested that the variables of number and variability of training exemplars of a given category increase the incorporation of features common to the category into the abstracted prototype (e.g., Homa & Vosburgh, 1976), while increasing the number of categories to be discriminated increases the weight associated with distinctive features of the category in the subjects' abstract representation (Homa & Chambliss, 1975). Consistent with this notion that the prototype abstracted by the subjects does not weight all features equally, is Reed's (1972) finding that for the classification of schematic faces, a prototype model that weights dimensions of the stimuli according to their cue-validity (i.e., category diagnosticity) provides a better quantitative fit to the data than does a prototype model in which all dimensions are weighted equally.

Attribute-frequency models have the problem of what to count in those cases, such as random-dot stimuli, in which there are no clearly definable features or dimensions of variation. The lack of defined features is no problem if it is simply assumed that subjects accumulate frequencies of idiosyncratically defined features for these stimuli since, as long as each subject is internally consistent, the basic predictions of the model will remain the same, but it does make attribute-frequency models difficult to test in these situations. Countering this problem for frequency models, however, is the problem for prototype models of how subjects average discrete dimensions to produce a prototype if the dimensions of the stimuli are non-continuous (e.g., Chumbley et al., 1978). Problems such as these have led some theorists (e.g., Reed, 1972) to suggest that whether the subjects abstract a mean or a mode is a function of the characteristics of the stimuli and tasks presented them.

Most investigations designed to discriminate among the various alternatives have been conceived as tests between prototype models and attribute-frequency models (e.g., Lasky & Kállio, 1978; Reitman & Bower, 1973; Chumbley et al., 1978; Neumann, 1974, 1977), and at least one investigator (Kellogg, 1981) has limited one of his recent investigations to attempts to discriminate among the various attribute-frequency models.[2.4] Attribute-frequency theorists have attempted to discredit prototype models by concentrating on the fact that most prototype models posit that the subjects abstract and use for classification an average representation of the category. They argue that the appearance that subjects are using an average is an artifact of the fact that, for the distributions of exemplars of these ill-defined categories that have been presented to the subjects, the average or prototypical stimulus is also the modal stimulus, and that it is some aspect of the mode rather than the mean of each of these distributions that the subjects are using for classification (e.g., Hayes-Roth & Hayes-Roth, 1977; Goldman & Homa, 1977) and recognition (e.g., Neumann, 1974, 1977). Consequently, a simple test that should discriminate between these two models is one in which the distribution of categorical exemplars (and, hence, the frequencies of the features) presented to the subjects is skewed or otherwise distorted (e.g., multimodal). Tests of this sort have generally favored attribute-frequency models (e.g., Neumann, 1977; Hayes-Roth & Hayes-Roth, 1977; Chumbley et al., 1978) in that modal predictions about the results were generally more accurate than predictions based on the mean. However, it is also the case that most of these tests have used stimuli with recognizably discrete features along clearly defined dimensions, rendering the evidence in favor of attribute-frequency models less than general. Moreover, some of these investigations have produced results consistent with an averaging or prototype model. Neumann (1977), for example, found that subjects given a recognition test of novel identikit faces following training with faces from the periphery of the category, were more likely to misrecognize as "old" the stimulus composed of average values not experienced during training than stimuli that contained even frequently occurring feature values.[2.5]

Neumann (1977) accounts for these results and those obtained with non-discrete dimensions in general by a version of an attribute-frequency model that he refers to as the "interval-storage" model. The model proposes that subjects accumulate the frequencies of overlapping intervals along the dimensions of variation of the stimuli, in which the size of the intervals varies with the extent to which "levels" along the dimension may be discriminated. It is the degree of overlap, he argues, that produces the appearance that the subjects have abstracted an average, since previously unseen values within the range of overlap on a given dimension will accumulate frequencies as surrounding values are presented. Neumann (1977) tested the model by varying either the size of the intervals or by reducing the degree of overlap between presented values. The results he obtained were generally in accord with the model. If it was unlikely that the frequency of values of the average stimulus would be incremented by presentations of surrounding values, subjects responded in a manner consistent with the mode rather than the mean. However, these results as well as those discussed earlier also are in accord with an instance-based model of structural learning, since memory for individual instances will indirectly reflect the distibutional characteristics of the presented exemplars.

Tests of the Individual Instances Model

Memory for individual instances as a model for structural learning in these situations has been rejected on a number of grounds. Its detractors have questioned the model on the basis of a priori arguments (e.g., Reed, 1978; Fried & Holyoak, 1982), poor quantitative fit to a number of different sets of data (e.g., Reed, 1972; Hayes-Roth & Hayes-Roth, 1977), and an inability to handle such effects as those produced by inserting a delay between training and test (e.g., Posner & Keele, 1970), and increasing category size (Homa et al., 1981). However, some theorists have suggested that memory for individual instances provides the best explanation for structural learning in these situations (e.g., Brooks, 1978; Medin & Schaeffer, 1978; Hintzman & Ludlam, 1981). There are a number of reasons for this apparent discrepancy. First, the a priori arguments that have been advanced involve a number of questionable assumptions about memory for individual instances. Primary among these assumptions is the notion that the memory for any given instance has to be perfect before it can influence classification and recognition. Second, tests of the model may have failed not because the basic principle is incorrect, but because the instantiation of the model may have been innapropriate for a given experiment. Primary among these problems of instantiation has been the definition of the similarity between events. Since this latter problem has been the most prevalent, it will be discussed first.

In most tests of the various models that have been proposed, the individual instances model has not fared as well as some versions of prototype and attribute-frequency models. Reed (1972) and Hayes-Roth and Hayes-Roth (1977) assessed the quantitative fit of a large number of different models of each of the three major types to the data of their experiments and found that either a weighted prototype model (Reed, 1972) or an attribute-frequency model that included the frequencies of combinations of features (Hayes-Roth & Hayes-Roth, 1977) provided the best fit to the data. In direct contrast, however, Medin and Schaeffer (1978) using essentially the same stimuli and procedure as Reed (1972), found that an individual instances model provided a better fit to their data than did either prototype models or attribute-frequency models.

One solution to the apparent contradiction produced by the results of Reed (1972) and Medin and Schaeffer (1978) may be found in how the instance models were instantiated across the experiments. In Reed's study, the similarity between any two faces for the most successful of the individual instances models that he tested (an average distance model) was calculated as the (cue-validity) weighted sum of differences along each dimension of variation of the stimuli. The classification for any given face according to the model was determined by the mean of these sums for the training exemplars of each category. Any given transfer stimulus was assigned according to the model to that category that produced the lowest mean (i.e., highest mean similarity). Medin and Schaeffer, however, calculated the similarity between faces in terms of their context-cue model. According to this model, each dimension of variation of the stimuli is weighted according to its <u>salience</u>, rather than by its category diagnosticity as in Reed's conception.[2.6] The major aspect of Medin and Schaeffer's (1978) model, however, is that it emphasizes the non-independence of features in the assessment of the similarity between stimuli. The similarity between any two stimuli is defined as the <u>product</u> of the weights associated with each dimension upon which the stimuli differ. That is, rather than combining independently, the dimensions interact. Since the calculation is multiplicative, differences on salient dimensions between two stimuli result in a greater degree of dissimilarity than do differences on less salient dimensions. One interesting component of the model is that dimensions that do not differ across two stimuli (or have no salience) have no effect at all on the assessed similarity between the stimuli. The model predicts that the probability that any given item will be assigned to a given category is a function of the "evidence value" for that category. Evidence for a given category is defined as the sum of the separate similarities of a given item to all members of the category divided by the sum of separate similarities of that item to all members of all categories. It is this calculation of "evidence values" that renders the context-cue model as a version of an average distance model. As Medin and Schaeffer (1978) stress, however, the evidence value is not an intrinsic component of the model; it is simply a device to produce quantitative predictions from the model. In many respects, because of the weighting produced by the multiplicative relation, the model may be considered to be a nearest neighbour model, and, in fact, Medin and Schaeffer (1978) consider it from the perspective of a nearest neighbour metric in a number of their discussions of the model. The important point of the context-cue model for the present discussion is that by altering the metric used to assess the similarity between events, an instance model can be shown to provide a better account of the data than can prototype and attribute-frequency models.

There are other examples of crucial differences being dependent on the "similarity metric" used. In one of their experiments, Posner and Keele (1968, Experiment III) were concerned about the possibility that the effects that they were observing were a consequence of generalization around the memory for individual exemplars. In an attempt to eliminate this possibility they produced a set of novel transfer exemplars for each category that had the same mean distance to their appropriate training exemplars as that of the prototypes of the random-dot categories. The idea was that if classification of the previously unseen prototypes was still superior to the classification of new exemplars under these conditions, then the subjects were probably abstracting and using information about the prototype, rather than responding on the basis of similarity to remembered training exemplars. The results were similar to those in which such similarity matching had not occurred. The prototypes of each category were classified more accurately than their matched sets of novel exemplars, leading Posner and Keele (1968) to conclude that the subjects had abstracted information about the prototype of each category from their experiences with the training exemplars. The problem with experimental designs of this sort is that the demonstration is only as good as the matching is effective in controlling the variable in question. Posner and Keele matched the mean similarity to training exemplars using a city-block metric to calculate the distance between patterns. A number of theorists, however, have suggested that a Euclidean metric would be more appropriate (e.g., Barresi, Robbins, & Shain, 1975; Reed, 1972). According to this metric, the novel exemplars and the prototypes would not be matched for their mean similarity to the training exemplars. The prototype of each category would have a lower mean distance than would the novel exemplars. Hence, if this metric more accurately captures the similarity judgements of the subjects than does the city-block metric, Posner and Keele's (1968) experiment does not provide unequivocal support for prototype abstraction (Robbins <u>et al.</u>, 1978).

The problem, however, is more general than a debate between city-block and Euclidean distance metrics in some psychological hyperspace. Tversky (1977) and Gati and Tversky (1982), for example, have argued that all <u>geometric</u> representations of psychological dimensions are questionable as models for the assessed similarity between events. They have proposed a <u>set-theoretic</u> approach that appears to be more consistent with many aspects of people's similarity judgements. The similarity between events according to this view is assessed by a feature-matching process. One particular component of this approach that brings even Medin and Schaeffer' (1978) context-cue metric into question is that it emphasizes the importance of common features between events. The model proposed by Gati and Tversky (1982) predicts that the similarity between two events increases as a function of the number of features they have in common, all other things (e.g., number of different features) being equal. For example, Gati and Tversky (1982) found that adding a common feature to two stimuli increased subjects' judged similarity of the stimuli relative to that when the feature was absent from both stimuli, despite the fact that, in both cases, the stimuli differed by the same number of features. None of the more common geometric metrics that have been proposed for assessing proximity, nor the metrics proposed by Medin and Schaeffer (1978) or Fried and Holyoak (1982) allow for this intuitively plausible notion. Thus, the results of investigations that attempt to test the various models either by matching sets of stimuli for similarity to training exemplars or by assessing the quantitative fit of the models to the data using the more common metrics of similarity are likely to be of limited utility, since the degree of quantitative fit, the ordering of models with respect to this fit, and the qualitative predictions of the various models in "matched" experiments critically depend upon the measurement of similarity.

Unfortunately, the problem is not one of simply choosing the best model of similarity. Even if a precise measure of similarity between stimuli could be found, one is left with the problem of knowing to <u>what</u> the measure should be applied. It is important to note that in almost all of the investigations, the predictions of the various models are based upon the similarity between stimuli as <u>presented</u> to the subjects. The calculations for the predictions of the various models assume that the representation of the training items in memory and the encoding of the items at test are veridical with respect to how the stimuli have been defined by the experimenter. Clearly, however, it is possible for the subjects to encode and retain only certain aspects of any given stimulus, elaborate the stimuli in some way, weight some features more heavily than others because of a priori theories about the stimuli or as a result of the testing of specific hypotheses, forget some aspects of a stimulus over a delay while retaining other aspects, and so on. In fact, there are many reasons why the representation of the exemplars in memory may differ quite markedly from that used in the test of any given model. Consequently, a given model may predict poorly not because it is fundamentally incorrect, but rather because it is instantiated using representations of the stimuli that may be quite different from those used by the subjects. For example, with the random-dot stimuli, the calculation of the similarity between items is based upon the movement of particular dots across patterns. As a measure of subjects' judgements of similarity between a test item and a training item (or prototype) in memory, even ignoring the other problems discussed above, the calculation implies that the processes that subjects are using "know", in some sense, which dots have moved either to assess the similarity between two events or form a prototype from their experiences. While this implication appears reasonable for patterns with small differences between them, it seems questionable for patterns that are large distortions of each other. In recognition of the problem with large distortion patterns, Homa (1978) used stimuli in which the dots were connected by straight lines, producing polygonal shapes. This operation has the effect of making it more obvious which dots (now verticies) have moved, potentially increasing the correlation between similarity as defined by the experimenter and similarity as used by the subjects, but the more general problem that the manner in which subjects represent the stimuli to themselves or remember them may not be in close correspondance to the experimenter's treatment of the stimuli still remains. Some researchers have acknowledged at least a limited version of this problem. Reed (1972) and Reed and Friedman (1973), for example, noted that the quantitative fit to the data of the various models they investigated would be only as good as the extent to which their representation of each stimulus matched that of the subjects. For some conditions, they found that using weights for each feature derived from the judgements of the subjects rather than the objective weights determined by such concerns as category diagnosticity improved the fit of a number of models, or at least provided an account of why a model that fit well in one situation failed to do so in another (also see Neumann, 1977). Similarly, multi-dimensionally scaled distances obtained from subjects' judgements of the similarity between stimuli rather than the objective distances determined by the experimenter improved the fit of some models to the data from the classification of schematic face stimuli. Thus, in asking what subjects learn in these tasks, we must also ask how the subjects interpret and represent the stimuli to themselves.

The context-cue model proposed by Medin and Schaeffer (1978) attempts to address this problem. As noted earlier, each dimension of the stimuli is weighted according to its salience in the subjects' judgements. If variation along a particular dimension has little or no influence on subjects' judgements in a given task, then it is treated for all intents and purposes as constant, and, therefore, irrelevant. Since the similarity metric is a multiplicative combination of the dimensions, non-salient (or truly constant) dimensions drop out of the calculation. On the other hand, highly salient dimensions have the greatest impact on the assessed similarity between events. If, for example, subjects consider a particular dimension to be important and two stimuli differ along this dimension, the judged similarity between the two events will be low, even if the two stimuli match on all other dimensions. If, however, the two events match on this important dimension, the similarity between the two events will be judged to be quite high, even in those cases in which the two stimuli differ on many less salient dimensions. Thus, stimuli that differ on many less salient dimensions may be judged to be more similar to each other than stimuli that differ on only a few highly salient dimensions. This prediction also may be found with Tversky's (1977) model of similarity and may play an important role in the memory for individual instances as a model for structural learning, as is discussed below.

The evidence most often cited as supporting abstraction over memory for individual instances is the effect of a delay between training and the classification test. As mentioned earlier, the accuracy of classification of the training exemplars reduces more with a delay than does the prototype or new exemplars of the category. With long delays, classification accuracy of the prototype may equal or actually exceed that of the training exemplars. If the reduction in accuracy for training items is attributed to forgetting, then why is there not a similar reduction for the prototype and new exemplars if memory for individual instances is mediating their classification? Medin and Schaeffer (1978) answer this question in their context-cue model. They argue that forgetting of the training exemplars may be modeled by changes in the salience associated with each dimension; over a delay, the salience of each dimension of the exemplars in memory drops such that all items both within and across categories become more similar and, hence, more confusable with each other. Since the prototype of a category generally has more highly similar exemplars within its own category than does any given training item, and since it is the pattern that is least likely to have many highly similar exemplars in an alternate category, the prototype should suffer less from the general increase in the confusability of exemplars in memory than should the training exemplars. The conclusion is perhaps more easily seen from the perspective of what happens to the basis of classification of the training exemplars according to this model. As mentioned earlier, there are some aspects of this model that make it similar to the nearest neighbour model proposed by Reed (1972). Primary among these is the notion that, while the "evidence value" for a particular category for the classification of any given item is calculated on the basis of a ratio involving the similarity to items in all categories, any given classification response is a function of which items in memory are contacted. That is, the "evidence value" proposed by the model provides an estimate of the probability that at least one item of that category will be contacted in memory given some particular item to be classified, but is not, in itself, the basis for classification. According to the model, then, the high classification accuracy of the training exemplars on an immediate test is a function of the fact that the most probable item in memory to be contacted for any given training item is the memory for that exact item. As the confusability of the items in memory increases with the delay, however, the probability that a given training exemplar will contact the memory of itself is reduced, while the probability it will contact the memory of some other exemplar, including that of exemplars in a different category, is increased. Thus, the training exemplars are classified on a delayed test in a fashion similar to new exemplars on either the immediate or delayed test. Since, however, new exemplars, including the prototype, never did have the benefit of contact with memory traces of themselves, they suffer less over the delay than do the training exemplars. This notion of what happens to the classification of an item when it contacts an item in memory different than its a priori "closest" match also provides an alternative explanation for the Posner and Keele (1968) "matched similarity" experiment discussed earlier, and a number of other results obtained in experimental designs of this sort. This notion is discussed below following a review of a number of findings relating to the effect of delay.

A similar memory for individual instances explanation of the effect of delay as that of Medin and Schaeffer (1978) has been proposed by Hintzman and Ludlam (1981). They used a computer simulation in which features of the exemplars in memory were randomly "forgotten" and in which classification decisions were on the basis of the closest exemplar (as determined by a version of Tversky's, 1977, similarity model) in memory. With increasing forgetting cycles, they produced essentially the same results as those found in the actual experiments. Note that which features were "forgotten" from the memory representation of each exemplar on each forgetting cycle were determined randomly and not with respect to category diagnosticity or the prototype. Despite this, the effect of such forgetting, as in Medin and Schaeffer's (1978) explanation, is to decrease the probability that a training exemplar will be classified on the basis of its own memory trace, while, paradoxically, increasing the probability that it will be categorized on the basis of memory for some other exemplar, possibly from a different category. This forgetting, up to some point, can actually enhance the classification accuracy of the prototype, since it can remove differences between the prototypical stimulus and memory for some exemplar that previously prevented the memory for that exemplar from being a "close" analogy of the prototype, and since this is more likely to be the case for items from the same category as a given prototype than items from a different category. Aside from providing support for the individual instances model, these results underline the importance of considering the representation of the individual instances in memory. No model that assesses the similarity between events in terms of the stimuli as presented can model results and speculations such as these.

It is possible to eliminate the differential effect of delay on training items and novel exemplars by varying either how the subjects encode the training items or by altering the items that the subjects are asked to learn. Hock, Tromley, Marshall, and Webb (Note 2) performed a series of experiments in which they varied a number of aspects of the delay paradigm in an attempt to determine when subjects were responding on the basis of prototypes and when they were responding on the basis of memory for individual instances. After replicating the original Posner and Keele (1970) experiment in which they found the usual convergence in the classification accuracy of training and novel exemplars over the delay, they repeated the experiment, but this time, motivated by the work of Brooks (1978), had the subjects learn the items during the training phase as a paired-associate task rather than the usual category discrimination task. The training exemplars from each of four random-dot categories were each assigned a label. Exemplars from one category, for example, were assigned colour names, while exemplars from another category were assigned the names of cities. Following training, either immediately, two weeks, or four weeks later, the subjects were given a test containing training and novel exemplars in which the subjects' task was to classify the items into their four categories using the category names of the individual labels used during training (e.g., items that the subjects thought should have a colour name were to be assigned to the "colour" category). Unlike the replication of the original study, subjects were more accurate at classifying the training exemplars than the novel exemplars even after a four week delay. Thus, by having the subjects learn each training pattern with a distinctive label, there was no evidence that the subjects had abstracted and were using prototypes even after a delay of four weeks. Lest it be concluded that this finding is peculiar to the paired-associate training procedure, Hock *et al.* in another experiment found that essentially the same results occur with category discrimination training if each exemplar of a given category is constructed to be distinct from other exemplars of the same category. By randomly moving a *single* dot from the prototype of each category to produce the training exemplars and the novel exemplars used in the classification test, Hock *et al.* produced patterns of a given category that were *identical* to their prototype and each other on all but one dot, but, as a consequence of the change in the one dot, the overall configuration for each pattern was quite distinct from that of any other pattern. Subjects were given standard discrimination training followed by a classification test either immediately, or after a two or four week delay. Even after a delay of four weeks, the training exemplars were classified more accurately than were the novel exemplars, with no evidence that they would converge over even longer delays.

The results of the Hock et al. experiments clearly demonstrate the importance of considering the representation of the individual exemplars in memory. By making the representation of each training exemplar distinct from that of any other, either by presenting the subjects with distinct items or by having them distinctively encode each item, it was possible to eliminate the primary evidence for prototype abstraction in these tasks. While these results obviously support a memory for individual instances view of structural learning, they do not rule out the possibility that prototypes were being abstracted and used in the original experiments since it is possible that paired-associate training or highly distinctive items may prevent the abstraction process from occurring during training or being used during the classification test. This is the view presented by Reber and Allen (1978) for the implicit learning of artificial grammars, and is discussed in a review of their work in the final section of this chapter.

Medin and Smith (1981) also investigated the effects of encoding on performance in the learning of category structure. They had subjects learn exemplars from ill-defined categories of schematic faces under three different encoding conditions. One of these conditions was the standard discrimination training procedure. Subjects in the other two conditions were instructed to learn which faces belonged to which category either by a "rule plus exceptions" strategy which would correctly classify all faces but two that would have to be memorized, or by attempting to form a prototype of each category. The different encoding conditions produced differences in the learning of the items, the classification accuracy of training and novel exemplars, and the speeded classification of the training exemplars. For each of these tests, however, the differences between the three encoding conditions could be accounted for by differences in the similarity parameters associated with the features of the stimuli. That is, analyzed in terms of the context-cue model, the results were consistent with the notion that the representation of the training exemplars in memory differed across the three encoding conditions, but that in each case responding was better described in terms of memory for individual instances than an abstracted representation of the instances. These results are important since they suggest, as noted by Medin and Smith (1981), that different encoding instructions affect only what subjects attend to, and retain about each individual exemplar they encounter in these tasks, and not the basic process, analogy to individual instances in memory, that they use to accomplish the demands of the tasks presented to them.

There have been several investigations in the recent literature that have been interpreted as incompatible with memory for individual instances as an explanation for structural learning. One of these (Reed, 1978), concentrated on the learning of individual and categorical labels of exemplars from two categories of schematic faces. While the results of this investigation, contrary to the claims of Reed (1978), present no problem for an individual instances approach, the experiment is discussed here because it highlights one of the more common, but questionable, reasons for the <u>a priori</u> rejection of the individual instances view of structural learning.

The basic idea of the experiments reported by Reed (1978) was to compare the learning of unique labels to each face of the categories with the simultaneous learning of the category labels. On each trial, which consisted of the presentation of each of the five faces from each category, subjects were asked to respond with the label for each face and then the category label, followed by feedback. The basic result was that the subjects produced fewer errors in their category label responses, and reduced these errors at a more rapid rate over trials, than they did for the unique labels for each face. From the perspective of memory for individual instances, these results are not surprising. Aside from the fact that the two types of label responses differ in their guessing accuracy (i.e., 50% for category labels as compared to 10% for unique labels), the fact that the subjects increased their category label accuracy more rapidly than their unique label accuracy indicates that subjects were more likely to make a within-category error than a between-category error for unique label responses on each successive trial of ten faces. Since the faces within a given category were more confusable with each other than they were with faces in the alternate category, the better "category" learning than "item" learning, as Reed (1978) refers to these measures, may be simply the result of analogy to the imperfect memory of other exemplars in the same category.

In his analysis of the experiments, Reed (1978) attempted to fit a number of different prototype and attribute-frequency models quantitatively to the data, but he did not attempt to fit an instance model to the data because, as he puts it: "Since the average distance model and nearest neighbor model require discrimination among the exemplars within a category, these models can be rejected in those cases in which subjects are unable to make within-categories discriminations." (Reed, 1978, p. 613) Although he does not make explicit why this must be the case, it is implied that unless a particular instance is perfectly recognizable (as indexed by assignment of the correct unique label) in his task, specific memory for prior experiences with the instance can not influence the encoding and classification of either the instance itself or any other exemplar. It is not necessary to assume such a proposition.[2.7]

A milder version of Reed's implied contention is common to many theorists, and, as discussed earlier, is the basis for the prototype interpretation of the delay effect. This contention, however, has been invoked more generally. Fried and Holyoak (1982), for example, suggest that memory for individual instances would be a poor explanation if the stimuli were presented rapidly during acquistion. Homa et al. (1981), to take another example, question the application of Medin and Schaeffer's (1978) instance model to situations in which the subjects have not been taken to an errorless learning criterion since "... there is no assurance that the old patterns were stored in memory." (Homa et al., 1981, p. 419) Similarly, Omohundro (1981) used subjects' recognition (i.e., old - new discrimination) data obtained following a category discrimination task to estimate the number of training exemplars the subjects had in memory for categories that had differed in the number of exemplars presented during discrimination training. By her calculation, the amount of specific exemplar information in memory did not vary as a function of category size, making memory for individual instances an unacceptable explanation in her view for the category size effect. In all of these cases, it is implied that classification accuracy should be a monotonically increasing function of how well the individual training exemplars are recognized, and that, if the subjects demonstrate poor memory for the training instances (as assessed by classification or recognition accuracy), memory for training exemplars can not be mediating much of the classification of new (or even old) exemplars. In a later section of this chapter, as well as in the experiments reported in subsequent chapters, this model of the relationship between demonstrated memory for individual instances and classification is challenged and a model in which good memory for individual instances is associated with <u>reduced</u> classification accuracy is presented. For the moment, however, the discussion will be limited to the general contention that poor memory for individual instances can not mediate classification.

If subjects in a category discrimination learning task were presented with each stimulus only once, it would appear reasonable to believe that their memory for each instance would be quite poor. Consequently, according to the general contention discussed above, these poor memories should be relatively ineffective in mediating the classification of subsequent stimuli. Medin and Schwanenflugel (1981), however, reported a category discrimination learning experiment in which each schematic face was presented only once, and found the specific memory for these faces could mediate classification. The task presented to the subjects was to attempt to categorize each successive new face on the basis of the feedback to prior faces. Errors in classification decreased as a function of increasing trials. That is, subjects were more accurate on faces presented later in the learning sequence than those presented earlier, indicating that they were transferring information from their experiences with the earlier faces to their classification of the later faces. The important point, however, is that this was the same for subjects given categories of faces that were linearally separable (i.e., capable of being defined in terms of average or protypical exemplars) as it was for subjects given categories of exemplars that were categories only in the sense that a face that had a higher similarity to its "closest" face in one category rather than the other was assigned to that category by the experimenter. Since the only basis for classifying any given face into one category rather than another for this latter condition was its similarity to specific exemplars, the results suggest that subjects in both conditions were responding on the basis of memory for individual instances. Moreover, since each face was presented only once, this classification was occurring on the basis of probably poor memory for each individual instance.

The possibility that memory for items that subjects fail to discriminate from each other may still influence the encoding and classification of other exemplars provides an interpretation in terms of memory for individual instances for another finding that has been cited as evidence against memory for individual instances. Robbins et al. (1978) investigated the learning of the reversal of exemplar–response pairings in a variant of the delay paradigm. Specifically, subjects learned to classify exemplars from two ill-defined categories and then were presented with a classification test of the category prototypes, old and novel exemplars either immediately or after a 24 or 72 hour delay. This was followed with a half-reversal shift of the assignments of category labels to items. The delay had no effect on classification, but did on the learning of the reversal shift. Subjects reversed immediately produced more errors on changed than unchanged pairings. Subjects given the half-reversal task after a delay, however, produced the same number of errors for both changed and unchanged pairings, and produced more errors and took more trials to criterion than did the subjects tested immediately. These results, as concluded by the experimenters, are consistent with the notion that, over a delay, specific-exemplar knowledge decays and responding comes more under the control of an abstracted prototype. However, as noted by Hintzman and Ludlam (1981), these results also would be expected if, over the delay, subjects simply became less able to discriminate among the items. If this were the case, then the subjects, after the delay, would be less able to detect which pairings had changed and which had not, producing the obtained results.

The last two investigations to be discussed in this section used "matched similarity" designs similar to that of Posner and Keele (1968, Experiment III) discussed earlier. In the discussion of Posner and Keele's experiment, the problem of the adequacy of the similarity metric was raised. The problem applies equally well to the present investigations, but will be ignored in preference to another problem with "matched similarity"designs. This latter problem, also raised earlier, is concerned with what happens to the classification and recognition of a training stimulus, or a stimulus that is a "close" analogy of training stimulus, when it fails to contact in memory the representation of its a priori "closest" analogy.

Consider an experiment by Omohundro (1981) in which subjects were tested for their recognition of training and novel exemplars following category discrimination training. During training, subjects were presented with exemplars from three categories in which the number of training exemplars from each category (i.e., category size) was varied across categories. Thus, one category was represented by only 4 training exemplars, another by 8, and the third by 12. On the recognition test, which included an equal number of training and novel exemplars from each category, false-positive ("old") responses to novel exemplars increased as a function of category size. The rate of false-positives for the different category sizes paralleled the classification accuracy obtained in another phase of the experiment. That is, the usual increase in classification accuracy associated with increasing category size was obtained. While this relationship between classification accuracy and recognition false-positives is interesting and will be discussed in detail in subsequent chapters, the point of present concern is the design of the recognition phase of the experiment in which these results were obtained. Specifically, the intention of the experiment was to demonstrate that subjects would be more likely to misrecognize as "old" novel exemplars that were "close" to the prototypes of the categories than those that were not when both types of novel exemplars were matched for their similarity to specific training exemplars. To accomplish this objective, Omohundro (1981) produced two novel exemplars for each training exemplar of each category such that, while they were matched for their similarity to the training exemplar, one of them would be much closer to the prototype of the category than would the other. On the recognition test, these three items (i.e., the training exemplar and its two matched novel exemplars) were presented together and the subjects' task was to choose from among the three items the one they believed to be a training exemplar. Under these conditions, subjects produced more false-positive responses to the novel exemplar of each triplet that was closer to the category prototype.

Omohundro (1981) concluded that the subjects had abstracted categorical information during training that they were using as at least a partial basis of their responses in the recognition task. But note, false-positive responses in this task may arise only when the subjects fail to recognize the training exemplar. This means that the subjects were misrecognizing any given novel exemplar on the basis of memory for something other than the training exemplar in question. In terms of subjects' memory for individual instances, what was contacted in these situations was the memory for some other exemplar. And since the novel exemplar that was "closer" to the prototype would be, on average, "closer" to any other training exemplar than would its matched novel exemplar, failing to contact the "correct" training exemplar in memory would favor the more protypical of the two novel exemplars. As mentioned, this interpretation highlights another major problem with the use of "matched similarity" designs in these investigations since, even if the the similarity metric is appropriate, the "matched" stimuli are <u>not</u> matched on all variables save the one of interest. Less prototypical stimuli, even when matched for their similarity to one training stimulus with that of a more prototypical stimulus, will tend, on average, to be less similar to other training exemplars than will the "matched", but more prototypical stimuli. If responding is based on memory for training stimuli other than the a priori "closest" analogy, less prototypical stimuli will be at a disadvantage relative to the more prototypical stimuli. In fact, the present investigation provides probably the best evidence that this can and does happen since the false-positive responses could not occur in Omohundro's task unless the a priori best analogy in memory for a given recognition triplet was not contacted.[2.8]

There is another aspect of the Omohundro (1981) results that have a bearing on the next experiment to be discussed. In her experiment, false-positive responses increased as a function of category size. According to the present interpretation, the relationship between false-positives and category size implies that as subjects obtain more experience with a category (through more training exemplars) their responses to particular stimuli apparently depend less upon memory for the a priori "closest" exemplar. In addition, since, as mentioned earlier, the tendency to overgeneralize is correlated with increasing classification accuracy, it appears that having a larger number of exemplars in memory makes the memory for any given exemplar less necessary for correct classification, but also less likely to be the primary memory contacted if the item that generated the memory is presented for recognition or classification. There are at least two reasons why this may be the case. First, having more items in memory for a given category increases the probability that there will be at least one item to contact if an exemplar from the same or even another category is presented. Having more items in memory also would be expected to increase the degree of similarity between the memory for an exemplar and some presented item because increasing the number of exemplars in memory will also tend to more completely represent the range of exemplars possible for the category. Second, having a large number of exemplars in memory may alter the encoding of any subsequently presented items because these new items may be encoded in terms of the prior experiences which are activated when the new items are presented. This activation of prior experiences may have the effect of increasing the degree of encoded similarity between events in memory, rendering them more confusable with each other and, hence, more likely to be contacted in place of one another. Although this latter point can be modeled in terms of Medin and Schaeffer's (1978) context-cue model as an increase in confusability due to a change in the similarity parameters, it is probably more easily seen as an increased "fuzziness" in the encoding and memory for the stimuli. A phrase to describe the notion is increased breadth of transfer. Breadth of transfer, which is elaborated below and in the next chapter, is intended to capture the notion that the manner in which subjects represent the exemplars to themselves may have effects on the range or breadth of other events to which the information (including category label and familiarity) associated with a specific instance in memory may be generalized. Category size, as evidenced above, may be a variable that influences breadth of transfer by affecting the encoding and memory representation of each exemplar, as is suggested by the results of the next experiment to be discussed.

Homa et al. (1981) performed an experiment in which they attempted to ascertain the limits of instance-based knowledge as an explanation for the learning of category structure. As with Omohundro (1981) and Posner and Keele (1968), they used an experimental design in which items were matched for similarity. The major difference between Homa et al.'s experiment and the others, however, is that the stimuli were matched for their similarity to the prototype rather than to the training exemplars. The variable upon which the novel stimuli (excluding the category prototypes) for classification differed in Homa et al.'s experiment was in their similarity to training exemplars. All training and novel were constructed to be the same distance from their category prototype, and for each training exemplar, novel exemplars were used that ranged from being virtual twins of the training exemplar to being no more similar to the training exemplar than they were to the category prototype. The two other variables of interest in the experiment were category size and immediate or delayed classification test. The results, particularly for the small category size (five training exemplars), attest to the power of memory for individual instances. For all category sizes and for both immediate and delayed classification, accuracy of classification of novel exemplars decreased as function of increasing distance between the novel exemplar and its "closest" training exemplar. The effect of similarity distance, however, interacted with both category size and delay, such that the magnitude of the effect was reduced as a function of increased category size and the delay in classification. Moreover, while the category prototypes were the poorest classified for the small category size, their classification accuracy increased with increasing category size such that, for the largest category size (20 training exemplars), the classification accuracy of the prototypes equaled that of the training exemplars.

Homa et al. (1981) acknowledge that memory for individual instances played a large role in the classification responses of their subjects, but they emphasize that the interactions that they obtained provide support for the notion that with large category sizes and a delayed test subjects shift the basis of their responding from analogy with memory for specific exemplars to an abstract representation of the category. There are several reasons, however, why their explanation may be unnecessary. First, the fact that the classification accuracy of the category prototype increases with increasing category size may reflect nothing more than, with larger category sizes, there are more exemplars from the category in memory to assist in classifying the prototype. That is, increasing the category size may increase the probability that presentation of the prototype will contact the memory for at least one of its categorical exemplars, simply because there are more available from its own category rather than some other. The same may be true for the reduction in the magnitude of the effect of exemplar distance with increased category size. The diminishing effect of exemplar distance is manifested as a reduction in the slope of the function relating classification accuracy to exemplar distance. Increasing category size exerts its effect by increasing the classification accuracy of distant novel exemplars that, with smaller category sizes, were classified poorly. Again, the increase may occur simply because there are more exemplars in memory from the same category to be contacted. Moreover, unlike the prototypes for which all exemplars are equidistant, increasing category size for novel exemplars also increases the probability that there will be more "close" exemplars in memory to contact. The second reason involves breadth of transfer. Both category size and delay may increase the range of exemplars over which the memory for a particular training exemplar may generalize. Since this increase in breadth would increase the probability of contact for any given item, increasing breadth of transfer could facilitate the classification accuracy for the more distant novel exemplars and the prototype of the category.

While breadth of transfer can explain Homa et al.'s (1981) results entirely in terms memory for specific exemplars, it needs some independent empirical support. Fortunately, there is an effect that is associated with increasing category size that Homa and his associates acknowledge to be somewhat of a puzzle. Specifically, as classification accuracy increases with category size, so does the tendency of subjects to over-generalize the category, despite the fact that the subjects are aware that there are an equal number of exemplars of each category on the classification test. Although the effect can be detected by the depressing effect of over-generalization on the classification accuracy for the remaining categories, it is most easily detected by including in the classification test a set of exemplars that belong to none of the training categories. A set of these unrelated items was included in the Homa et al.'s (1981) study, and the tendency of subjects to classify these into one or another of the training categories increased as a function of both category size and delay. This is consistent with the notion that with increasing category size and time, the breadth of transfer of memory for specific exemplars increases to encompass a wider range of exemplars. Breadth of transfer is used again below to account for what, from the perspective of individual instances, appears to be a paradoxical result in the learning of artificial grammars.

Implicit Learning of Artificial Grammars

With many of the abstraction models that have been proposed to account for the learning of ill-defined concepts, the theorists are not claiming that their subjects are explicitly abstracting and using the underlying structure, any more than they would advance such a claim for the learning of natural categories which their artificial categories were designed to model. If such was the claim, it would be expected that most investigators would naturally ask their subjects to report the bases of their classification responses, as is commonly done in traditional concept formation research. Few investigators report having done so, however, and of the few that have (e.g., Elio and Anderson, 1981), the results have been relatively uninformative, and not considered to be of any great importance as evidence either for or against any particular model.

The work of Arthur Reber and associates on the learning of artificial grammars is an important exception in this regard. Reber and his associates (e.g., Reber & Allen, 1978) make a very explicit claim that their subjects are implicitly abstracting the underlying rules or structure of the grammars presented to them through their experience with grammatical instances. According to their view, the high level abstractions of linguistics and other complex cognitive domains are learned by an "... unconscious, non-rational, automatic process whereby the structural nature of the stimulus environment is mapped into the mind of the attentive subject." (Reber & Allen, 1978, p. 191). They argue that the process can be suppressed by explicit attempts on the part of the learner to abstract the underlying rules, or by tasks and procedures that direct the learner's attention to the idiosyncratic properties of the particular instances. In this way, they provide an account of how people handle the complexity of many stimulus domains, as well as the difficulty people have in explicitly forming and testing abstractions in concept formation experiments and formal education.

Over a series of experiments on the learning of artificial grammars, Reber and his associates have provided a wealth of evidence supporting their position. Most clearly supported is their claim that the knowledge that subjects gain in these tasks is largely implicit. Even in an experiment with hand-picked, highly motivated subjects, subjects' explicit justifications for their predominantly correct decisions were limited to relatively simple aspects of the stimuli and then on only about 40% of the test trials (Reber & Allen, 1978). Many of the other experiments had subjects claiming no explicit knowledge at all, and protesting, as a consequence, that they would be unable to accomplish the tasks set for them. However, they performed relatively well (e.g., Reber, 1976; Reber & Lewis, 1977). Moreover, as both Brooks (1978) and Reber (1976) have demonstrated with these materials, explicit attempts on the part of subjects to discover the complicated, grammatical rules, result in poor explicit knowledge and little or no implicit knowledge to use in subsequent tasks. Thus, the contention that attempts at explicit rule induction in complex domains disrupts the implicit process is supported. The primary question, however, is whether the process necessarily is one of implicit abstraction.

The materials that Reber and his associates have used in their demonstrations of implicit learning of structure are letter strings generated from artificial grammars similar to the one shown in Figure 4.1. The grammars are taken to be analogous to the regularity and structure that people must learn in natural languages. As Reber and Allen (1978) describe these materials, "... an artificial language presents the subject with a large and flexible store of stimulus materials which reflect a structure based on probabilistic and often remote contingencies which are not immediately discernable" (p. 190), and also not found in the artificial material generated for traditional concept learning tasks (e.g., Bruner et al., 1956).

In his earlier experiments with these materials, Reber (1967, 1969) used the free recall performance of his subjects to assess what they learned about the materials. Relative to a control group that received randomly generated letter-strings, subjects receiving items generated from the grammar demonstrated significantly better learning of letter strings presented later in the learning sequence. That is, similar to the results of Medin and Schawenflugel's (1981) experiment discussed earlier, subjects learning items from the grammar were transferring what they had learned from their experience with earlier letter strings to their learning of items later in in the learning sequence. Since the control condition did not evidence such transfer, it seems clear that what the subjects in the grammar condition were learning was related to the grammatical consistency of the items (Reber, 1967). In a related experiment, Reber (1969) demonstrated that subjects could transfer the learning of one set of items to the learning of letter-strings of similar structure, but composed of different letters. Transfer to the learning of letter-strings with a different underlying structure, but the same letters, or different structure and different letters was much less successful.

Thus, Reber concluded that it was the grammatical structure of the items that was mediating the transfer, and that his subjects were implicitly abstracting this structure from their experience with the letter-strings.

In later experiments, Reber and his collegues have assessed transfer using a variety of tasks. Following training which consisted of memorizing and recalling letter-strings from an artificial grammar in which no mention is made of the existence of the grammar, subjects can successfully discriminate novel grammatical strings from non-grammatical strings in a surprise classification test (Reber, 1976), and can solve anagrams composed of scrambled, grammatical letter-strings (Reber and Lewis, 1977). In all of these tasks, subjects can not account for their performance by citing explicit rules, and explicit attempts to discover rules during training, result in subsequent classification performance that is little better than that expected on the basis of chance (Brooks, 1978; Reber, 1976).

Brooks (1978) challenged the notion that subjects were implicitly abstracting the underlying grammatical structure in these tasks by presenting subjects, again with no mention of there being rules for the items, with letter strings from artificial grammars in a learning task in which it was virtually impossible for them to abstract the rules of the underlying grammar. However, his subjects could discriminate novel grammatical from non-grammatical letter strings on a subsequent surprise classification test, and also could discriminate from which of two grammars the grammatical letter strings had come. Brooks (1978) suggested that the most likely explanation for these and Reber's results was that subjects were examining each new letter string to see if it resembled any of the items they had been asked to memorize. If it did, they treated it as obeying the same set of rules as the similar remembered item and classified it accordingly. If some new item failed to resemble any of the items they had been asked to learn, then the item was called "non-grammatical". Since items from the same grammar would tend to resemble one another whereas non-grammatical items would tend not to resemble items from either grammar, a strategy of analogy to prior instances could successfully sort the items.

In a reply to Brooks (1978), Reber and Allen (1978) acknowledged that under some circumstances analogy to prior instances could account for subjects' ability to discriminate grammatical from non-grammatical items, and that this ability, by itself, does not provide unequivocal support for their position. They argued, however, that Brooks' (1978) results were obtained under special circumstances in which abstraction was not feasible, and that implicit abstraction of structure represented the more general case. They demonstrated their point by contrasting the performance of paired associate training with that of an observation training condition. The paired associate condition was like of one of the conditions used by Brooks (1978), and consisted of requiring that the subjects learn a unique response to each training stimulus. The training stimuli were letter strings from an artificial grammar. As in all of the previous experiments, subjects were not informed about the existence of the grammatical rules during training. The observation condition was described as "non-directive", and was intended as an analogy to the unstructured and incidental manner in which people apparently learn natural languages and concepts. In this condition, again no mention was made of the grammatical rules, and subjects were instructed simply to observe the letter strings as they were presented successively in a random order. Following training, subjects were given a classification test in which they were to discriminate some old, but mostly novel, grammatical letter strings from non-grammatical letter strings. The primary result was that the observation condition resulted in better performance on this task than did the paired-associate condition. From the perspective of Brooks' (1978) analogy explanation, this result is paradoxical since it implies that the observation condition led to better learning of the individual training stimuli than did paired associate training. Not only is this implausible on a priori grounds, it is also contradicted by the results of a recognition test in which the paired associate condition was found to lead to significantly better recognition of the training stimuli. If the observation condition did lead to poorer memory for the training stimuli, then it seems unlikely that it should have resulted in better performance on the classification test if subjects were using analogy to remembered training items as their basis for classification.

The two training conditions resulted in other differences as well. Following observation training, classification accuracy was more a function of correct responses to grammatical items than non-grammatical items. In fact, the two training conditions led to roughly equal classification accuracy for non-grammatical items. Moreover, although it was rare following either type of training, paired associate training led to more justifications for particular responses in which analogy to training stimuli was given as the reason than did observation training. Finally, paired associate training led to a greater degree of consistent (subjects classified the items twice) rejection of grammatical items than did observation training. This latter result was interpreted by Reber and Allen (1978) to suggest that paired associate training was more likely to lead to the apprehension of inappropriate rules than was observation training.

Reber and Allen (1978) discuss their results in terms of three different modes of aquiring and using structural information. The first, explicit rule induction occurs if subjects explicitly analyze the material in search of relatively simple, invariant aspects of the stimuli. If they succeed in isolating some relatively stable component of the training stimuli, they can use it to aid in the discrimination of grammatical and non-grammatical items because there is a reasonably good chance that many grammatical items, but not many non-grammatical items, will share this component. In Reber and Allen's (1978) analysis of the experiment, the strategy of explicit rule induction described at least part of their subjects performance on about 40% of the trials. For the most part, however, subjects' reported use of explicit-rule information was limited to describing the initial and terminal aspects of the letter strings, and then primarily for detecting non-grammatical items as violations of the grammar. This result is consistent with another finding of the experiment. Subjects were more likely to reject non-grammatical items corectly if the grammatical violation occurred at the beginning or end of a letter string than if it occurred in the middle. Second, in agreement with Brooks (1978), Reber and Allen suggest that at least part of their subjects' classification preformance could be attributed to analogy to individual instances, although they suggest that the strategy applies more to paired associate training than observation training. As mentioned, the explicit use of an analogy strategy occurred most often following paired associate training, and was associated with a high rate of false-rejections of grammatical items. This result suggests that there were many grammatical items for which no specific analogy was available in memory that, consequently, were erroneously labelled non-grammatical. Third, Reber and Allen argue that a large part of their subjects' performance was mediated by the implicit abstraction of the underlying grammar. They associate this process, primarily with the observation training condition in which, they argue, the subjects unconsciously abstract a partial, but veridical representation of the grammar from their experience with the training stimuli. The implicit representation is then used, again implicitly, to facilitate subjects' classification of the items. According to Reber and Allen's (1978) analysis, apprehension and use of implicit abstraction was relatively effortless and resulted in the greatest accuracy on the classification test.

Reber and Allen (1978) argue that the task requirements associated with attempting to learn a unique label for each training item in the paired associate condition disrupts the implicit learning process primarily by forcing the subjects to individuate each exemplar from its similar neighbours. Thus, the subjects respond on the basis of analogy to similar training exemplars because that is about all they can do. The observation procedure, on the other hand, by not requiring that the subjects discriminate similar items from one another, allows for the implicit abstraction of regularities across the exemplars. Following experience of this sort, subjects respond on the basis of the implicit application of these abstracted regularities to the new items. Like subjects given paired associate training, however, they are probably limited to the almost total use of one process, because they have poor memory for individual exemplars to use as analogies for new items. Since the abstracted regularities are veridical with respect to the underlying grammar, and since using these regularities frees the subjects from the capriciousness of item similarity and memory for specific events, classification accuracy following observation training should exceed that following paired associate training. Reber and Allen's (1978) evidence for this claim of implicit abstraction rests primarily on their finding that observation training does, in fact, lead to better classification performance than does paired associate training, although other differences between the two conditions are held to buttress the conclusion.

For the most part, I agree with Reber and Allen's (1978) proposal, particularly that what subjects learn and use in these tasks is a mixture of strategies and types of information. However, I disagree with their contention that their experiment demonstrates that different training procedures result in the use of qualitatively different processes – from the use of analogy to specific training exemplars in the paired associate condition to the apprehension and use of implicit abstraction in the observation condition. Challenging this claim of implicit abstraction with respect to Reber and Allen (1978) in particular, and the concept learning literature in general, represents the bulk of the work reported in this thesis. However, a few of the more pertinent points will be raised here.

Encoding and Breadth of Transfer

The strength of Reber and Allen's (1978) argument lies in the fact that observation training leads to poorer memory for individual training exemplars than does paired associate training, while resulting in superior classification accuracy. This result appears to contradict memory for individual instances as the basis of classification performance for the observation condition. However, there are at least two ways that a subject might fail to contact similar items in memory and, hence, fail to exploit the similarity of training and transfer items on the classification test. The first, and most obvious, is that the subject may have few items in memory. If so, then it would be unlikely that any given transfer item would find a close analogy in memory. This is the explanation that Brooks (1978) proposed for the poor transfer performance of subjects asked to look for rules during training. In their unsuccessful search for rules, the subjects failed to learn many of the individual items and, consequently, arrived at the transfer test without adequate rules or individual instances in memory to guide their classification of the items. The same failure to learn the training items adequately was cited by Reber and Allen (1978) to account for the high false-rejection of grammatical items following paired associate training. Since subjects given observation training, according to Reber and Allen's (1978) view, were using implicit abstraction, their even poorer memory for the training exemplars would not leave them susceptible to this failure to find an analogy in memory.

As mentioned in the previous section of this chapter, poor memory for instances coupled with reasonably good classification accuracy, particularly over a delay, has often been cited as evidence against analogy to memory for individual instances as an explanation in complex or ill-defined concept tasks. However, it was argued earlier in the discussion of Medin and Schwanenflugel's (1981) experiment, that subjects could use analogy to poorly remembered stimuli in their classification of novel stimuli. It was also mentioned that some of the effects that have been observed could be attributed to the breadth of transfer around what is known about specific exemplars. What was not discussed was the relation between these two notions. It is this relation that provides the second reason for a failure to exploit the similarity between memory for specific exemplars and the transfer stimuli.

Consider one possible consequence of good memory for specific events. As usually measured (e.g., recognition), this typically means that that the subject discriminates well between events. That is, the subject does not confuse some previously seen event with some novel, but similar event. But since no two events, even repetitions of the same nominal event, are ever truly identical in all respects, good recognition memory implies that the memory for some particular event has enough breadth to capture most presentations of the same nominal event, but is restricted enough to exclude all but the most similar new events. Good recognition memory, then, should be associated with reduced breadth of transfer. In contrast, poor recognition memory for specific events may be seen to be associated with increased breadth of transfer, possibly to the extent that most events in the same domain or context fail to be discriminated from one another. A mechanism for breadth of transfer differences will be discussed in the next chapter. For the moment, the point is that <u>poor</u> memory for specific exemplars may lead to <u>better</u>

classification accuracy of novel items than does good memory because the breadth of transfer of these poor memories may be broad enough to capture most new instances of the category, while narrow enough to exclude most non-instances.

The relationship between breadth of transfer and performance is complicated. At some high degree of breadth most new items might strike the subject as similar to those in memory resulting in little basis for rejecting any item and reducing classification accuracy because of a high false-positive rate. Too little breadth would hurt performance because the subject would reject most items, producing a low hit rate. Some degree of breadth, however, will be optimal for a given set of transfer stimuli. Applying this argument to the results of Reber and Allen's (1978) experiment, the observation condition results in superior classification accuracy not because of implicit abstraction, but rather because the breadth of transfer around the subjects' memory for the specific training stimuli was closer to the optimum for the set of classification stimuli they received than was the case for the subjects in the paired associate condition.

The present claim is that Reber and Allen's (1978) encoding manipulation affected the breadth of transfer of their subjects' knowledge and not, as they claim, the type of knowledge and processes that the subjects were using to classify the stimuli. There are a number of aspects of Reber and Allen's (1978) data that are consistent with this notion. Shown in Table 2.1 are the results, in a modified form, from the classification phase of Reber and Allen's (1978) experiment. First, if instead of accuracy, the statistic used by Reber and Allen (1978) in their analysis, we concentrate of the rate of responding that an item is grammatical, it can be seen that the two encoding conditions lead to different tendencies on the part of the subjects to accept transfer stimuli as "grammatical". Following observation training, subjects were more likely to label any given item as "grammatical" than

Table 2.1
Reber & Allen's (1978) Transfer Results

The results of the transfer phase of Reber & Allen's (1978) experiment adapted from Reber & Allen's Table 1. The results from each training condition for both grammatical and non-grammatical items are presented in two ways: once as the mean percentage of items of each type labelled "grammatical" by the subjects, and again as the mean percentage of responses of each type correctly assigned rather than, as Reber & Allen presented these data, the mean percentage of items of each type correctly labelled.

	Item Status					
	Grammatical		Non-grammatical		All Items	
Learning Procedure	Labelled "G"	Correctly Assigned	Labelled "G"	Correctly Assigned	Labelled "G"	Correctly Assigned
Observation-1st	87.2	77.9	24.8	85.5	56.0	81.2
Observation-2nd	82.0	79.5	21.2	81.4	51.6	80.2
Paired Associate-1st	68.4	74.3	23.6	70.7	46.0	72.4
Paired Associate-2nd	74.4	76.2	23.2	75.0	48.8	75.6

following paired associate training, particularly if only first-task performance is compared.[2,9] That is, observation training does appear to lead to increased breadth of transfer relative to paired associate training. Moreover, the aforementioned greater accuracy on grammatical items for observation training appears, to a large degree, to be a consequence of this increased positive response tendency. Reversing the conditionalization from the number of stimuli of a particular type correctly labelled to the number of responses of a particular type correctly assigned, it can be seen that paired associate training led to greater relative accuracy in the assignment of grammatical responses than did observation training. That is, in terms of assigning transfer responses to items, observation training accuracy seems to be associated with non-grammatical responses. Following paired associate training, however, subjects were more judicious in their acceptance of items as "grammatical", as would be expected if they were generalizing with reduced breadth of transfer around their memory for specific training exemplars (i.e., only accepting those items that were that much "closer" to those in memory and, hence, that much more likely to actually be grammatical).

There is an interesting addendum to this story. In a follow-up study (Allen & Reber, 1980) that occurred two years after the original investigation, eight of the original 10 subjects were asked to repeat the classification phase of the experiment. Remarkably, the subjects were able to classify the items with an accuracy of about two-thirds correct! While this result adds to the growing list of demonstrations of the robustness of memory for arbitrary material, the important point for the present discussion is that there were no differences between the two encoding conditions in either accuracy or the tendency to accept items as "grammatical". If Reber and Allen's (1978) interpretation of the original results is accepted, then the present results suggest that memory for individual exemplars is at least as robust as is an abstract representation of the items. But if abstractions are no less susceptible to the effects of forgetting over time than are memories for individual exemplars, then one of the more often cited advantages of abstraction loses its force.

Reber and his collegues have investigated the effects of other manipulations on the learning of artificial grammars. Reber et al. (1980), for example, found that by presenting the training stimuli such that the commonalities (common paths through the grammar) among the items were made more obvious, explicit abstraction could be made to result in superior classification accuracy in these tasks. These and other findings will be discussed in subsequent chapters as they relate to the results of my investigations.

Summary and Conclusions

The trend in concept learning research in recent years has been to focus on the learning of "ill-defined" category structure. This trend has been associated with a tendency to equate what people learn in these tasks with the structure of the domain in question. However, "ill-definedness" might better be described as a property of the manner in which people treat a complex stimulus domain than a property of the domain itself. It is clear, however, no matter how construed, that people are sensitive to the distributional structure of events that they experience and, through this, the distributional structure of some concept or domain as a whole. Despite the myriad of demonstrations of this ability and its primarily implicit effect on performance within concept learning, the evidence for it being the result of implicit abstraction of structure is at best questionable. As Hintzman and Ludlam (1981) noted in the closing line of their recent paper, "...unequivocal support for a prototype theory as opposed to an exemplar-based theory of classification learning still appears to be lacking." (p. 382). Relative to the findings of the literature as a whole, no one model that has been proposed, abstractive or otherwise, can adequately account for the entirety of systematic variation that has been found. At least part of the problem stems from some of the more questionable assumptions about similarity and memory that have been employed in tests of the various models. Rather than move in the direction of weak eclecticism, however, it is my belief that with the appropriate addenda, such as the notion of breadth of transfer, an individual instances approach to the implicit learning of categorical structure can provide a framework for the investigation of what it is that people learn in these tasks and how they use the information they obtain. The artificial grammar paradigm appears most easily adapted to this approach, and is the one used in the research to be reported. The first step in this direction requires the development of a classification learning paradigm that is less susceptible to the equivocality inherent in many of the experiments discussed. This is the subject of the next chapter.

Chapter 3

THE BASIC PARADIGM

Most of the research reviewed in Chapter 2 was an attempt to demonstrate that, through their experience with a limited number of categorical exemplars, subjects are sensitive to variations in the categorical structure reflected in subsequently experienced events. A common conclusion has been that this sensitivity is a consequence of the subjects' developing an abstract representation of the structure of the category from their experience with exemplars of the category. This conclusion has had an unfortunate biasing effect on the design and analysis of the experiments used to investigate the learning of categorical structure. With few exceptions, researchers have attempted to investigate what subjects learn by varying the structure inherent in the sets of items used both to train and test their subjects. If these structural differences are reflected in performance, as assessed in relation to the structural variation, then abstract knowledge of the structure, usually in the same form as that manipulated by the experimenter, is attributed to the subjects.

Categorical structure, however, is not a simple variable. Two items that differ in their relation to the structure of some set of items or the category as a whole will also tend to differ in other properties, especially their specific similarity to other items in the set or category. It is the possibility that subjects are responding to these covariates of categorical structure, and thereby indirectly reflecting structural variations, that renders the results of much of the work reviewed in Chapter 2 equivocal with respect to abstraction. Attempts to deal with these covariation problems, either by attempting to match items with respect to the suspected covariates while varying categorical structure, or by assessing the quantitative fit of models that emphasize different covariates of categorical structure to the data of a single experiment have been unsuccessful partly because there is little agreement as to how the covariates should be defined or measured.

In contrast to the difficulties that researchers have encountered in their attempts to demonstrate that subjects abstract and use at least some aspects of categorical structure, researchers attempting to demonstrate the effects of some of the covariates of structural variation, particularly the similarity between transfer items and specific training items, have had relatively little difficulty. Brooks (1978) and Medin and his collegues (Medin & Schaeffer, 1978; Medin & Schwanenflugel, 1981; Medin & Smith, 1981) have demonstrated the important role of memory for specific instances in concept tasks. However, as argued by Homa et al. (1981) and Reber and Allen (1978), the results were obtained in tasks in which abstraction was not feasible, either because few training items were used, or because the requirements of the training task biased the subjects' processing toward the learning of individual events. If strong evidence for the role of individual instances arises only in tasks with few training items or unique training procedures, the evidence favoring the individual instances model as a general explanation of what subjects learn and use in conceptual tasks is also not unequivocal.

The general problem associated with research in this area may be conceived as one that is in search of a paradigm for its resolution. Instance theorists object to the research involving structural variations as demonstrating very little because of a confounding with the specific similarity between training and transfer items. Abstraction theorists object to the research cited by instance theorists because the results were obtained in situations that would be expected to mitigate against the apprehension and use of abstraction. This chapter develops a basic paradigm involving the learning of artificial grammars that I believe answers the major objections of theorists on both sides of the issue. This basic paradigm, with modifications and addenda, is used in all of the experiments reported in subsequent chapters.

The Basic Problem: Covariation

Attempts to separate performance arising from the memory for individual instances from that arising from some abstract representation of the category often include a confounding between the categorical status of any given exemplar and the similarity between that item and some one or more "known" or training exemplars in memory. A near twin of some member of a category is more likely to be a member of that category than some other category for which no near twin exists. Consequently, unless these two variables can be uncounfounded, any evidence of categorical transfer is equivocal. The covariation problem is exacerbated by the related problem of how to define or measure the similarity between events since similarity is more a property of how subjects encode and represent events than a property of the stimuli that are used. Given that the functional relationship between a response measure and the similarity between items is unknown for most tasks, a posteriori attempts to adjust performance in conceptual tasks to control statistically for similarity (e.g., assessing the quantitative fit of various models to the data) are unlikely to be convincing. The same is also true for experimental designs that attempt to control for the similarity between events by matching test stimuli on the basis of some a priori metric for their similarity to training items. Unless the metric is perfectly correlated with the similarity between items as used by the subjects, the finding of reliable correlation between categorical status and the response measure remains equivocal.

There are several ways to demonstrate the covariation problem in these tasks.[3.1] The one presented here is in terms of what has been called Simpson's

Paradox (Simpson, 1951). Simpson's paradox is a special case of covariation problems that arise from the collapsing of multidimensional frequency tables in the presence of non-additivity (see Bishop, Feinberg, & Holland, 1975; Fienberg, 1977). It has received recent attention in the psychological literature (Hintzman, 1980; Flexser, 1981) because of the common practice among memory researchers to construct and analyze contingency tables that have been collapsed across subjects (the experimental unit) and items (the experimental sub-unit) where the marginal distributions of the variables of interest are determined by the subjects and/or items rather than fixed by the experimenter. To facilitate transfer to the experiments in subsequent chapters, the example is presented in terms of subjects' classification of grammatical and non-grammatical items in an hypothetical artificial grammar experiment, although the arguments apply equally well to the investigation of the learning of concepts of any form.

Consider the categorical transfer materials used by Reber and Allen (1978). One-half of these items were constructed to contain violations of the artificial grammar used to generate the training items and the other half of the transfer items. Thus, Allen and Reber (1978) viewed the transfer list as consisting of samples of two populations of items: items that were grammatical according to the artificial grammar, and items that were non-grammatical according to the same artificial grammar. The transfer items, however, may be divided into two sets of items based upon their similarity to specific training items. Specifically, we can imagine separating the transfer items into those that are similar to specific training exemplars, the "close" items, and those that are relatively dissimilar to any training item, the "far" items.[3.2] The question is whether these two different ways of dividing the items are orthogonal to each other. In Reber and Allen's (1978) experiment, there is good reason to believe that these two variables were not orthogonal. Aside from the general argument that members of any given, complex category are more likely to resemble other members of the category (i.e., the training items) than are non-members, many of the operations used by Reber and Allen (1978) to construct their sets of transfer items probably increased this general tendency. For example, 20% of the grammatical transfer items were items actually experienced by the subjects in the training phase and, hence, were not just similar to the training items, but identical to them. In the other direction, 40% of the of the non-grammatical transfer items differed from the grammatical transfer items by having grammatical violations in the initial two letters of the items, while another 40% had these violations in the terminal two letters. Studies of proofreading (e.g., Erhlich & Rayner, 1981; Haber & Schindler, 1981) have indicated that relative to errors in deep, internal letter positions, spelling errors in the initial and terminal letters of words are most easily detected. Since Reber and Allen (1978) found the same pattern of results for the detection of grammatical violations in their experiment, it appears that changes in these positions are highly salient and, hence, likely to be more heavily weighted in subjects' assessment of the similarity between items. Another 12% of the non-grammatical items were grammatical items "spelled backward", rendering them very dissimilar to any of the training items. In short, almost independently of how the similarity between training and transfer items is to be assessed, it appears that the "close" items in Reber and Allen's (1978) materials would be composed primarily of grammatical items, while the "far" items would be predominantly non-grammatical items.

An idealized version of the confounding between the grammatical status of transfer items and the close-far similarity between training and transfer items is shown in the top two tables in Figure 3.1., in which 80% of the "close" items are grammatical and 80% of the "far" items are non-grammatical. Given the confounding,

CLOSE ITEMS

Item Type	Response G	Response NG	
G	64	16	80
NG	16	4	20
	80	20	100

FAR ITEMS

Item Type	Response G	Response NG	
G	4	16	20
NG	16	64	80
	20	80	100

ALL ITEMS

Item Type	Response G	Response NG	
G	68	32	100
NG	32	68	100
	100	100	200

the question becomes one of its effect on the observed relationship between categorical status and the response measure if, as in Reber and Allen's (1978) analysis, similarity is ignored. Suppose that responding that an item is grammatical or non-grammatical for either "close" or "far" items is entirely independent of the grammatical status of the items, and that subjects produce most of their "grammatical" responses to "close" items and most of their "non-grammatical" responses to "far" items, as is shown in Figure 3.1. For convenience, equal numbers (100 in each case) of grammatical and non-grammatical items, "close" and "far" items, and "grammatical" and "non-grammatical" responses have been used. Note that in both the "close" and "far" sub-tables, $P(R=G|S=G) = P(R=G)$, where R indicates response, S indicates item type, and G indicates grammatical. That is, in neither table is there any contingency between categorical status and response. Yet, as is shown in the bottom table in Figure 3.1, the table produced by collapsing over "close" and "far" items demonstrates a large, positive (and spurious) contingency between categorical status and the response measure, corresponding to a "percentage correct" of 68% and a correlation of $r = 0.36$. That is, in the collapsed table, $P(R=G|S=G) > P(R=G)$, an effect that, because of the equal marginal frequencies, is not a consequence of either a stimulus or a response bias. It also should be noted that the effect occurred with the correlation between the response measure and whether or not an item was "close" or "far" being only $r = 0.60$.

In most of the recent concept formation research, the relationships between categorical status and similarity, and responding and similarity are simply unknown. For any given concept formation experiment, in the absence of any knowledge about the relationships between categorical status and specific-item similarity and between the response measure and specific-item similarity, little may be said about any observed relationship between categorical status and the response measure, regardless of the reliability and direction of observed relationship (see note 3.1). Similarly, little may be said about any differences in the observed relationship across treatment conditions. Thus, the fact that Reber and Allen (1978), for example, reported a reliable interaction between their encoding conditions and the observed relationship between grammatical status and the response measure is not conclusive about the source of the interaction. The results may have occurred independently of any changes in, or even the existence of, a direct relationship between categorical status and the response measure. In the experiments reported in the next chapters, it is demonstrated that, through a judicious choice of transfer stimuli, one can produce a negative, a positive, or no relationship at all between the categorical status of the items and the subjects' responses depending upon the experimenter-induced correlation between specific-similarity and categorical status.

The solution to the problem of covariation requires experiments in which both variables, categorical status and specific-item similarity, are orthogonally manipulated. Unfortunately, a major problem in attempting to do this for a number of the stimulus domains that have been used (i.e., "ill-defined" or "fuzzy" categories) is that it is not clear how to measure either variable such that all potential confounding between them within the sets of stimuli that are used may be removed.

One partial solution to the problem is to use well-defined categories, like those produced with artificial grammars. In this way, at least the categorical status of any given transfer stimulus is precisely known. Any measure of specific-item similarity that is orthogonal to the well-defined categorical status will be independent of the categorical status of the items. Unfortunately, this does not mean that any resultant effect of categorical status in the experiment reflects a true measure of a direct relationship between categorical status and the response measure. As is argued throughout subsequent chapters of this thesis, and demonstrated in Experiment 4, any effect of the categorical status of the items on the response measure may reflect a form of residual similarity that is not captured by the similarity variable used in the experiment. What the orthogonal manipulation does mean, however, is that any effect of the similarity variable will be independent of the categorical status of the items.

Some researchers in the area might object that the cost of obtaining even this partial degree of control is too excessive, since it requires that the experiments be conducted with well-defined rather than ill-defined categories - a case of throwing away the domain of interest in the service of "scientific" rigour. However, as I argued at some length in Chapter 2, the emphasis on "ill-definedness" as a property of concepts may have been misplaced. If, as I suggest there, "ill-definedness" is better viewed as a property of the manner in which people deal with a complex domain, rather than the domain itself, the advantages to obtaining a greater degree of experimental control using well-defined categories, outweigh the possible costs of not using ill-defined categories. At the very least, the experiments to be reported directly address the implicit abstraction interpretation of the learning artificial grammars.

Specific Similarity

It is difficult to determine the basis of categorical transfer because it is difficult to separate the properties of interest of a given transfer item (e.g., its categorical status and its similarity to training items) from the transfer item itself. Consequently, if the properties in which we are interested are inherently correlated within the sets of items that we use, any result remains equivocal. If we rely upon item variation to produce our experimental "manipulation", as is commonly done in classification learning research, the problem remains. A solution is to hold the items constant and vary their experimental status through some extra-item, experimental manipulation. While this is not possible for the categorical status of items (i.e., with a well-defined category such as an artificial grammar, an item either is or is not a grammatical item), it is possible to manipulate at least some aspects of the similarity between any given transfer item and the training items, independently of the set of transfer stimuli used. Three different methods are used in the experiments reported in this thesis. One of them, the manipulation of <u>specific similarity</u>, represents the core of the basic paradigm, and was used in all of the experiments. The other methods are discussed in the next section.

As discussed earlier, the transfer items in Reber and Allen's (1978) experiment may be divided into those items that are "close" to the training stimuli and those that are "far". This segregation of the items is unlikely to be independent of their categorical status. However, it is possible to produce a second <u>training list</u> of items from the same grammar that reverses the orginal "close" or "far" status of each item such that those transfer items that were "close" to the training items in the original list, are now "far" or dissimilar to the training items in the new training list, and those that were originally "far", are now "close". Thus, the "close" transfer items for a subject trained with one list would be the "far" transfer items for a subject trained with the alternate list. Across subjects, each transfer item would serve equally often as a "close" item and as a "far" item <u>independently</u> of its status as a grammatical or non-grammatical item; each item would serve as its own control.[3.3]

A large effect of the "close-far" manipulation does not guarantee that at least part of the subjects' categorical judgements arose from non-abstractive processes. Although the "close-far" manipulation would be independent of the generative grammar, it may not be independent of some abstracted grammar that is peculiar to a given training list. One way to ensure a large "close-far" effect, for example, would be to produce training lists that were remarkably different from each other. Since a subject sees only the one training list, any representation of the items, abstractive or otherwise, would be likely to be quite different than that arising from experience with the alternate list. All that a large "close-far" effect would indicate under these circumstances is that what a subject learns is a function what items were experienced during training. For this reason, it is necessary that, with the exception of being composed of different items, the two training lists be as closely matched as possible with respect to the grammar. However, the more the two lists are matched with respect to the grammar, the more likely it will be that transfer items that are "close" (or "far") to one list also will be "close" (or "far") to the other. Indeed, this is one of the major arguments for the ecological validity of individual instances. But the attendent problem is that not only is it highly likely that there is a high degree of covariation between categorical status and training-transfer, inter-item similarity for natural experiences, it is almost impossible to eliminate it in controlled experiments.

The specific similarity manipulation that is used throughout the experiments that follow, incorporates the "close-far" manipulation produced by using a constant set of transfer items and two, matched training lists, but includes a specificity component that effectively circumvents the problem just discussed. The solution is that each transfer item, when it serves as a "close" item, is similar to one and only one training item. In all other respects, including similarity to the remaining training items, "close" items are virtually identical to "far" items. Thus, if subjects demonstrate a preference for any particular transfer item when it is used as a "close" item this preference must be a consequence of experience with its "close" training item and not any general property peculiar to the training list or grammatical items as a whole, since, if that training item were removed, the transfer item would be a "far" item.

The precise details of the specific similarity manipulation are given in the procedure sections of the experiments. In general, though, the similarity between any two items was defined as the number of common letters in position. A "close" transfer item, whether grammatical or non-grammatical according to the artificial grammar used, is one that differed from its closest training item by a single letter in position. "Far" transfer items of a given training list differ from all of the training items of that list by at least two letters in position, as do the "close" items to all but their closest training item.

Encoding Operations

One of the most important findings of Reber and Allen's (1978) experiment is that variations in how the subjects learn the training items influences the effectiveness of the information thus gained for categorical transfer decisions. In terms of breadth of transfer, this occurs because certain ways of encoding the training stimuli are more appropriate for the tasks and materials set for the subjects than are others. For example, Reber and Allen's (1978) paired-associate condition proved more effective than did their observation condition for a test of item recognition, but was less effective for a test of categorical transfer. The result is important for our present purposes because it suggests several methods of investigating the relationship between training and transfer independently of the exact transfer stimuli used.

Two different types of encoding manipulations were used in the experiments to be reported. The first of these may be referred to as global or general encoding operations, similar to that used by Reber and Allen (1978) and to memory research (e.g.,rote vs. elaborative rehearsal, etc.). The basic idea here is to manipulate how the subjects encode the training items as a whole, and then to assess the effect of this manipulation on some test of memory or transfer. A major problem with this approach is that it is difficult to localize the source of the encoding effect. Thus, for example, Reber and Allen (1978) attributed the effect of their encoding manipulation on categorical transfer to a shift from implicit abstraction in one condition to memory for specific training stimuli in the other, whereas, as I argued in Chapter 2, the effect also may be attributed to variations in only a single process; namely, the breadth of transfer of the memories for specific training items. One partial solution to this problem is to track the effects of some global encoding manipulation across a number of different tasks. Any training-task interactions that arise may be informative in terms of the source or sources of the encoding effect for the original task of interest. This was done for all of the experiments to be reported, and is discussed below.

The second approach has more in common with recent research on episodic memory (e.g., Light & Carter-Sobell, 1970; Hunt & Ellis, 1974; Donaldson, 1981), and requires varying the encoding of specific items during training and separately at testing. The manipulation is a <u>local</u> encoding operation which is directly concerned with the importance of <u>encoding specificity</u> as a source of transfer. The idea is that by manipulating the encoding of items across training and test, it should be possible to enhance or reduce the degree or breadth of transfer from specific training items to specific transfer items. As implemented (see Experiments 5 and 6), the manipulation was designed to emphasize that "similarity" and, hence, transfer, is more a function of how the subjects encode and represent the items than how the various items are related to each other on strictly formal grounds.

The next section discusses the basic training procedure and transfer tasks used in each of the experiments to be reported. Included is a discussion of the dependent measures of transfer used in these tasks. In discussing these measures, an individual instances approach that relates the various tasks and measures to each other in terms of breadth of transfer is discussed.

<u>Experimental Tasks and Measures</u>

The basic experimental design used in all of the experiments to be reported consists of three phases. In the first phase, subjects are asked to learn, by various encoding procedures, a list of training items selected from those produced by an artificial grammar. No mention is made of the underlying grammar, and, as far as the subjects are aware, their only task is to learn the training items according to the encoding task they have been given. Following training, the subjects are given either a surprise recognition test or a surprise categorical transfer test, depending upon the experiment, followed by the one of these two tests not yet received. As discussed in Chapter 2, subjects' performance on each of these three tasks has been interpreted to provide evidence for implicit abstraction. Their inclusion within the design of a single experiment allows for the investigation of the possible relationships between them. To anticipate the results of the experiments, these tasks do appear to be related in an orderly, systematic fashion, but implicit abstraction does not appear to be the explanation.

<u>The Training Phase: Item Learning</u>

The first phase of each experiment consists of a learning or training phase. The general procedure is the same as that used by Reber and Allen (1978), and contrasts with the discrimination training procedure used in most concept learning research.[3,4] In both Reber Allen's (1978) paradigm and the experiments reported in this thesis, the subjects are requested to learn a list of training items, generated from an artificial grammar, according to one of a variety of encoding procedures. To institute the specific similarity manipulation for the subsequent tests of transfer, one-half of the subjects are trained with one list of training items, while the remainder are trained with a list of different training items that is matched with the first list with respect to the rules of the grammar.

In each experiment, the training items in a given training list are divided into several sub-lists. The items from each sub-list are presented in random order for a number of trials before the next sub-list is presented, and the subjects are asked to study and then free recall the sub-list items on each trial. Thus, the number of items correctly recalled per trial may be viewed as arising from three different systematic sources: (1) the effect of trials, or the repetition of the same sub-list items, a measure I refer to as <u>intra-list</u> transfer; (2) the effect of sub-list, or the effect of prior experience with other sub-lists of items, a measure I refer to as <u>inter-list</u> transfer; and (3) the interaction of these two factors. These three sources are logically independent in the sense that it is possible, for example, for subjects to demonstrate increasing recall as a function of trials (i.e., intra-list transfer) while demonstrating no transfer from one sub-list of items to the next.

We have reason to be interested in the patterns of item learning that subjects demonstrate during training for a test of categorical transfer since training performance has been used as evidence for abstract learning. Reed (1978), for example, has argued that the learning of individual items as individuals occurs independently of, and more slowly than, the learning of individual items as exemplars of a category, the latter of which he attributes to the abstraction of structure. Similarly, Reber (1967, 1969, 1976) has argued that list to list transfer during training is indicative of the apprehension of abstract structure. The notion that both Reber and Reed seem to have in mind is that the increment in recall from one trial to the next of the same nominal items (intra-list transfer) is indicative of exemplar learning, while the increment from one list of items to another list of items from the same category (inter-list transfer) is indicative of abstract structural learning. More generally, according to the general contention discussed in Chapter 2, it would appear that most theorists would predict that if inter-list transfer (either within the learning task from one sub-list to the next, or across the learning task and the transfer task) is a consequence of exemplar knowledge, as instance theorists contend, then inter-list transfer should be positively correlated with, or at least a monotonically increasing function of, how well-learned the training, or first list items are. That is, intra- and inter-list transfer should be positively related.

The breadth of transfer notion discussed in Chapter 2, however, suggests a different relationship between item learning and inter-list transfer. Inter-list transfer is a measure of transfer from the learning of one set of items to the learning of a different set of items. Thus, differences in this measure for the same sets of items (say, across different encoding conditions) may be interpreted in terms of differences in the breadth of transfer. The greater the breadth (up to some point) the greater the degree of inter-list transfer that should be observed. Intra-list transfer, in contrast, is a measure of the transfer from one experience with a given item to another experience with the same item and, as such, provides a measure of the extent to which the breadth of transfer is restricted to the same nominal event. Consequently, in contrast with the general contention discussed earlier, intra-list transfer and inter-list transfer, by the present interpretation, should be inversely related. To the extent, for example, that one encoding condition relative to another leads to good intra-list transfer during training, the degree of inter-list transfer for that condition should be reduced. Similarly, the degree of transfer from training to categorical transfer and recognition should also be reduced. Thus, breadth of transfer differences as indicated by intra- and inter-list differences during training should manifest themselves in both recognition and categorical transfer tests.

The major factor responsible for breadth of transfer differences during training, as I see it, is <u>encoding variability</u>. This component of repetitive exposure to the same nominal stimulus has been proposed to account for such phenomena during recall as the spacing effect (e.g., Glenberg, 1977), and has been variously considered both to facilitate (e.g., Anderson & Bower, 1972) and hinder recall (e.g., Young & Bellezza, 1982) relative to consistent encoding. Rather than review the research concerned with encoding variability on recall (see, e.g., Hintzman, 1976), I will discuss simply some likely relations between encoding variability and measures of breadth of transfer.

Errors in free recall during training can occur in at least two different ways. First, and most obviously, subjects may fail to contact the representation of a given item in memory and, hence, fail to recall anything about the particular item. Second, subjects may recall only certain (correct) aspects of the item while failing to recall others, or recall as components of the item other (incorrect) aspects that arise from the subjects' elaborations, inferences, or confusions with other items. These latter components may be seen to be associated with increased encoding variability since they are concerned with what the subject does with a given item from one repetition to the next. For example, inter-item confusions would be expected to occur if the subject encodes the current presentation of an item in terms of the item immediately preceding it. Thus, if repetitive experience with a given nominal item is represented in memory as multiple traces (e.g., Hintzman & Stern, 1978), increased encoding variability would be expected to be associated with increased differences among the memory traces of the item, only a few of which would be expected to lead to veridical recall. Consistent encoding across repetitions, in contrast, would be expected to be associated with relatively similar memory traces of a given item and, hence, unless the subject consistently misrepresents the item, veridical recall.

However, consistent encoding would be expected to have a cost associated with it. Specifically, while it should lead to veridical recall, reducing recall errors due to inter-item confusion and the like, consistent encoding would also be expected to lead to a greater sensitivity to variations in encoding context. That is, variation in the recall cues from training to test would be expected to have a greater detrimental effect on the recall of items consistently encoded than on those whose encoding was relatively more variable (e.g., Tulving & Thomson, 1973). Variable encoding, then, should generalize to more recall situations than consistent encoding, while leading to less veridical recall. Consistent encoding should lead to veridical recall, but generalize to fewer recall situations. In breadth of transfer terms, consistent encoding should be associated with reduced breadth of transfer, and variable encoding with increased breadth of transfer. Thus, consistent encoding during training should be associated with increased intra-list transfer, but reduced inter-list transfer, while variable encoding should be associated with increased inter-list transfer since the greater variability of the traces in memory should increase the probability of their being contacted when new, but highly similar items are presented in subsequent lists. This relationship between encoding variability and breadth of transfer is discussed in more detail in Experiment 1, where two encoding conditions, expected to differ in encoding variability during training, evidenced such an inverse relationship between intra- and inter-list transfer during training.

Categorical Transfer: "Grammatical" Responses

In each of the experiments in this thesis, the categorical transfer phase is the first phase in which the subjects are informed about the existence of the underlying grammatical rules. As in Reber and Allen's (1978) and experiment, the subjects' task is to discriminate novel items with respect to their grammaticality. The dependent variable used in the tests of categorical transfer is the rate of responding "grammatical" to different types of transfer stimuli. Using the subjects' relative frequency of responding "grammatical" allows for a direct comparison of the magnitude of the effects of the categorical status (i.e., grammaticality) and specific similarity of the items. To the extent that subjects use similarity to the memory for specific instances rather than some abstracted version of the grammar as the basis of their transfer judgements, the effect of specific similarity on the relative frequency of "grammatical" responses should exceed that of the grammaticality of the items.

Given an effect of specific similarity on subjects' responses, the concern is one of the effect of encoding on the relationship between specific similarity and responding. According to Reber and Allen's (1978) implicit abstraction hypothesis, the effect of the encoding manipulation in their experiment was to shift the basis of responding from memory for individual instances in the paired-associate condition to the (implicit) use of some abstract representation of the grammar in the observation condition. Thus, the difference in performance between the two encoding conditions was seen as a difference in sensitivity to the grammatical status of the items. As well, had specific similarity been systematically manipulated in Reber and Allen's (1978) experiment, it would be expected that specific similarity would have been associated more with the responses of subjects given paired-associate training than subjects given observation training. In contrast, according to the breadth of transfer notion, the actual categorical status of the items is irrelevant, and the difference in performance between the two encoding conditions arose as a consequence of differential senstivity to the (unmeasured) specific similarity of the transfer items, but in the opposite direction to that suggested by Reber and Allen's (1978) explanation. According to this notion, the observation procedure during training would be expected to lead to greater encoding variability of the items than would the paired-associate procedure, primarily because the associates in the latter condition provide a consistent, unique context within which to encode each item. As discussed above, this should result in a greater breadth of transfer of the memory for the individual training exemplars following observation training. The effect of this for a test of categorical transfer should be that more of the transfer items should strike observation-trained subjects as similar to the training items and, hence, likely to be grammatical. Importantly, however, this increased tendency to be seen as similar to the training items should occur primarily for "close" transfer items since on a priori grounds they are the more confusable with the training exemplars. Thus, observation training should be associated with larger effects of the specific similarity manipulation than should paired-associate training and, more generally, should lead to a higher rate of "grammatical" responses to the transfer items, as in Reber and Allen's (1978) experiment. Since variations in breadth of transfer are not tied to Reber and Allen's (1978) particular encoding manipulation, the experiments to be reported compared a number of different encoding conditions, expected to differ in breadth of transfer, for their sensitivity to the specific similarity manipulation.

Recognition: False-positive Responses

In each of the experiments to be reported, subjects were given a test of item recognition. Each test was constructed to contain the items with which a subject was trained and a sample of categorical transfer items. The sample of transfer items was selected to conform to the orthogonal structure of the categorical transfer test. In fact, with the exception of the presence of the training items, each recognition test may be thought of as a reduced version of a categorical transfer test disguised as a test of recognition.

Breadth of transfer would be expected to manifest itself in two ways during a test of recognition. First, following training with an encoding procedure associated with decreased breadth of transfer, subjects should demonstrate a reduced tendency to accept recognition items as "old". That is, as with the acceptance of items as "grammatical" in a test of categorical transfer, reduced breadth of transfer should lead to a tendency to accept as "old" only those items that are that much "closer" to the training items in memory. Since the "closest" items are, in fact, the training items, this reduction in responding should occur primarily as a reduction in false-positive responses to non-training items, rather than as an overall reduction in the tendency to respond. Second, as in categorical transfer, subjects should be more likely to respond positively to (i.e., mis-identify as "old") "close" items than "far" items, and this tendency should covary with breadth of transfer.

In general, subjects' performance on the training tasks, their distribution and rate of "grammatical" responses on tests of categorical transfer, and their distribution and rate of "old" responses during recognition should be related in terms of the breadth of transfer of the memories for individual training stimuli induced by the encoding procedure used to learn the items. To the extent that these relationships are demonstrated, particularly in relation to specific similarity, the necessity to invoke implicit abstraction of categorical structure as an explanation for performance in these tasks is reduced.

Chapter 4

EXPERIMENT 1

The first experiment to be reported was an attempt to replicate the categorical transfer results reported by Reber and Allen (1978), using the basic paradigm discussed in the previous chapter. There are three variables of primary interest: the grammatical status of the transfer items, the specific similarity of the transfer items, and the encoding procedure used during the learning of the training items. It was expected that specific similarity would prove to be a greater determinant of responding than would grammaticality, and that the encoding manipulation would exert its effect primarily through affecting the subjects' sensitivity to the specific similarity rather than the grammaticality of the transfer items.

The experiment used three encoding conditions rather than the two used by Reber and Allen (1978), in an attempt to localize the effective difference between encoding conditions. It is Reber and Allen's expressed belief that their results as well as those of Brooks' (1978) investigations are a consequence of the task demands associated with paired-associate training. The implication is that the results are unique to paired-associate training. However, without contrasts other than observation training, it is not clear what aspects of paired-associate training are responsible for the reduced categorical transfer performance of subjects given such training, or even whether paired-associate training is necessary to produce the major effects they observed. In his earlier work, Reber (1967, 1969, 1976) used a free recall procedure to train his subjects. Since the results obtained from these studies provided the initial empirical support for Reber's implicit abstraction hypothesis, a free recall training procedure is a natural control condition by which to contrast the effects of observation and paired-associate training. It contrasts with observation training by requiring that the subjects memorize the training items, and contrasts with paired-associate training by instituting this requirement in the absence of the additional requirement to associate or otherwise memorize identifiers for each item.

In the present experiment, then, two memory conditions were contrasted with Observation training. One of them, No Label training, required that the subjects memorize the items, but no associates were presented with the items. The subjects' task was simply to study and then recall the items, as in Reber's earlier experiments. The second memory training condition, Overlapping Mnemonic training (the name "Overlapping" is used to contrast with mnemonic conditions in subsequent experiments), was a variation of the paired-associate training procedure used by Reber and Allen (1978) and Brooks (1978). The subjects were required to study and then recall the training items and an associated phrase of each training item. It was not a paired-associate task in the usual sense. The items and their associated phrases were presented and recalled together; the associated phrases were not learned as responses to the items.

According to the implicit abstraction hypothesis, the differences between encoding conditions observed during categorical transfer are a function of a shift from responding on the basis of implicitly abstracted rules in the observation condition to responding on the basis of memory for individual instances in the paired-associate condition. As discussed earlier, the encoding effects that Reber and Allen observed may have been a consequence of quantitative variation across encoding conditions of only one process. In particular, their results may be accounted for in terms of differing degrees of breadth of transfer across their two encoding conditions. Categorical transfer is only one of a number of tasks that should be sensitive to differences in breadth of transfer. Intra- and inter-list transfer during training, and the rate and distribution of false-positive responses during recognition provide additional measures. Accordingly, all subjects in the present experiment received three tasks, a training phase, followed by a categorical transfer phase, followed by a final recognition test. Each task was performed independently of knowledge about any subsequent task.

Method

Materials

The materials consisted of two training lists of 8 items each and a 64-item transfer list. These are shown in Table 4.1. All of the grammatical items were selected from the pool of all possible items, 3 to 8 letters in length, generated from the artificial grammar depicted in Figure 4.1. Since the items differed in length, there are two ways in which an item could be considered to be a "close" match of another item according to the criteria outlined earlier. Two items were considered to be "close" to one another if one item was identical to another except for the <u>substitution</u> of one letter (eg., VXM and VXR) or one item was identical to another except for the <u>addition</u> of one letter (eg., VXM and VXRM). Both types of "close" items were used. The training items were selected from the grammatical items that were 3, 5, and 7 letters in length. Consequently, the substitution transfer items were 3, 5, and 7 letters in length while the addition transfer items were 4, 6, and 8 letters in length. Since one of the primary concerns of the experiment was to demonstrate an effect of specific similarity, an item was chosen as a training item only if it was at least two letters in position different from any other training item. In this way it was possible to generate 16 triplets of closely matched items from the pool of grammatical items such that each transfer item would be "close" to one and only one training item.[4.1] Each triplet, then, contained a training

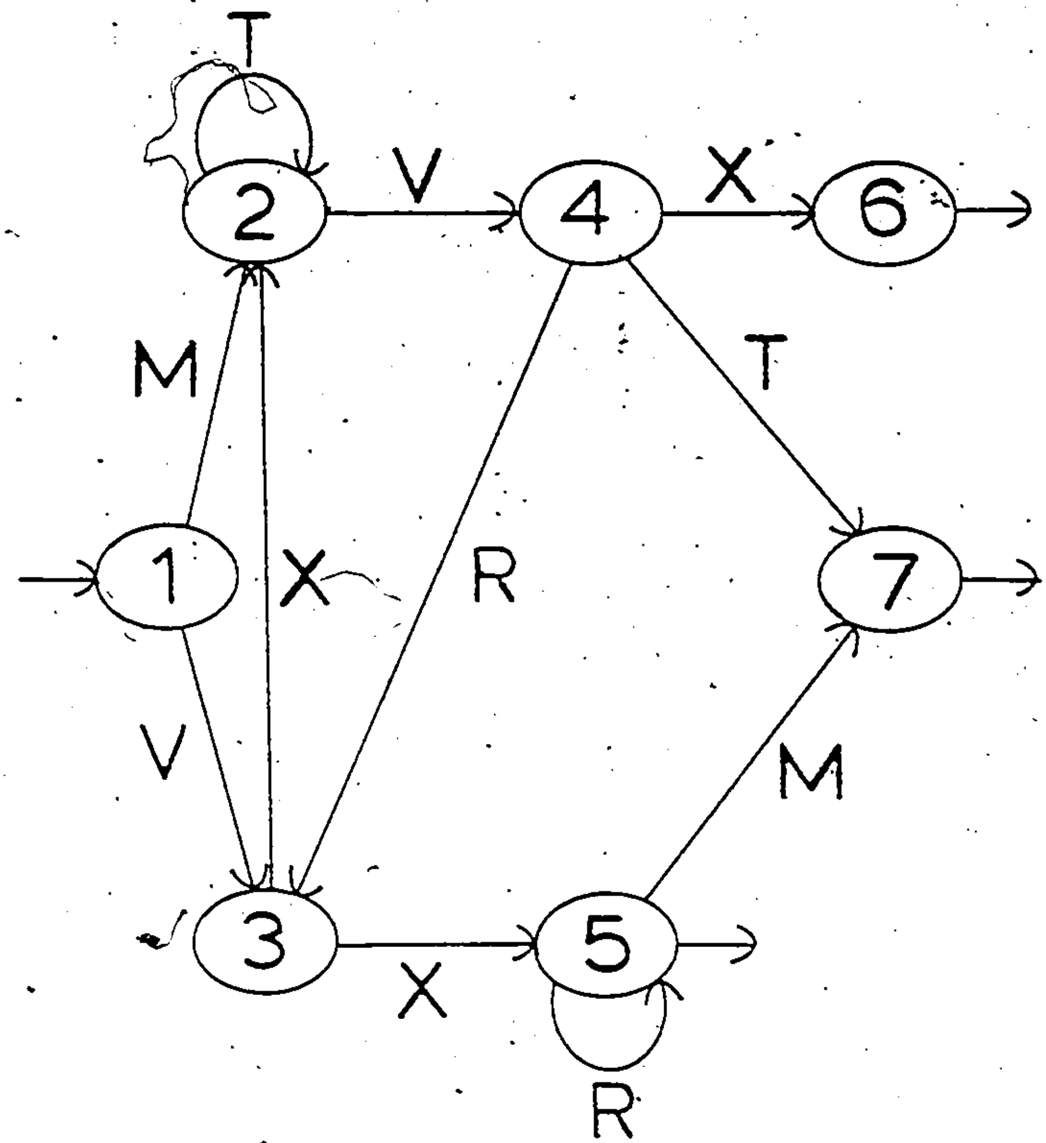

Figure 4.1: The artificial grammar used to generate the items in Experiments 1, 2, and 3. A grammatical string of letters is produced by any sequence of moves from node 1, following the arrows, to any exit node (5, 6, or 7). For example, the letter-string VXTTTVT is grammatical since it can be produced in this way (i.e., 1-3-2-2-2-2-4-7), while MXRVRTT is non-grammatical since it cannot be produced as an ordered sequence from this grammar.

Table 4.1
Training and Transfer Items

The training and transfer items used in Experiments 1, 2 and 3. One-half of the subjects were trained with training list one, and the remainder with training list two. All subjects were tested, twice, with all 64 of the transfer items. Transfer items in the same row as a given training item are the "close" items for a subject trained with that item. Transfer items associated with the alternate training list are the "far" items for a subject trained with a given list.

Training Items		Transfer Items			
		Substitution		Addition	
		Gram	Non-gram	Gram	Non-gram
List 1	VXM	VXR	VXT	VXRM	VXTM
	MVRXR	MVRXM	MVRXX	MVRXRR	MVRXRT
	VXTVX	VXTVT	VXTVR	VXTTVX	VXTRVX
	MTTTTVT	MTTTTVX	MTTTTVM	MTTTTTVT	MTTTTRVT
	MTTVRXM	MTTVRXR	MTTVRXT	MTTTVRXM	MTTRVRXM
	VXRRRRR	VXRRRRM	VXRRRRX	VXRRRRRR	VXRRRRRT
	VXTTTVT	VXTTTVX	VXTTTVM	VXTTTTVT	VXTTTRVT
	MVRXTVX	MVRXTVT	MVRXTVR	MVRXTTVX	MVRXTRVX
List 2	MVX	MVT	MVR	MTVX	MRVX
	MTTVT	MTTVX	MTTVM	MTTTVT	MTTRVT
	VXRRM	VXRRR	VXRRT	VXRRRM	VXRRTM
	MVRXRRM	MVRXRRR	MVRXRRT	MVRXRRRM	MVRXRRTM
	MTVRXRR	MTVRXRM	MTVRXRX	MTVRXRRR	MTVRXRRT
	MTVRXVX	MTVRXVT	MTVRXVR	MTTVRXVX	MTRVRXVX
	VXTVRXM	VXTVRXR	VXTVRXT	VXTTVRXM	VXTRVRXM
	VXVRXRR	VXVRXRM	VXVRXRX	VXVRXRRR	VXVRXRRT

item, a substitution transfer item and an addition transfer item, all of which were grammatical according to the grammar.

The corresponding non-grammatical transfer items for each triplet of grammatical items were produced as is shown in Table 4.2. For example, the grammatical substitution item VXR differs from its matched training item VXM by an R in the third letter position. The non-grammatical substitution item was produced by replacing the R in VXR with a T to produce VXT. Similarly, the non-grammatical addition transfer item VXTM was produced by replacing the R in VXRM with T. Note that a T occurring in this position is illegal according to the grammar.

The 16 training items were divided into the two training lists shown in Table 4.1. These lists were constructed such that the lengths, the initial and terminal letters, and the grammatical sub-rules of the items were reasonably matched across the two lists. This was done in an attempt to produce two training lists that would be equally representative of the generative grammar. Each training list was randomly divided into two sub-lists of four items each, and five random orders of the items of each sub-list were generated.

The 32 substitution transfer items were randomly divided, twice, into two 16 item lists, as were the 32 addition transfer items, yielding two pairs of 16 item lists for both substitution and addition transfer items, for a total of eight transfer lists. Across the pairs of 16 item lists, then, each item occurred twice.

Two final recognition lists were constructed, each containing 8 training items and a sample of ~~16~~ transfer items. One recognition list contained the 8 list 1 training items, while the other contained the 8 list 2 training items. The sample of 16 transfer items for each recognition list were chosen so as to maintain the orthogonal structure of the transfer lists. Thus, two items for each combination of grammaticality x alteration type (substitution or addition) x specific similarity

Table 4.2
Production of Non-grammatical Items

The non-grammatical substitution operations used to produce the non-grammatical transfer items for Experiments 1, 2 and 3.

Where the grammatical transfer stimulus differed from its closest training stimulus by a:	The non-grammatical transfer stimulus differed in the same letter position by a:
X	M
M	X
T	R
R	T

("close" or "far") were used. Except for satisfying these restrictions, the items were chosen at random.

Three different encoding conditions were used. For two of them, Observation and No Label training, the training items were presented with no modification. For the Overlapping Mnemonics training condition, each training item was associated with a mnemonic phrase. These mnemonic phrases, shown in Appendix A, were constructed such that the associated training items were acronyms of their respective phrases. For example, the list 1 training item MTTTTVT was associated with the phrase "Montreal's Thousands Take The TV Times", while the list 2 training item MVX was associated with the phrase "Mandy Viewed X-rays". These phrases were constructed to partially overlap with at least one other phrase in both thematic content and the words used, although no attempt was made to relate the consistency of the phrases to the grammatical consistency of the associated items. By making the associated phrases less distinct, it was hoped that they would not necessarily destroy any chance that might otherwise have been possible to see common patterns across the training items. The two oppositely extreme operations, of which this is a mixture, are discussed in Experiments 2 and 3.

Subjects

Thirty-two McMaster University undergraduates served as subjects in each of the three encoding conditions, resulting in a total of 96 subjects.[4.2] All subjects participated in the experiment to fulfill a course requirement in introductory psychology. From 4 to 20 subjects were tested at a time in a classroom setting. The subjects in the two memory conditions were deliberately mixed within a session. The subjects in the observation condition were tested separately in order to control the time that they had to study the items.

Procedure

All instructions and materials were presented to the subjects in booklets. For the training phase, the pages of the booklets were back-printed to mask the contents of the previous and subsequent pages from the subjects. The subjects were asked to read the instructions on the first page of the booklets and to start at a signal from the experimenter. These instructions described the training task to the subjects, and cautioned them to work on a page at a time and to not flip back to previous pages.

One-half of the subjects received training list 1 while the remainder received training list 2. Which sub-list of a given training list was presented first was counter-balanced across subjects within a given training list, and the order of presentation of the five random orders of a given sub-list was systematically rotated across subjects within sub-list and training list. The subjects received all five randomizations of one sub-list before proceeding to the next sub-list. Between sub-lists the training instructions were repeated and the subjects were informed that they would be receiving a list of four, new items. At no time during the training phase were the subjects informed about the number of sub-lists they would receive, or that their experimental task consisted of anything more than the training task.

Observation training consisted of allowing the subjects to study the four items on each page for 90 seconds, and then having them advance to the next page upon a signal from the experimenter. Subjects in the No Label training condition were asked to study the four items on each page and then to print all of the four preceding items that they could remember on a following blank page. Rate of presentation was subject-paced. Subjects in the mnemonics condition were similarly treated except that they were also required to print the mnemonic associates of the training items. This latter requirement was to ensure that the subjects would attempt to use the mnemonic phrases during training.

Following training, all subjects were presented with the categorical transfer test. The subjects read instructions to the effect that all of the items they had seen during training were constructed according to a complex set of rules and that their task was to indicate which of the new transfer items were probably constructed according to the same set of rules. As in the observation and memory conditions in Reber and Allen (1978) and Brooks (1978) this was the first mention of the rules that underlay the letter strings. Each transfer item was presented twice, for a total of 128 separate judgements, although the subjects were not informed of this fact. Each 16 item transfer list was presented on a separate page, and each item was preceded by a blank in which the subjects were to indicate their judgement of each item. The subjects were instructed to place a checkmark next to each item that they believed to obey the rules and to leave all other items blank. One-half of the subjects in each training condition received the substitution transfer test and re-test before the addition transfer test and re-test, while the remainder received these tests in the opposite order. The sequence of items within each transfer test was fixed across subjects.

After completion of the eight, 16-item lists of the transfer test, all subjects received one of the two recognition tests that was appropriate to the training list they had received. The eight training items and the sample of 16 transfer items were randomly ordered on a single page. The subjects were instructed to indicate which of the 24 items were the items they had seen during training. At no time were the subjects informed about the frequency of the various alternatives in either the categorical transfer task or the recognition task.

Results and Discussion

Training

The booklets for the two encoding conditions (No Label and Overlapping Mnemonics) for which training responses could be collected were scored for the number of training items correctly recalled. These data were subjected to mixed analysis of variance with subjects nested within encoding condition. The combination of 5 training trials for each of two sub-lists resulted in 10 cells per subject. The various between-subject counterbalancing factors were not tested. All effects were tested for reliability at the 0.05 alpha level, the significance level used for all experiments. The analysis of variance summary tables containing mean-square error terms and obtained probability values for this experiment as well as those that follow, are presented in Appendix C.

The mean percentages of items correctly recalled per trial per sub-list are shown in Figure 4.2 for each of the two encoding conditions. For both training conditions, the mean number of items correctly recalled increased significantly over training trials. More items were correctly recalled on the last trial of each sub-list (78%) than on the first trial (43%), although this increment due to trails was significantly larger for the first sub-list (trial 1 =35%; trial 5 = 76%) that the subjects received than the second (trial 1 = 50%; trial 5 = 80%). There was also a significant main effect of sub-list. Subjects in both training conditions correctly recalled significantly more items per trial from the second sub-list they received (71% correct) than the first (59% correct). While in line with the results in Reber's (1967, 1969, 1976) earlier work, this result contrasts with the memory training results in Reber and Allen (1978) in which no significant improvement was found across lists of different training items.

Collapsed across both trials and sub-lists, the mean number of items correctly recalled did not differ significantly across the two encoding conditions. Subjects in both conditions attained essentially equivalent levels of performance.

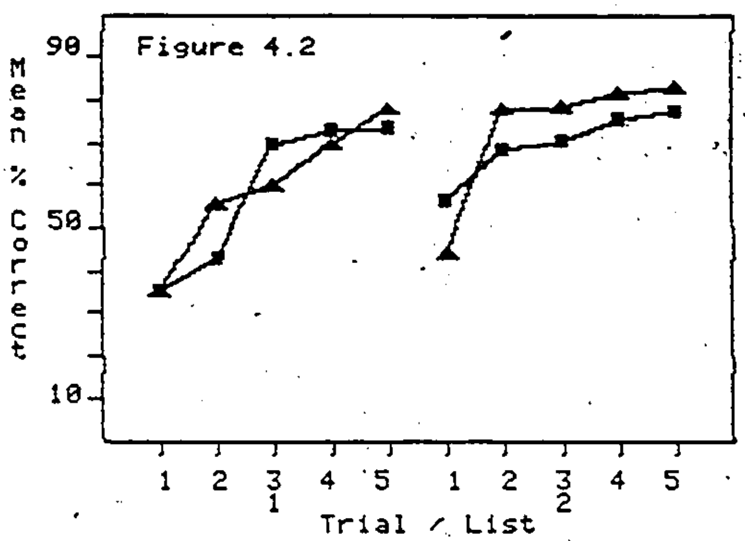

Figure 4.2: Training Phase. Mean percentage of items correctly recalled as a function of trials and sub-list for No Label (squares) and Overlapping Mnemonic (triangles) training.

They did so, however, for entirely different reasons. As indicated by a significant training condition by trials interaction, subjects in the mnemonics condition achieved their overall level of performance primarily as a function of intra-list transfer. That is, for both sub-lists, subjects in the mnemonics condition demonstrated a greater increment due to trials than did subjects trained without mnemonics. Since the overall level of performance was statistically identical for both conditions, it follows that subjects in the No Label condition achieved their level of performance more as a function of inter-list transfer than intra-list transfer. This is borne out by the finding that encoding condition was significantly related to the trials by sub-lists interaction mentioned above. The greater improvement due to trials on the first sub-list than the second was more pronounced for the No Label condition than the mnemonics condition. Thus, performance on the second sub-list for the No Label condition was more a function of transfer from the first sub-list than was true for the mnemonics condition. Consequently, while both encoding conditions evidence intra- and inter-list transfer, the relative contribution of each to overall performance differs significantly across the two training conditions.

These results suggest that the consequences for recall of repeating the same nominal stimulus vary as a function of the task requirements associated with attempting to memorize a given stimulus. Overlapping Mnemonic training, relative to No Label training, appears to enhance the degree of transfer from one encoding of a given item to the next, while limiting the breadth of transfer across different items. These differences in performance between the two training conditions may be conceptualized as arising from differences in encoding variability. The notion here is that the encoding of any given item in the mnemonics condition should be relatively stable from one trial to the next since the mnemonic associated with the item provides a consistent context for encoding the letters of the item. Thus, any advantage in specific item recall accruing to consistent context should be enhanced. Moreover, the mnemonic associated with each item is relatively unique in the sense that there is a smaller degree of overlap among the mnemonics, both intra- and inter-list, than there is among the items with which they are associated. Consequently, if inter-item confusion is a source of recall errors, these errors should be reduced for the mnemonic condition relative to the No Label condition. There is a cost associated with this increased encoding specificity and reduced inter-item confusion, however. Unlike the No Label condition, any benefits associated with the consistency across items will be lost, or at least reduced, in the mnemonics condition. Thus, in learning the items of the second sub-list, subjects in the No Label condition are able to transfer more of what they learned on the first sub-list than are subjects in the mnemonics condition. That is, the inter-item confusion that may hinder the trial to trial performance of the same items for subjects in the No Label condition may enhance their performance on a list of new items by providing them with less to learn about each new item. In the present experiment, the advantages of encoding stability and reduced inter-item confusion in providing performance gains from item repetitions for subjects in the Overlapping Mnemonics condition were neatly counter-balanced by the performance gains provided by an increased ability to generalize across items in different lists for subjects in the No Label condition. Note that this interpretation of inter-list transfer, unlike that of Reber (1967, 1969, 1976), attributes inter-list transfer to how the subjects distribute their experiences with individual training items rather than to abstraction of the underlying grammar.

The breadth of transfer interpretation of the training phase results provides at least two possible explanations of the lack of inter-list transfer with paired-associate training in Reber and Allen's (1978) experiment. The unique label associated with each training stimulus in that experiment and the requirement to individuate the items sufficiently to assign the correct label to each item may have been sufficient to remove any degree of inter-list transfer. That is, Reber and Allen's subjects may have performed in a manner similar to what would be expected from an extreme version of mnemonic training in the present experiment. This possibility is investigated in Chapter 5. Alternatively, as is discussed in more detail in Chapter 6, the high degree of simlarity between items across lists may have led to too much inter-list transfer, reducing performance in a manner similar to the intra-list performance of the No Label subjects in the present experiment. While this is clearly possible, it seems unlikely for Reber and Allen's (1978) paired-associate condition since it implies that their subjects did not attend to the associated labels, whereas, from the fact that they were able to assign the labels correctly to 50% of the items (with chance assignment at 1/25) by the end of training, it is clear that the subjects did attempt to fulfill the requirements of the paired-associate task. Thus, the former explanation of reduced breadth of transfer seems more likely.

The differences observed for the training conditions in the present experiment suggest a number of consequences for the categorical transfer and recognition phases of the experiment. The fact that the No Label subjects appear to have a greater degree of inter-item transfer than do the Overlapping Mnemonics subjects suggests that such will also be the case for transfer from the learning phase to both the catergorical transfer and the recognition phases. That is, No label subjects should demonstrate both a higher rate of responding that an item is grammatical in the classification task and a higher rate of false-positive responses in the recognition task than do Overlapping Mnemonic subjects. Moreover, the fact that mnemonic associates do not occur with the items in either of these latter tasks, suggests that the lower rate of positive responses expected in both tasks for the mnemonics condition will be further reduced since, without the stabilizing presence of the mnemonics, encoding of the items should be more variable than that during training leading to a further reduction in contact with the training experiences. On the other hand, the variability in the encoding of the training items for subjects in the No Label condition would be expected to leave them with more potential traces as "close" matches to the encoding of the categorical transfer and recognition items, thereby increasing their positive response rate.

Similar arguments may be advanced for the subjects in the Observation condition. If Observation training is considered to be a more extreme version of No Label training (i.e., high encoding variance of each training item leading to a high degree of inter-item transfer), it would be expected to lead to the highest rate of positive responses on both categorical transfer and recognition, and the greatest sensitivity to the specific similarity of the items. That is, subjects in the Observation condition would be expected to show the greatest breadth of transfer in these tasks. Note that these predictions are based upon the subjects' expected memorial representation of each of the individual items they experienced during training, and not upon differences in some abstraction of the underlying grammar.

Specific Similarity and Grammaticality

The dependent variable used in the analysis of categorical transfer was the mean frequency of items labelled as "grammatical" or as "obeying the rules" by the subjects. These data were subjected to a mixed analysis of variance with subjects nested within encoding condition. The orthogonal manipulation of the four within-subject factors (specific similarity, grammaticality, alteration type, and pass) resulted in 16 cells per subject.[4.3] The various between-subject counter-balancing factors (training list, test order of alteration type, and order of sub-list during training) were not tested. The major results of the categorical transfer phase of the experiment are summarized in Table 4.3, which presents the mean percentage of items labelled "grammatical" by the subjects as a function of encoding condition, and the grammatical status and specific similarity of the items.

The specific similarity of transfer items to training items was not varied over an extreme range in this experiment. There is nothing in these materials as deviant as the 12% of Reber and Allen's (1978) transfer items that were grammatical items "spelled backward", and none of the training items were repeated in the transfer test. This was done deliberately to prevent large specific similarity differences from overwhelming any potential effect of the grammaticality of the items. Thus, except for being formally similar to one specific item of the training stimuli, "close" transfer items are no more similar to the training items of a given training list as a whole than are the "far" transfer items. However, the specific similarity of the transfer items was by far the largest effect in the experiment. The effect of grammaticality, while significant, was much smaller. The ratio of estimated variance components of the two effects indicates that the relative effect size of specific similarity was almost 12 times that of grammaticality – a highly significant difference as is indicated by comparison of the partial correlations with responding for the two variables ($z = 5.52$).[4.4] This difference in the relative size of the two effects is reflected in the magnitude of the mean difference in the percentage of "grammatical" responses produced by the two variables. As is shown in column 5 (G–NG) of Table 4.3, grammatical transfer items increased responding by an average of only 4.3% over that of non-grammatical items. In contrast, as is shown in column 6 (C–F) of Table 4.3, "close" items recruited an average of almost 14% more "grammatical" responses than did "far" items. The importance of the specific

Table 4.3
Categorical Transfer Phase

The mean percentage of items labelled "grammatical" for each training condition as a function of the specific similarity and the grammatical status of the items. Column G-NG refers to the mean percentage difference of positive responses between grammatical and non-grammatical items. Column C-F refers to the mean percentage difference of positive responses between close and far items. The column labelled Actual Pc refers to the mean percentage of transfer stimuli correctly labelled. Pseudo Pc refers to the mean percentage of close grammatical and far non-grammatical items correctly labelled. Relative Frequency refers to the mean percentage of items receiving positive responses.

	Close		Far		G-NG	C-F	Actual	Pseudo	Rel.
	Gram	Non-G	Gram	Non-G			Pc	Pc	Freq.
Observation	43.9	37.7	26.3	22.8	4.9	16.3	52.4	60.6	32.7
No Label	37.3	33.5	20.8	18.5	3.1	15.7	51.5	59.4	27.5
Overlapping Mnemonics	33.5	29.3	24.6	18.8	5.0	9.7	52.5	57.4	26.5
	38.2	33.5	23.9	20.0	4.3	13.9	52.2	59.1	28.9

similarity variable is attested to by more than its large main effect. It also interacted significantly with all of the variables in this experiment except grammaticality, while grammaticality did not interact significantly with any of the remaining variables either alone or in combination.

It is important to remember at this point that, unlike grammaticality, the specific similarity of the transfer items involves a comparison of responding to the same items across subjects. Thus, the large effect of specific similarity cannot be attributed to a confounding of specific similarity with the generative grammar, and can only be a result of the specific experience of the subjects with a particular training list. This does not mean, however, that, therefore, the specific similarity effect is necessarily independent of abstract, grammatical knowledge. It could be argued that since each subject was exposed to only 8 items from the grammar during training, and since the "close" items for any given subject would tend to mimic the structure of these training items, the subjects had abstracted a grammatical structure that rejected more "far" transfer items than "close" transfer items. While this argument appears plausible, it suggests that the items in the different training lists were remarkably different in structure. That is, for subjects given training with one list to have abstracted a grammatical structure that rejected the "close" items for subjects trained with the alternate list, the two lists would have to have given rise to quite different abstracted rules to produce the obtained results. Since the items in the two training lists are composed of exactly the same letters and since the two lists were matched for both initial and terminal letters of the items, it is clear that if the two lists gave rise to different abstracted structures, these differences must have occurred at a level beyond that of individual letters or the more mundane types of positional constraints. At the level of bigrams, which was the degree of abstraction that Reber and Lewis (1977) suggested as the level that corresponded to the bulk of their subjects' abstracted knowledge, the two training lists barely differ at all. In fact, the complete set of bigrams that describe training list 1 is a subset (amounting to 12/14) of the bigrams that describe training list 2. Similarly, two-thirds of the trigrams that characterize list 1 items also characterize list 2 items. The trigram level and any higher-order approximations to any hypothetical grammars underlying the two lists begin to reflect differences across the two training lists that occur because of the specific individual items composing each list which, of course, is the point of the specific similarity manipulation. Thus, at least in this regard, the training lists could not have produced abstracted rules that differed to the degree required to account for the specific similarity effect in this experiment.

There is a further aspect of the design of the present experiment that seriously challenges the notion that subjects were using <u>any</u> general structure (e.g., prototypes, feature lists) as the primary basis of their categorical decisions. In fact, the results suggest that an average distance model (i.e., responding that an item is grammatical if its mean similarity to all of the training exemplars in memory exceeds some criterion) is also an unlikely explanation for the results of the present experiment. Remember that a "close" transfer item is referred to as such not because it is similar to a list of training items as a whole, but rather because it is similar to one and only one training item. "Close" transfer items differ from "far" transfer items not by being similar to any general characteristic of a training list, but by being similar to a specific item of the training list. In all other respects, including similarity to the remaining items in a training list, "close" items are virtually identical to "far" items. Hence, any advantages in recruiting "grammatical" responses that accrue to "close" items must have occurred as a consequence of the specific as opposed to the general experience of the subjects on a given list.

The most neutral conclusion that may be made with respect to a combination of a large effect of specific similarity and the small effect of the grammaticality of the items, is that grammaticality is a poor measure of whatever it is that subjects "know" about the grammar. If, as in the present experiment, the grammaticality (or any general summary of categorical status) of the items is unconfounded with the degree of specific similarity between training items and transfer items, its effectiveness as a predictor of the subjects' transfer performance is extremely limited.

This latter point is clearly demonstrated by assessing the accuracy of the subjects' responses. As is shown in column 7 (Pc) of Table 4.3, the 96 subjects in this experiment correctly labelled 52.2% of the transfer items, which is just better than the blind-chance level of 50%. While normally such a low percentage correct value would be taken as evidence for little more than random responding, it should be remembered that that this result was obtained with a set of test items in which specific similarity was set orthogonal to the grammatical status of the items. In fact, the large effect of specific similarity attests to the fact that the subjects were judicious when labelling the transfer items. If "far" grammatical items and "close" non-grammatical items are removed from the calculation, thereby producing a positive correlation between specific similarity and grammaticality, the percentage correct rises to almost 60% (psuedo Pc in Table 4.3). That is, what has typically been cited as evidence for abstract knowledge in these tasks, emerges to a reasonable extent in the present experiment only if the categorical status and the specific similarity of the items are positively confounded.

Subjects in the present experiment demonstrated a significant preference for transfer items that were produced by the substitution alteration (32.2% labelled "grammatical") over those produced by the addition alteration (25.6% labelled "grammatical"). This finding suggests that the subjects were influenced by the length of the items, in itself irrelevant to the grammar, responding more to items that were the same length as those they had learned (substitution items of 3, 5 or 7 letters in length) than to those that were not (addition items of 4, 6 or 8 letters in length). Alteration type was not independent of the specific similarity of the items. The effect of specific similarity was significantly larger for substitution items than for addition items. Thus, the tendency of the subjects to respond that an item was "grammatical" fell off with increasing difference between the training and transfer items as measured by the conjunction of specific similarity and item length. In addition, the subjects produced significantly more "grammatical" responses on their first pass through the substitution and addition items (29.6%) than they did on the second pass (28.3%). The miniscule size of the effect suggests that the reduction in "grammatical" responses is trivial. But note that this effect, like that of alteration type, also was not independent of the specific similarity of the transfer items. Accompanying the drop in "grammatical" responses over the two presentations of the transfer items was a significant drop in the size of the effect of the specific similarity of the items. Moreover, this interaction, in turn, was significantly related to the alteration type of the items. In general, as we move from those conditions in which overall "grammatical" responding is highest (substitution items on the first pass: 33.2%) to those conditions where it is lowest (addition items on the second pass: 25.3%), the effect of the specific similarity of the items undergoes a significant reduction.

The point is that there appears to be a positive correlation between the rate of "grammatical" responses and the size of the specific similarity effect. Reductions in the rate of responding as a function of pass or alteration type are associated with reductions in the magnitude of the specific similarity effect. Importantly, however, higher rates of "grammatical" responding appear to occur more as a function of increased "grammatical" responses to "close" items than a general increase in the tendency to respond. This also was the case for the small difference in "grammatical" responses to grammatical and non-grammatical transfer items, although the effect failed to reach significance.[4.5] Thus, there appears to be a non-linear, positively accelerated relationship between the tendency to respond that an item is grammatical and the perceived similarity of that item to a training stimulus. This is even the case if, as noted in the discussion of the difference between the two passes through the items, the item is the same nominal item encountered at a different time, or in a different context. Generally speaking, then, it appears that any variable that affects the perceived similarity between training and transfer stimuli should also affect the magnitude of the specific similarity effect. The effect of encoding condition, manifested as differences in breadth of transfer, is another variable that would be expected to modify the effect of specific similarity. This is discussed next.

Encoding and Breadth of Transfer

In Reber and Allen's (1978) experiment, the major difference on categorical transfer between the observation and paired-associate training conditions was manifested as a difference in sensitivity to the grammatical status of the items. However, given the likely confounding of grammaticality and specific similarity in their experiment, it is not at all clear that the difference that they observed was not actually due to a difference in sensitivity to specific similarity rather than grammaticality. In contrast, the results from the present experiment are unequivocal in this regard.

As is shown in the right-hand column (relative frequency) of Table 4.3, the three encoding conditions differed in the predicted direction in their overall tendency to respond that the items were grammatical, although this trend was not significant. However, the curvilinear relationship discussed above suggests that unless there are very large differences among the encoding conditions in terms of the perceived similarity of the transfer items to the training stimuli, or unless the variation occurs at the high end of the similarity dimension, simple response frequency is unlikely to be a powerful measure of encoding effects. Variation in the effect specific similarity, however, should be more sensitive since it provides a relative estimate of where on the similarity dimension the results of a particular encoding condition are located. Accordingly, as expected, the three encoding conditions differed significantly in their sensitivity to variation in the specific similarity of the transfer items, as is shown in column 6 (C–F) of Table 4.3. Thus, Observation training resulted in the greatest effect of the specific similarity of the transfer items while Overlapping Mnemonic training demonstrated the least. To say the least, the size of the encoding effect is unimpressive, but given the overall low rate of responding in this experiment (a mean of only 28.9% of the items were labelled "grammatical"), it is surprising that it occurred at all.

The important point, however, is that while encoding condition was related significantly to the magnitude of the specific similarity effect, it was not related significantly to variation in the magnitude of the effect of grammaticality. As is shown in column 5 (G–NG) of Table 4.3, the three encoding conditions did not differ significantly in their sensitivity to the grammatical status of the transfer items. Thus, if grammaticality is unconfounded with specific similarity, an effect that has been attributed to variation in sensitivity to the categorical status of the items may be seen to be due to variation in sensitivity to the specific similarity between training and transfer items. If the confounding between specific similarity and grammaticality is re-introduced, as was done for the calculation of Pseudo Pc in Table 8.3, the results of the transfer phase of the present experiment may be seen to replicate those of Reber and Allen (1978).

Recognition and Breadth of Transfer.

One of the key components to Reber and Allen's (1978) argument that the differences they observed between their two encoding conditions resulted from differences in implicit abstraction of the grammar is that observation training resulted in markedly poorer recognition of the training items than paired-associate training, while demonstrating superior categorical transfer. As discussed in Chapter 2, from the perspective of memory for individual instances as typically construed, such a result is paradoxical if memory for individual instances is the basis of responding during categorical transfer. However, precisely the same relationship between item recognition and categorical transfer was found in the present experiment although, as indicated above, encoding differences during categorical transfer were clearly not a function of differences in implicit abstraction.

The results of the recognition phase of the experiment may be found in Table 4.4 which presents the mean percentage of items labelled "old" as a function of encoding condition, and the grammatical status and the specific similarity of the items. The recognition data were analysed in two ways, as a recognition task and as a categorical transfer task.

Considered as a recognition task, the results are not surprising. Mnemonic training led to better recognition than the other conditions, and actively attempting to learn the items, as subjects were requested to do in the No Label condition, resulted in better recognition of the items than did simply observing them, as is shown in the last column (recognition Pc) of Table 4.4. That is, as in Reber and Allen's (1978) experiment, recognition accuracy was inversely related to categorical

Table 4.4
Recognition Phase

The mean percentage of items labelled "old" for each training condition as a function of the specific similarity and the grammatical status of the items. The first eight columns are the results of the recognition phase analysed as a categorical transfer task, and their interpretation is the same as that for categorical transfer. The remaining three columns refer to the mean percentage of recognition hits, the mean percentage of recognition false-positives, and the mean percentage of recognition items correctly labelled, respectively.

	Close		Far				Actual	Pseudo	Recognition		
	Gram	Non-G	Gram	Non-G	G-NG	C-F	Pc	Pc	Hits	Fa	Pc
Observation	18.2	12.5	18.0	7.8	7.9	2.5	54.0	55.2	61.7	14.1	73.8
No Label	20.3	7.0	3.1	2.3	7.0	10.9	53.5	59.0	62.1	8.2	77.0
Overlapping Mnemonics	12.2	1.6	4.7	2.3	5.5	3.4	53.3	54.9	63.3	5.2	79.0
	16.9	7.0	8.6	4.2	7.2	5.6	53.6	56.4	62.4	9.2	76.6

transfer performance. The source of these recognition differences, however, is interesting. The various encoding conditions did not differ significantly in hits, but they did differ significantly in false-positive responses, as is shown in columns 9 (recognition hits) and 10 (recognition false-positives) in Table 4.4. That is, recognition accuracy as a function of encoding condition was inversely related to the tendency to emit "old" responses and, more specifically, to the rate of false-positive responses. The significant variation in false-positive responses across encoding condition suggests that the various encoding operations result in differing degrees of breadth of transfer of the training items. Subjects in the Overlapping Mnemonics condition, for example, achieved superior recognition accuracy by restricting the breadth of transfer around the memory for the individual training items such that it encompassed fewer non-training items than was the case for subjects in the Observation condition. In addition to providing a manipulation check for the categorical transfer phase of the experiment, this result is consistent with the breadth of transfer differences found in the analysis of the training data and confirms the conclusion that the effect of the encoding manipulation is categorical transfer is to restict the breadth of transfer around what is known about specific, individual training items. It also suggests that, at least for the conditions of this experiment, the processes underlying categorical transfer and specific item recognition may not be all that different, as is detailed next.

In analyzing the data of the recognition phase of the present experiment (as well as those that follow) as a categorical transfer task, only false-positive responses were used so that in both tasks all that is being considered is the frequency of responding positively to the various kinds of "new" items. Unfortunately, the false-positive rate was so low in all of the experiments that many of the effects discussed for categorical transfer were not testable; for many of the cells of the design, the modal frequency and in some cases even the total frequency was zero. To circumvent this problem, the data were collapsed over the variable of alteration type, and, for the most part, only the main effects of the remaining variables are interpreted.

Since the actual task facing the subjects was a recognition task, it should come as little surprise that when they did produce errors, the errors were significantly more likely to be to items that were highly similar to the items to be recognized than to items that were dissimilar. That is, the frequency of false-positive responses was related to the specific similarity of the items; more responses were made to "close" items than to "far" items, as is shown in column 6 (C-F) of Table 4.4. Similarly, in line with the investigations discussed in Chapter 2 that used recognition as a test of categorical transfer, false-positive responding in the present experiment was significantly related to the categorical status of the items. The subjects were significantly more likely to consider grammatical items to be "old" than non-grammatical items, as is shown in column 5 (G-NG) of table 4.4. Somewhat unexpectedly, however, the ratio of the estimated variance components indicates that the relative effect size of grammaticality was almost twice as large as that of specific similarity, although as indicated by comparison of the partial correlations of the two variables with responding, the difference in magnitude was not significant ($z = 0.99$). The finding of an effect of grammaticality on recognition at least as large as that of specific similarity contrasts with the results of the categorical transfer phase in which the effect of grammaticality was much smaller than that of specific similarity. This result is explicable if it is assumed that the manipulation of grammaticality involved a covariate manipulation of item similarity that is not captured by the specific similarity manipulation, and that the demands of the recognition task emphasize this covariate similarity to a greater extent than the demands of the categorical transfer task. That is, there is more to the perceived similarity between training and transfer items than the number of common letters in position, which despite the intentions of the experimental design, remains correlated with the grammatical status of the items, and this additional form of similarity is more appropriate for a test of recognition than categorical transfer. This explanation is intuitively plausible and attests to the difficulty encountered in attempting to disambiguate specific item similarity from general principles, rules, prototypes and the like. It also underlines the potential task dependency of inter-item similarity; that what proves to be a good measure of similarity in one task, may not be so for another task involving the same items. Finally, it emphasizes the potential usefulness or heuristic value of decisions based upon specific item similarity across a wide domain of situations in the world, since only rarely will specific similarity and some general summary of he events of interest be completely uncorrelated.

The finding of significant effects of specific similarity and grammaticality in both categorical transfer and recognition confirms the notion that categorical transfer and recognition probably involve similar processes, at least for the conditions of this experiment. Moreover, the results of the recognition phase reaffirm the point about item specific transfer. As is shown in columns 7 (actual Pc) and 8 (Pseudo Pc) of Table 4.4, even in a recognition task in which generalization is supposed to be at a minimum (i.e., subjects are requested to respond on the basis of episodic rather than semantic information) one finds evidence of what looks like categorical transfer as long as specific similarity is positively correlated with the categorical distinction. Clearly, in a situation where the task is to generalize, such as the categorical transfer task, subjects can and probably do use transfer on the basis of specific similarity to even greater effect.

Chapter 5

EXPERIMENT 2

Experiment 2 was an attempt to replicate the results of Experiment 1 with greater control over the subjects' encoding of the training items. While the results of Experiment 1 are in agreement with the interpretation that the encoding manipulation resulted in differing degrees of breadth of transfer, it is clearly possible that differences other than degree of breadth among the encoding conditions were responsible for the group differences in categorical transfer and recognition. For example, the presentation of the training items for the Observation and No Label encoding conditions more closely approximated the appearance of the items during testing than was the case for the mnemonics condition. Hence, the lower sensitivity to the specific similarity of the transfer items and the lower false-positive rate of the subjects in the mnemonics condition may have occurred not because they had restricted breadth of transfer of the training items to a greater extent, but simply because the test items were less similar (lacking the associated phrases) to the training items as presented than was true for the other encoding conditions.

A related problem has to do with the Observation condition. Reber and Allen (1978) were interested in this training condition as an analogy to the non-directed, incidental manner in which people learn about natural grammars. Admittedly, much of our knowledge about the structure of our natural language is acquired incidentally, but the analogy between Observation training and natural language learning beyond this point seems a bit strained. Besides, even if the analogy were as close as Reber and Allen (1978) intended, the non-directiveness of the observation procedure renders the effect of this encoding condition open to many interpretations. That is, it is difficult to know to what to attribute the performance of the Observation subjects since we have very little idea as to what they were doing during training, or even whether the majority of the subjects were doing the same thing. Both this problem and that relating to the mnemonic training condition speak to same issue: If we wish to attribute differences in performance to varying degrees of breadth of transfer, then it would aid the task immensely if we could be reasonable confident that the encoding manipulation was directly associated with varying breadth and nothing else.

In their paper, Reber and Allen (1978) argued that the difference between their two encoding conditions was qualitative rather than quantitative. The observation training procedure was said to lead to a qualitatively different form of knowledge than the paired-associate procedure. In this claim, they were supported by the introspective reports of their subjects who claimed that the two training conditions and the transfer therefrom were experienced quite differently. Thus, if we wish to make a strong argument for a single quantitative dimension of variation underlying the differences due to encoding condition, it is again necessary that the encoding conditions be reasonably matched in all aspects including phenomenology, and differing only in the breadth of transfer produced. That is, we want tasks that are formally identical. On approach would be to use two observation training procedures and vary some parameter associated with breadth of transfer across them. Unfortunately, such non-directive training procedures do not readily admit to a parameter of this sort. A second approach, using two paired-associate procedures, is the one used here.

For the present experiment, two encoding procedures were used. In one, the Label condition, each training item was associated at random with a unique animal name. In the other, the Unique Mnemonic condition, each training item was associated with a unique phrase where the item was an acronym of the phrase. The associated phrases are unique in the sense that all the key words leading to the acronym training item were used only for that item. This contrasts with the mnemonic training procedure used in the previous experiment where the phrases were designed to partially overlap in both thematic content and words used.

During training, subjects in the two conditions were required to recall both the items and their associates. The required recall of the associates was to ensure that the subjects would use the associates when encoding and recalling the items. As in Experiment 1, following training, the subjects received identical transfer lists. Hence, the encoding conditions appear to be reasonably matched, in terms of both the training task and the relationship of the training task to the transfer task. Where the two encoding conditions differ, however, is in the specificity of the item associates. In the label condition, while it is true that each item has a unique associate, the fact that the items and their associates have been arbitrarily paired does relatively little to constrain the encoding of the training items. The particular encoding that a subject attempts for any given item and any given presentation of that item is fairly unconstrained by the animal name associate. In the Unique Mnemonic condition, in contrast, the mnemonic associate of each item was determined by the letters of the item. Hence, given that the mnemonic is used when the item is encoded, the encoding of any given item and any presentation of that item should be similarly determined or constrained and, therefore, should involve relatively little variance. Since the mnemonic associates are designed to be unique from one another, the mnemonic training procedure should result in relatively unique encodings for each item with relatively little variance from one encoding episode of a given item to the next.

For both conditions, the constraints imposed by the associates are removed during categorical transfer and recognition since no associates of the items accompany their presentation during these phases. Consequently, the encoding of the items during recognition and categorical transfer should be less constrained and more variable than that during training. Subscribing to the principle of encoding specificity (Tulving & Thomson, 1973), this increase in variability should decrease the chances of reinstating the conditions of encoding at training and thus decrease the chances of retrieval during transfer. However, since the mnemonic training would be expected effectively to result in fewer, different memorial representations of each item than that of the arbitrary label training, the perceived similarity of the transfer items to the training stimuli, and, hence, breadth of transfer, should be lower following Unique Mnemonic training than following Label training. To the extent that specific similarity is the underlying basis of the subjects' responses, performance on the categorical transfer and recognition tasks should reflect these differences.

Method

Materials

The materials were the same as those used in Experiment 1, with the exception of the modifications necessary to institute the new encoding conditions. The mnemonics used with Unique Mnemonic training are presented in Appendix A. These were constructed by producing a unique phrase for each of the 16 training items such that each item was an acronym of the key words of the phrase with which it was associated. None of the key words of each phrase occurred in any of the other phrases, and the thematic content of each phrase was unique to that phrase. For the Label condition, each of the 16 training items was associated with a unique animal name (e.g., MOOSE, BEAVER, RAT, FOX), such that the animal referent was also unique (e.g., since MOOSE was used, ELK was not).

Subjects

Thirty-two McMaster University undergraduates served as subjects in each of the two encoding conditions, resulting in a total of 64 subjects. Most subjects participated in the experiment to fulfill a course requirement in introductory psychology. The remainder, assigned at random across the two conditions, were paid $2.00 each to participate. As in Experiment 1, from 4 to 20 subjects were tested at a time in a classroom setting, and encoding conditions were deliberately mixed within a session.

Procedure

The procedure for both training conditions was the same as that used with mnemonic training in Experiment 1. During training, the subjects in both conditions were instructed to study carefully the four items and their associates on each page of their booklets, and then to recall both the items and their associates on the following blank page. Following training, the subjects received the categorical transfer and recognition tests as outlined in Experiment 1. All phases of the experiment were subject-paced.

Results and Discussion

Training

The data from the training phase were analyzed as in Experiment 1. The mean percentages of items correctly recalled per trial per list for each of the two encoding conditions are shown in Figure 5.1. Collapsed across training conditions, the results are a virtual replication of the training phase in Experiment 1. Significantly more items were corectly recalled on later trials (trial 5 = 78%) than on earlier trials (trial 1 = 46%), and this increment was significantly larger for the first sub-list the subjects received (trial 1 = 39%; trial 5 = 78%) than the second (trial 1 = 53%; trial 5 = 79%). Subjects correctly recalled significantly more items per trial from the second sub-list they received (69%) than from the first (61%), although the above effect of trials was the much larger effect. Thus, both

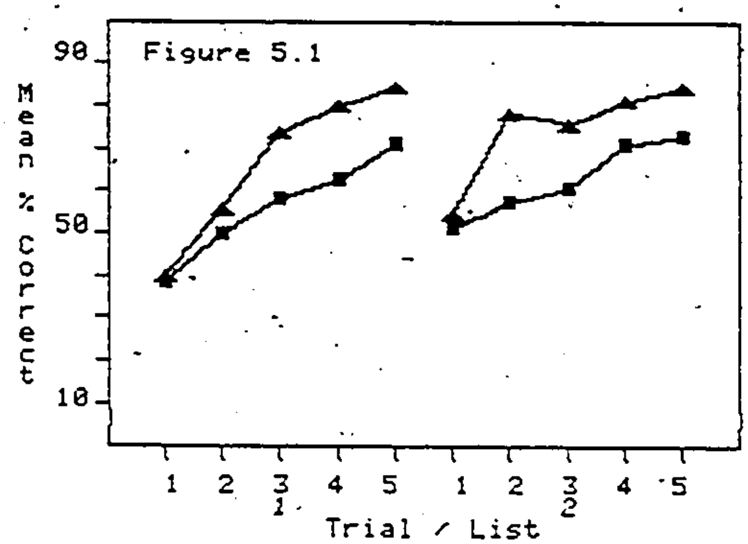

Figure 5.1: Training Phase. Mean percentage of items correctly recalled as a function of trials and sub-list for Label (squares) and Unique Mnemonic (triangles) training.

encoding conditions evidenced intra-list transfer and, to a much smaller extent, inter-list transfer.

Overall, subjects in the Unique Mnemonic training condition correctly recalled significantly more items per trial per list (71%) than did subjects in the Label condition (59%). As indicated by a significant training condition by trials interaction, and by the absence of any other significant effects involving training condition, the main effect of encoding was entirely a result of Unique Mnemonic training leading to increased intra-list transfer. Subjects in the Unique Mnemonic condition gained more per sub-list from successive trials than did Label trained subjects, and, consequently, obtained a higher overall level of performance. Thus, the greater specificity of the mnemonic phrases to their item associates significantly enhanced the subjects' ability to correctly recall the items. As discussed earlier, this effect probably arises from a combination of reduced breadth of transfer and a greater specificity of item encoding. The greater specificity of the associated mnemonics probably reduces the variability of encoding of the same item while increasing the difference in encoding between items, reducing confusions between items and enhancing contact with prior experiences of the same event. As before, however, the reduced breadth of transfer in this condition should lead to reduced categorical transfer, reduced recognition false-alarms, and, possibly, increased recognition hits. The qualification for recognition hits is motivated by the fact that during recognition the item associates are removed, increasing the chance that an "old" item will be encoded sufficiently differently that contact with with the learning experiences of that item will be reduced. While this should be true for both training conditions, the suspected greater encoding variability of each item during Label training should attenuate the effects of this increased recognition variability.

Categorical Transfer and Recognition

The results of the categorical transfer and recognition phases of the present experiment, analyzed as in the previous experiment, are a virtual replication of the major results of Experiment 1 with one notable exception. Table 5.1 presents the mean percentage of items labelled "grammatical" for categorical transfer as a function of encoding condition, and the grammatical status and specific similarity of the items. Table 5.2 presents the results of the recognition phase in terms of the mean percentage of items labelled "old" as a function of encoding condition, and the grammatical status and specific similarity of the items.

As in Experiment 1, the specific similarity of the items was by far the largest significant effect of the categorical transfer phase of the experiment, as is shown in column 6 (C–F) of Table 5.1. Importantly, however, there was no significant effect of the grammatical status of the items on categorical transfer; grammatical items were no more likely than non-grammatical items to be labelled "grammatical" by the subjects (column G–NG in Table 5.1). Yet, as in Experiment 1, there was a significant effect of grammatical status on recognition. Subjects produced significantly more false-positive responses to grammatical than non-grammatical items (column G–NG in Table 5.2), although, also as in Experiment 1, the magnitude of this effect on false-positive responses was not significantly different than that of the significant effect of specific similarity ($z = 0.50$). These results provide further support for the contention that the small, but significant effect of grammaticality in Experiment 1 was a consequence of a form of specific item similarity not captured by the specific similarity manipulation rather than the result of transfer on the basis of some general structure of the items. In fact, the only effect of grammaticality on categorical transfer in the present experiment occurred in an interaction with specific similarity and alteration type (where, as in Experiment 1,

Table 5.1
Categorical Transfer Phase

The mean percentage of items labelled "grammatical" for each training condition as a function of the specific similarity and the grammatical status of the items. Column G-NG refers to the mean percentage difference of positive responses between grammatical and non-grammatical items. Column C-F refers to the mean percentage difference of positive responses between close and far items. The column labelled Actual Pc refers to the mean percentage of transfer stimuli correctly labelled. Pseudo Pc refers to the mean percentage of close grammatical and far non-grammatical items correctly labelled. Relative Frequency refers to the mean percentage of items receiving positive responses.

	Close		Far		G-NG	C-F	Actual	Pseudo	Rel.
	Gram	Non-G	Gram	Non-G			Pc	Pc	Freq.
Label	46.6	45.8	29.1	26.7	1.6	18.3	50.8	60.0	37.0
Unique Mnemonics	35.6	39.6	26.8	27.1	-2.1	10.7	48.9	54.3	32.3
	41.1	42.7	27.9	26.9	-0.3	14.5	49.9	57.1	34.7

Table 5.2
Recognition Phase

The mean percentage of items labelled "old" for each training condition as a function of the specific similarity and the grammatical status of the items. The first eight columns are the results of the recognition phase analysed as a categorical transfer task, and their interpretation is the same as that for categorical transfer. The remaining three columns refer to the mean percentage of recognition hits, the mean percentage of recognition false-positives, and the mean percentage of recognition items correctly labelled, respectively.

	Close		Far				Actual	Pseudo	Recognition		
	Gram	Non-G	Gram	Non-G	G-NG	C-F	Pc	Pc	Hits	Fa	Pc
Label	17.4	10.9	7.0	3.9	4.8	8.7	52.4	56.8	62.5	9.8	76.3
Unique Mnemonics	6.5	2.3	3.1	0.8	3.3	2.5	51.6	52.9	73.0	3.2	84.9
	12.0	6.6	5.1	2.3	4.0	5.6	52.0	54.8	67.8	6.5	80.6

alteration type was significant as a main effect). Once again, there was a positive correlation between the rate of responding that an item of a given type was "grammatical" and the size of the specific similarity effect. Since the interpretation of an interaction of this sort in the previous experiment was in terms of a reflection of a curvilinear relationship between perceived similarity and responding, it appears reasonable to conclude that the present interaction is a manifestation of the same relationship with grammaticality simply functioning as another similarity measure. Again, as in Experiment 1, a comparison of the subjects' actual accuracy on the categorical transfer task (Actual Pc in Table 5.1: 49.9%) with their psuedo accuracy (Pseudo Pc in Table 5.1: 57.7%) suggests that unless the catgorical distinction (grammaticality) is confounded with specific similarity, the accuracy of the subjects' categorical transfer judgements appears essentially random.

Having eliminated grammaticality as a major determinant of subjects' responses during categorical transfer, the key question is whether or not it was still possible to demonstrate a version of Reber and Allen's (1978) encoding effect. As is shown in column 5 (G–NG) in Table 5.1, neither encoding condition showed any effect of the grammatical status of the items. Yet, as expected, they did differ significantly, in the predicted direction, in their sensitivity to the specific similarity of the items. They also differed in the predicted direction in the relative frequency of responding that the items were "grammatical" (column 9 in Table 5.1), although, as in Experiment 1, the trend was not significant. Expressed in terms of accuracy, the differences between the two encoding conditions replicate, in a fashion, the encoding effect in Reber and Allen's (1978) experiment, but provide a clear demonstration that this effect, as in Experiment 1, is a function of changes in sensitivity to the specific similarity of the items rather than variation in some implicit abstraction of the underlying grammar. As is shown in columns 7 (Actual Pc) and 8 (Pseudo Pc) of Table 5.1, unless specific similarity is positively confounded with grammaticality, encoding condition has no effect on the accuracy of the subjects' transfer judgements.

It is important to remember at this point that the two encoding conditions of the present experiment are both versions of only one of the encoding procedures used by Reber and Allen (1978). Yet, it was still possible to produce an effect of encoding condition on categorical transfer. Since the only treatment difference between the two encoding conditions was the specificity of the training item associates, it seems unlikely that the categorical transfer differences between the two encoding conditions were a function of qualitatively different processes underlying responding in the two groups. Rather, it seems more likely that the differences reflect nothing more than variation in only a single process, namely, the breadth of transfer of the memory for individual training items. The significant difference, discussed earlier, in intra-list transfer during training between the two conditions supports this interpretation. The results of the recognition phase of the experiment add further support. As is shown in columns 9, 10, and 11 in Table 5.2, subjects given Unique Mnemonic training correctly identified more of the recognition items than did subjects given Label training, but as in Experiment 1, the increased recognition was because the mnemonic trained subjects produced significantly fewer false-positive responses rather than because of significant variation in hits, as would be expected if they were making their "old"-"new" judgements with less breadth of transfer around their memory for individual training items than was the case for Label trained subjects.

In summary, the results of Experiment 2 reinforce the conclusions advanced for Experiment 1. Labelling a transfer item as "grammatical" is more a function of the specific similarity of the item than the grammatical status of the item. In fact, in the present experiment, categorical transfer performance was independent of the grammaticality of the items. Differences in categorical transfer performance as a function of encoding appear to be related to differences in breadth of transfer rather than variation in implicit abstraction. Finally, since grammaticality was related to false-positive responses during recognition but not to categorical transfer judgements, it would appear that grammaticality is a measure of specific item similarity that is not captured by the specific similarity variable used in these experiments. To argue differently would be to suggest that the subjects in the present experiment were sufficiently obtuse that they used non-episodic (i.e., grammatical) information to fulfill the requirements of an episodic task (i.e., recognition), while using only episodic information to meet the requirements of what is nominally a semantic task (i.e., categorical transfer). This seems highly unlikely. As is demonstrated in Experiment 4 (see Chapter 7), when subjects are explicitly requested to judge the similarity between training and transfer stimuli, grammatical transfer stimuli are judged to be significantly more similar to training items than are non-grammatical stimuli, even when the items are "matched" for specific similarity as defined in the present experiments.

Chapter 6

EXPERIMENT 3

In the previous two experiments, the effect of the mnemonic encoding conditions was interpreted to reflect a restriction in the breadth of transfer of the memory for individual training items. The source of this effect was attributed to the specificity of the mnemonic phrases to their associated training items. In the next experiment, the specificity of the mnemonic associates during mnemonic training was eliminated in an attempt to investigate the effects of mnemonic training in these tasks if the mnemonics associated with the items emphasized the commonalities rather than the dis-similarities among the training items.

As Reber and Allen (1978) have argued, paired-associate training conditions such as those used by Brooks (1978) and the one used in their own experiment, may lead subjects to employ analogy to individual instances during transfer because the requirement to individuate the items during training prevents the subjects from implicitly abstracting the commonalities among the items. Similarly, the requirements of the training task for subjects given label and mnemonic training in the previous experiments may have led them to emphasize specific item information at the expense of inter-item (grammatical) information. This possibility was investigated in the present experiment through the use of a mnemonic encoding condition that I refer to as Grammatical Mnemonic training. This condition was an extreme version of the Overlapping Mnemonic training used in Experiment 1. For Grammatical Mnemonic training, the mnemonic phrases associated with each acronym item were constructed such that the words used reflected the path taken through the grammar to produce the item. Thus, items that were produced by similar paths through the grammar would also have similar mnemonic phrases associated with them. For example, the mnemonic phrase associated with the training item VXTVX was "Virgins eXamining Terrified Varmints eXplode". The mnemonic associated with VXTTTVT, which was produced via a similar path through the grammar (except for repeating the T at node 2, and exiting the grammar differently), was associated with the phrase "Virgins eXamining Terrified, Timid, Tense Varmints Tremble".

To provide a contrast to Grammatical Mnemonic training, a second group of subjects were given observation training similar to that used in Experiment 1. In this case, however, the viewing time per page of items was reduced to 15 s (hence, Short Observation) from the 90 s used in Experiment 1. According to Reber and Allen's (1978) notions about observation learning, this reduction in viewing time should decrease the degree of item specific information that subjects gain from their training experiences, while possibly increasing the apprehension of commonalities among the items. Thus, by markedly different routes, both the Grammatical Mnemonics and the Short Observation conditions in the present experiment should decrease the subjects' learning of specific items while, according to Reber and Allen's (1978) conception, increasing implicit abstraction of the underlying grammar.

In terms of breadth of transfer, the conception of the effects of these two encoding conditions is quite different from that of Reber and Allen (1978). For the Short Observation condition, reducing the amount of time the subjects have to study the items on any given exposure would be expected to increase the variability of encoding of the items and, hence, increase transfer since 15 s for a four-item page of items should leave enough time for little more than a cursory glance at each item on the page. Hence, the encoding of an item on any given presentation should be more dependent on such item-irrelevant factors as the position of the item on the page or what item occurred before it than would be true if time was available to encode more item specific or item-relevant information.

The Grammatical Mnemonics condition requires a modification of the arguments presented in Experiment 2. The unique mnemonics in Experiment 2 were constructed so that each item associate would have a stable, but unique encoding. That is, the condition was designed such that there would be maximum overlap in the encoding of each presentation of the same item, but minimal overlap in the encoding of different items. Each item, through being encoded with its unique mnemonic, could be thought of as giving rise to a unique domain of similarity. That is, the artificial world as presented to each subject during training consisted of eight overly-determined, highly distinct categories, corresponding to the eight different types of encoding episodes of a training list. In a sense, much of the work confronting the subjects during training had already been done for them. There was no need, for example, for the subjects to look for differentiating information or cues that are diagnostic of one item as opposed to another since that information has been supplied by the unique mnemonic associated with each item.

For Grammatical Mnemonic training in the present experiment, the subjects would have to search actively for differentiating information in their attempt to accomplish the training task. The here is that to reduce the confusion between items introduced by the similarity of the grammatical mnemonics, the subjects would have to develop their own unique encoding patterns that would capitalize on the differences between items. In a sense, the difference between the Unique Mnemonics condition and the Grammatical Mnemonics condition may be seen as a distinction between a priori and a posteriori individuation of the training items. For the Unique Mnemonic training, the differences in the formal characteristics of the items have been enhanced by the differences in the associated phrases. Each item has been unambiguously labelled, as has every component of each item. The letter V that begins the item VXM, for example, is not the same as the letter V that begins the item VXTVX; one is indicative of "Viruses" and the other "Vertical". In contrast, a letter V in the first position of an item for the Grammatical Mnemonics condition is always indicative of "Virgins". Therefore, to distinguish one item from another, the subjects in the Grammatical Mnemonics condition are forced to encode the initial V in VXM, for example, differently than the initial V in VXTVX, despite the extended similarity of the components brought on by the associated grammatical mnemonics. To accomplish the learning task, the subjects must actively individuate the items from one another. To the extent that they can accomplish this individuation during training, the breadth of transfer that they display during categorical transfer should be reduced. On the other hand, failing to individuate the items from one another during training because of the extended similarity induced by the grammatical mnemonics would be expected to leave the subjects with little basis for discriminating among the items, resulting in little difference between their performance and that expected from Short Observation subjects. As will be seen, the results of the experiment support neither of these possibilities, but neither do they support in any obvious way the interpretation suggested by Reber and Allen's (1978) implicit abstraction hypothesis.

Method

Materials

The materials were the same as those used in the first two experiments with the exception of the modifications required to institute the Grammatical Mnemonics encoding condition. The mnemonics used with each of the two training lists are given in Appendix A. These were constructed by assigning a unique word to each legal node to node transition in the artificial grammar. The key letter (typically the first letter) of each word corresponded to the letter assigned to that transition in the grammar. For example, the node 1 to node 2 transition was assigned the word "Many", while the node 1 to node 3 transition transition was assigned the word "Virgins". For those transitions that allowed repeated cycling (node 2 to node 2 and node 5 to node 5), each successive repetition was assigned a unique word. Thus, for example, the first node 2 to node 2 transition was assigned the word "Terrified", the second the word "Timid", the third the word "Tense", and so on. The only constraints on the words used were that each word had to be unique, and that the resultant phrase for each training item had to be meaningful and correct according to English syntax. Where necessary, both the syntax and the meaning of the phrases were clarified by the appropriate use of commas.

Subjects

Thirty-two McMaster University undergraduates served as subjects in each of the two training conditions, for a total of 64 subjects. Most subjects participated in the experiment to fulfill a course requirement in introductory psychology. The remainder, assigned at random across the two conditions were paid $3.00 each to participate. The two conditions were run separately in order to control the time that subjects in the Short Observation condition had to study the items. For both conditions, from 2 to 10 subjects were tested at a time in a classroom setting.

Procedure

The training procedure for the Short Observation condition was the same as that used for the Observation condition in Experiment 1, with the exception that the time per page of training items was reduced from 90 s to 15 s. The subjects were allowed to study the four items on each page for 15 s and then were required to turn to the next page of items on a signal from the experimenter. For the Grammatical Mnemonics condition, the training procedure was the same as that for the mnemonics conditions in the first two experiments. Subjects were instructed to study the items and their associated phrases on each page and then to recall the items and their associated phrases on a following blank page. Following training, all subjects received the categorical transfer and recognition tests as outlined in Experiment 1.

Results and Discussion

Training

The first notable result of the experiment was that attempting to learn the items in the Grammatical Mnemonics condition proved to be quite difficult. Depicted in Figure 6.1 is the mean percentage of items correctly recalled per trial per sub-list for the 32 subjects in the mnemonics training condition. As can be seen, mean overall performance (49% correct) was quite poor. This result contrasts sharply with the performance of subjects in the mnemonics conditions in the first two experiments in which the mean percentage of items correctly recalled averaged 67% and 71%, respectively. Although across experiment comparisons should be viewed with some caution, since the training items are identical for all three experiments, it does seem unlikely that the poor performance of subjects in the present experiment is entirely due to subject differences. Be that as it may, the poor training performance is associated with a number of findings that are unique to this experiment.

Unlike the previous experiments, the mean number of items correctly recalled on the second sub-list (50%) was not significantly different than on the first (48%). Moreover, there was no sub-list by trials interaction. Statistically, performance on the two lists was virtually identical. Clearly, in contrast to the first two experiments, the subjects in the present experiment showed no evidence of inter-list transfer. While this result is similar to the results of the paired-associate condition in Reber and Allen's (1978) experiment, I believe it occurred for a different reason.

As usual, the mean number of items correctly recalled increased significantly

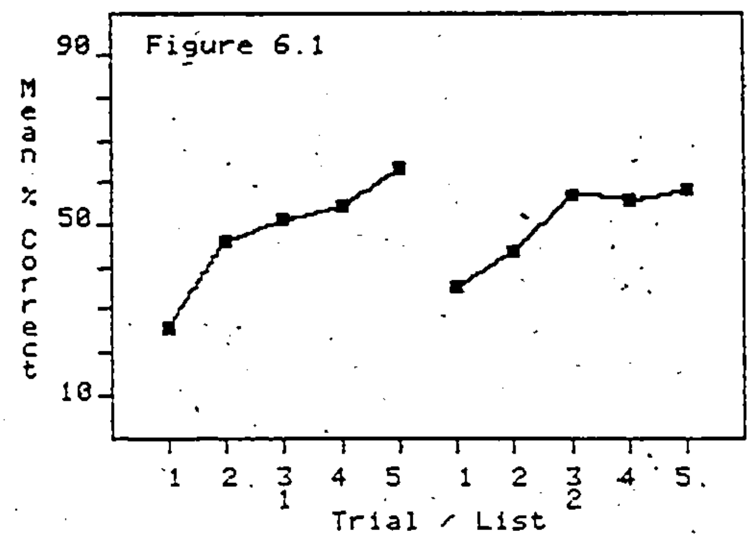

Figure 6.1: Training Phase. Mean percentage of items correctly recalled as a function of trials and sub-list for Grammatical Mnemonic training.

as a function of training trials. More items were correctly recalled on the later trials with a given sub-list (trial 5 = 61%) than on the earlier trials (trial 1 = 30%). But even this effect, relative to that of the previous experiments, is reduced. Subjects in the present experiment appear to have gained less from the repetitions of the items than subjects in the previous experiments, despite the fact that from their low trial 1 performance, they had more to gain. Thus, the poor overall performance of subjects trained with Grammatical Mnemonics appears to be a consequence of both no inter-list transfer and reduced intra-list transfer.

There is a problem, however. The dependent variable for the training analyses is an accuracy measure and not, as in recognition and categorical transfer, a simple frequency measure. Accuracy can be reduced because of either under- or over-generalization. Thus, both the lack of inter-list transfer and the reduction in intra-list transfer may have occurred as a consequence of <u>too much</u> transfer from the encoding of one item to the encoding of another, rather than too little. At an extreme, this would be equivalent to the subjects treating nominally different items as the same event. Thus, the bundle of instances in memory that, to the subject, are his or her representations of experiences with a given item, may actually be composed of experiences arising from nominally different items.

There remains the question as to what, exactly, Grammatical Mnemonic training has done. Relative to the memory training conditions in the earlier experiments, it appears that Grammatical Mnemonic training reduces the subjects' ability to discriminate the letter strings from one another. The subjects apparently find the items to be more alike one other than do subjects in the earlier memory training conditions, including the No Label condition in Experiment 1. Thus, while Overlapping Mnemonic training, Unique Mnemonic training and Label training relative to No Label training appear to render the items less alike one another, Grammatical

Mnemonic training appears to increase the similarity between items to an extent even greater than that when the items are encountered with no labels or mnemonics at all (i.e., No Label training in Experiment 1). Since the items are the same for all training conditions, it is clear that these differences between encoding conditions are not a function of differences in the formal properties of the training items. Thus, the differences must occur as a function of changes in the informal properties, or encoding, of the items. But how is it possible to make items that are clearly perceptually disimilar, as indexed by the No Label condition, so similar that they are almost impossible to discriminate from one another? The answer to this question, I believe, is to be found in how the subjects interpret or represent the items to themselves.

Consider, as an example, a collection of perceptually discriminable events such as dogs. By perceptually discriminable I mean only that most observers if confronted with the animals will acknowledge that while they may all have, say, two ears, four legs and a tail, the types of ears, legs and tails vary considerably from one dog to the next. Hence, at that level, the dogs are quite discriminable. However, if through some encoding task, say an itemized check list, each observer was asked to score each animal on whether on not it had ears, a tail, four legs, etc., the resulting representation of each dog (on the check list) would be almost indistinguishable from that of any other dog. If the observers were asked to discriminate the dogs strictly on the basis of the check list information, they would have great difficulty doing so.

I suspect that something quite similar to this example is happening in the Grammatical Mnemonics training condition. Unlike the other memory training conditions in which the item representations are probably at the level of specific letters in position (particularly for Unique Mnemonics), the representations of the items in the Grammatical Mnemonics situation are probably at the level of higher-order features, groups of letters or patterns. As in the dog example, representing the items in terms of higher-order features would have the effect of increasing the similarity of one item to another, despite the fact that, at the lower or letter level, the items are discriminable.[6.1]

One way of viewing the present conception is to think of the mnemonic associates of the items as instructions on how to encode each item. If the instructions for each item are quite different, as in the Unique Mnemonics situation, the representation of each item should also be unique. If, on the other hand, the instructions are quite similar, as in the Grammatical Mnemonics situation, to the point of requesting non-discriminating information about each item, the resulting representations of the items should be quite similar to one another and, hence, quite confusable with one another.

By itself, the finding that the subjects in the Grammatical Mnemonics condition failed to discriminate the items from one another during training would suggest an extreme breadth of transfer and, hence, little basis for discriminating among the items during transfer. In this regard, they would be expected to perform in much the same way as subjects given Short Observation training. That is, subjects in both conditions would be expected to perceive most transfer items as similar to the training stimuli and, hence, fail to sort the transfer items successfully, as well as do relatively poorly on a test of item recognition. However, the suspected source of the failure to discriminate among the items during training for subjects in the Grammatical Mnemonics condition – the encoding of the items in terms of higher-order features rather than letters in position – suggests that they will be able to discriminate among the transfer items, particularly with respect to grammatical status. To see why this may be the case, it is necessary to consider the similarity between transfer and training stimuli in terms of the higher-order features discussed above rather than in terms of specific letters in position as is done for the specific similarity manipulation.

Since the encoding of the items in terms of higher-order features is suggested to be a consequence of the grammatical mnemonics, the reader is invited to attempt the following demonstration. Study the items and associated grammatical mnemonics (in Appendix A) for one of the two training lists (say, List 1) for a short period of time, and then attempt to generate grammatical mnemonics for the transfer items shown in Table 4.1. If the reader's experience is anything like mine, he or she will have found that it is relatively easy to generate mnemonics for grammatical transfer items, and quite difficult to do so for non-grammatical items. Moreover, the specific similarity of the transfer items to the studied training items seems quite irrelevant to this ability. It is no more difficult to generate mnemonics for the "far" grammatical transfer items (those in the bottom half of Table 4.1 if training list 1 was studied) than it is for the "close" grammatical transfer items. For example, the "close" grammatical transer items VXR and MVRXM are readily represented as "Virgins eXpect Rewards" and "Many Varmints and Rats eXpect Miracles", respectively. Similarly, the "far" grammatical transfer items, MVT and MTTVX, are easily seen as acronyms for "Many Varmints Tremble" and "Many Terrified, Timid Varmints eXplode". In contrast, the non-grammatical items fail to give rise to any reasonable grammatical mnemonics. For the non-grammatical transfer items VXT and MVRXX, for example, the best that came to mind for me were "Virgins eXpect (to) Tremble" and "Many Varmints and Rats eXpect (to) eXplode"; neither of which, in contrast to those readily generated for the grammatical transfer items, seems highly similar in either meaning or formal structure to the mnemonics for the training items in the first (or second) list. Of course, this result is a function of the formal procedure used to produce the grammatical mnemonics in the first place. But, to the extent that the subjects represent the items to themselves in this way, their judgements of the similarity between training and transfer stimuli and, consequently, their transfer judgements should show similar effects.

If the above analysis of the effects of grammatical mnemonic training is valid, then subjects given grammatical mnemonic training should demonstrate an enhanced ability to discriminate the categorical status of the items during transfer. It is important to recognize, however, that this ability would not represent implicit abstraction of the underlying grammar. Rather, it would represent transfer on the basis of similarity to the memory of the individual training stimuli, where this similarity is different than that captured by the specific similarity manipulation. That is, such a result would not mean, necessarily, that the subjects had abstracted rules, just analogy to individual instances at a higher level of similarity than individual letters in position.

Categorical Transfer and Recognition

The data from the categorical transfer and recognition phases of the experiment were analysed as in the previous experiments. The major results of both phases, summarized as in the previous experiments, are shown in Tables 6.1 and 6.2.

With respect to the primary contention that specific similarity is a more important determinant of responding than is grammaticality in these tasks, the results of this experiment are in agreement with the results of the previous two experiments. Specific similarity was the largest significant effect in the categorical transfer phase of this experiment. As in Experiment 1, there was a small, but significant effect of grammaticality, but the magnitude of this effect was significantly smaller ($z = 3.49$) than that of specific similarity. In fact, the ratio of the estimated variance components indicates that the relative effect size of

Table 6.1
Categorical Transfer Phase

The mean percentage of items labelled "grammatical" for each training condition as a function of the specific similarity and the grammatical status of the items. Column G-NG refers to the mean percentage difference of positive responses between grammatical and non-grammatical items. Column C-F refers to the mean percentage difference of positive responses between close and far items. The column labelled Actual Pc refers to the mean percentage of transfer stimuli correctly labelled. Pseudo Pc refers to the mean percentage of close grammatical and far non-grammatical items correctly labelled. Relative Frequency refers to the mean percentage of items receiving positive responses.

	Close		Far		G-NG	C-F	Actual	Pseudo	Rel.
	Gram	Non-G	Gram	Non-G			Pc	Pc	Freq.
Short Observation	35.0	39.5	24.1	26.0	-3.2	12.2	48.4	54.5	31.1
Grammatical Mnemonics	37.1	23.8	25.1	17.3	10.6	9.3	55.3	59.9	25.8
	36.1	31.6	24.6	21.7	3.7	10.7	51.9	57.2	28.5

Table 6.2
Recognition Phase

The mean percentage of items labelled "old" for each training condition as a function of the specific similarity and the grammatical status of the items. The first eight columns are the results of the recognition phase analysed as a categorical transfer task, and their interpretation is the same as that for categorical transfer. The remaining three columns refer to the mean percentage of recognition hits, the mean percentage of recognition false-positives, and the mean percentage of recognition items correctly labelled, respectively.

	Close		Far				Actual	Pseudo	Recognition		
	Gram	Non-G	Gram	Non-G	G-NG	C-F	Pc	Pc	Hits	Fa	Pc
Short Observation	25.3	20.3	21.1	21.1	2.5	1.7	51.2	52.1	44.5	21.9	61.3
Grammatical Mnemonics	26.3	13.2	12.5	8.6	8.5	9.2	54.2	58.9	47.7	15.2	66.2
	25.8	16.8	16.8	14.8	5.5	5.5	52.7	55.5	46.1	18.6	63.8

specific similarity was over 10 times that of grammaticality. Both variables were significantly related to false-positive responding during recognition to statistically identical degrees ($z = 0.09$). As a whole, the 64 subjects in the experiment produced more false-positive responses to grammatical items than to non-grammatical items, and more to "close" than to "far" items.

As would be expected on a priori grounds for the Short Observation condition and on the basis of the training performance of the Grammatical Mnemonics condition, neither encoding condition was associated with good recognition performance. In fact, as is shown in column 11 (recognition Pc) of Table 6.2, the recognition performance of subjects in both conditions was less accurate than that of subjects in any of the conditions of the previous experiments. Importantly, however, despite their relatively low recognition accuracy, subjects in the Grammatical mnemonics condition were still more accurate than subjects in the Short Observation condition. Again, however, the difference was due to a significant reduction in false-positive responses (column 10 in Table 6.2) for the mnemonics condition rather than to significant variation in hits (column 9 in Table 8.2). As in the previous experiments, this result suggests that the encoding manipulation was effective in producing differences in breadth of transfer across the two conditions. Consistent with this conclusion, Short Observation subjects were more likely than Grammatical Mnemonics subjects to label the categorical transfer items as "grammatical" (column 9 in Table 6.1), although, as in the previous experiments, the trend was not significant. Similarly, as is shown in column 6 (C-F) in Table 6.1, the two encoding conditions differed in the appropriate direction in their sensitivity to the specific similarity of the items. Unlike in the previous experiments, this difference was not significant. However, encoding condition was significantly related to the interaction of alteration type and specific similarity, suggesting that the encoding conditions did differ in their perception of the specific similarity of the items, although the differences were quite small. Overall, while it appears reasonable to conclude that the two encoding conditions differed in breadth of transfer, it also appears that the effect of the Grammatical Mnemonics training procedure is to reduce these differences relative to those found in the previous experiments.

The results of the preceding analysis suggest that, for the most part, the two conditions were not all that different despite their markedly different training. However, this was not the case. The major difference between the two encoding conditions occurred in their sensitivity to the grammatical status of the items. In fact, as is shown in column 5 (G–NG) in table 6.1, the significant main effect of grammaticality during categorical transfer was entirely due to the performance of subjects in the Grammatical Mnemonics condition. The consequences of this result are two–fold. First, as is shown in column 7 (Actual Pc) of Table 6.1, the subjects in the mnemonics condition demonstrated an above–chance ability to determine correctly the grammatical status of the items, something that definitely was not true of the subjects in the observation condition. Second, while both conditions demonstrated the typical increase in accuracy associated with Pseudo Pc (column 8 in Table 6.1), again highlighting the importance of specific similarity, it is the condition with the lower breadth of transfer that produced the higher Pc. Not only is this last result the opposite of what occurred in the first two experiments, it can also be interpreted as inconsistent with the results of Reber and Allen's (1978) experiment since it is the memory training rather than the observation training in the present experiment that produces the superior transfer.

The results of the transfer phase of this experiment provide a clear demonstration that observation training in this situation does not lead to implicit abstraction of the underlying grammar. In the absence of a confounding between grammatical status and specific similarity, there is no evidence in the present experiment that subjects given observation training have abstracted anything about the artificial grammar used to generate the items. It is important to note that the conditions of the Short Observation training procedure in the present experiment more closely approximated those of the observation procedure in Reber and Allen's (1978) experiment than did those of the observation training in Experiment 1. In particular, the total viewing time per item in the present experiment was reduced from that of Experiment 1 to more closely match the 30 s (over three presentations) per item used by Reber and Allen (1978). Yet, in doing so, the effect of grammaticality was eliminated rather than enhanced. Thus, it seems highly unlikely that the successful transfer performance following observation training in Reber and Allen's (1978) experiment had anything to do with abstraction of the underlying grammar, implicit or otherwise.

The results from the Grammatical Mnemonics condition present a different story altogether. The fact that subjects in the Grammatical Mnemonics condition did not differ significantly from their observation counterparts in their sensitivity to specific similarity suggests that this variable was important for subjects in both conditions. However, the large effect of grammatical status for subjects given Grammatical Mnemonic training could be interpreted to suggest that they had managed to induce a partial representation of the grammar that they used in conjunction with specific item information during categorical transfer. Such a conclusion would be consistent with the conclusions suggested by Reber and his co-workers for the results of another experiment. In this experiment, Reber <u>et al.</u> (1980) compared the transfer performance emanating from a number of different implicit and explicit learning conditions. As discussed earlier, instructions to the subjects to search explicitly for rules while learning the training items typically results in the subjects doing less well on a test of transfer than subjects simply asked to memorize the training items. Reber et al. (1980) replicated this basic result, but also demonstrated that if the training items were presented in a manner that emphasized the grammatical commonalities among the training items, subjects instructed to look explicitly for rules while learning the items produced the best transfer performance of any training condition. While Reber et al. (1980) produced this result by presenting the items grouped according to the paths taken through the grammar to produce them and by requesting that their subjects explicitly look for rules during training, the parallels with the Grammatical Mnemonics condition of the present experiment are clear. In both cases, emphasizing the grammatical commonalities among the training items resulted in superior transfer accuracy. The question, of course, is whether this improved performance arose as a function of increased abstraction of the underlying grammar.

As in all of Reber's artificial grammar experiments, there is good reason to believe that grammaticality and inter-item similarity were confounded in the Reber et al. (1980) experiment. Thus, the high transfer accuracy of the subjects in the Reber et al. (1980) experiment can not be taken as unequivocal evidence for abstraction of the underlying grammar. Similarly, the large effect of grammaticality for subjects given Grammatical Mnemonic training in the present experiment is also not unambiguously interpretable, despite the fact that the specific similarity of the transfer items was set orthogonal to their grammatical status. The reason for this was presented earlier when the results of the training phase were discussed. Specifically, if the Grammatical Mnemonics subjects are encoding the items in terms of higher-order features rather than specific letters in position, the similarity between training and transfer items, assessed at this level, is no longer orthogonal to the grammatical status of the transfer items. As discussed in Chapters 2 and 3, this highlights the importance of considering how the subjects represent the items to themselves and how this may affect the assessment of similarity between items. The orthogonal manipulation of specific similarity and grammaticality only controls the confounding between categorical status and inter-item similarity to the extent that the similarity metric underlying the measurement of specific similarity is correlated with inter-item similarity as perceived by the subjects. In the present experiments, the degree of control exercised by this variable has been quite good, particularly for Experiment 2 and for the Short Observation procedure in the present experiment. However, in all of these experiments, grammaticality has been found to exert some effect, at least for recognition if not categorical transfer. To this point, I have represented this result as occurring as a function of a form of residual similarity not captured by the specific similarity manipulation that, in light of the results of Experiment 2, is emphasized more during recognition than during categorical transfer. The results of the training and transfer phases of the present experiment are consistent with a similar interpretation and further suggest that the residual similarity is a function of representing the items at a higher featural level than specific letters in position. Grammatical Mnemonic training and possibly the structured training procedure used by Reber et al. (1980) may emphasize this form of similarity to a greater extent than the other encoding conditions that have been used. The possibility that the effect of grammaticality in these experiments may be a consequence of a confounding with a residual form of inter-item similarity was investigated in the next experiment.

Chapter 7

EXPERIMENT 4

In the preceding three experiments, subjects were trained with only eight items from the grammar. In the experiments of Reber and his colleagues subjects typically have received 25 training items. Consequently, while it appears that implicit abstraction is an unlikely explanation for the results of the present experiments, it is possible that had the subjects been trained with more items, evidence for implicit abstraction might have been obtained, as has been claimed by Homa et al. (1981). In addition, the rate of positive responses in the present experiments has been low, particularly relative to that of subjects in, for example, Reber and Allen (1978). As mentioned earlier, some investigators have found that increasing the number of training exemplars for a given category increases both the accuracy of subsequent generalization of that category (Homa, Cross, Cornell, Goldman and Shwartz, 1973; Homa and Chambliss, 1975; Homa et al., 1981; Omohundro, 1981) and, a fact acknowledged by the investigators to somewhat of a puzzle, the rate of false-positive responses of that category in categorical transfer (Homa et al., 1973; Homa et al., 1981) and recognition (Omohundro, 1981). Consequently, increasing the number of training exemplars within the present experiments may facilitate the rate of responding in both categorical transfer and recognition, as well as increase the accuracy of the subjects' categorical transfer decisions. For these reasons, the next three experiments trained each subject with 16 items, twice the number used in the first three experiments.

Another factor that may have reduced the response rate in the preceding experiments is the response required from the subjects. For both categorical transfer and recognition, the subjects were to indicate their judgement of the items with a unitary or "yes" response. Only those items that the subjects believed obeyed the rules (or believed were "old" for recognition) were to be marked. All other items were to be left untouched. This procedure would be expected to bias against the production of positive responses since it requires that the subject explicitly react to an item only when he or she feels that the item "obeys the rules" (or is "old" for recognition). Thus, for most subjects a positive response probably has a singular meaning. This is probably not the case for items left blank. These latter decisions during categorical transfer, for example, may arise because the subject is confident that a particular item does not obey the rules or because the subject is unsure of the rule status of the item. That is, while responding positively to an item probably indicates that the subject thought the item to be grammatical, leaving an item blank does not mean necessarily that the subject believed it to be non-grammatical. However, with the present response measure, we have no way of knowing for sure how the subjects distribute responses that arise from a "don't know" or "unsure" state, or even whether the majority of subjects distribute them in the same way. Consequently, in the next experiments to be reported, the subjects were asked to indicate their judgement of each item on a six-point response scale that allowed the subjects to indicate their degree of belief that an item is or is not grammatical.

In the preceding three experiments, the recognition task was subsequent to the transfer task. This resulted in the possibility that the specific similarity and grammaticality effects observed during the recognition phase were a consequence of the subjects' experience during the categorical transfer phase of the experiment rather than a consequence of similar processes underlying both tasks. That is, the subjects may have been more likely to judge as "old" those items that were "close" or grammatical because they had judged these same items to be "grammatical" during categorical transfer. To ensure that the subjects' recognition judgements would be independent of any effects arising directly from categorical transfer, the next three experiments reversed the order of these two tasks such that recognition judgements preceded categorical transfer.

Reversing the order of the recognition and categorical transfer tasks raises the possibility of investigating whether prior experience with some of the transfer items during recognition affects how they are judged during categorical transfer. The subjects' prior experience with these items might increase the familiarity of these items relative to those not experienced during recognition, possibly resulting in a higher rate of "grammatical" responses to them. This possibility was investigated in each of the next three experiments.

Another issue investigated in the next experiments to be reported is that raised by Reber and Allen's (1978; see also Reber et al., 1980) finding that their subjects more readily detected (i.e., correctly labeled as non-grammatical) those non-grammatical items that had violations in the initial or terminal letter positions of the transfer items than those items where the grammatical violations occurred in deep, internal letter positions. While I have no quibble with their finding, I do take issue with Reber and Allen's representation of this result as detection of violations of the underlying grammar. Since no grammatical item, and, hence, no training item, may have these violations, each "violation" in the transfer stimuli is a guaranteed difference between any given transfer item and any or all training items. It may be these differences or specific dissimilarities between training and non-grammatical transfer stimuli, and not the fact that they are grammatical violations, that are responsible for the result. If this is the case, then grammatical transfer stimuli, where such changes at the different letter positions would not, by definition, be violations of the grammar, should evidence the same pattern as that obtained with the non-grammatical transfer items. This possibility is investigated in the next experiments by orthogonally varying the position of change from training to transfer stimuli across the grammatical status of the transfer items.

Two further changes were made in the design of this and subsequent experiments. First, a new artificial grammar was constructed to produce the items. The grammar used in the preceding experiments was too limited in the number and types of items produced from it to accomplish the objectives of the next three experiments. Second, an eight-item, interview questionnaire was constructed to assess subjects' judgements about how they fulfilled the requirements of the categorical transfer task. This questionnaire was not intended to be an exhaustive list of potential strategies; rather it was included simply to determine whether subjects thought that they were using rules or similarity to training stimuli when making their categorical judgements.

In the present experiment, the focus was shifted from the training task received by the subjects, to the requirements of the categorical transfer task. Specifically, the present experiment directly tackled the notion that subjects' judgements during categorical transfer are mediated by specific similarity. If specific similarity is the basis of subjects' judgements, then requesting subjects to sort the items on the basis of their similarity to training stimuli should produce patterns of responding to the various types of transfer stimuli that are similar to those obtained from subjects given the usual categorical transfer task. In particular, if, as suggested previously, the occasional significant effect of grammaticality represents a form of residual similarity not captured by the specific similarity manipulation, subjects given the similarity transfer task should show grammaticality effects comparable to those of categorical transfer subjects.

Consequently, the present experiment compared the transfer performance of subjects given the usual categorical transfer task with that of subjects asked to judge the similarity of each transfer item to the training stimuli.

Method

Materials

The new artificial grammar is shown in Figure 7.1. To accomplish the objectives outlined above, it was necessary that this grammar be far more complex than that used in the first three experiments. This complexity allowed for the production of many more short letter strings (less than 8 letters in length), as well as increased numbers of item pairs that differed by only one letter at initial and internal letter positions. Since subjects in the preceding experiments consistently judged items produced by the addition alteration to be less grammatical and less "old" than substitution items, only substitution items were used in the present experiment.

The generation and selection of the items was done with the aid of an Apple II microcomputer. The grammar was programmed into the computer[7.1] and all possible items, 3 to 7 letters in length, were generated. The computer was then programmed to produce pairs of items of the same length that differed by only a single letter. To counteract the low rate of initial and internal letter differences in the stimuli of the previous experiments, the selection of pairs was biased toward differences in the initial position, followed by differences in the second letter position, and so on. In this way, the number of item pairs in which differences occurred in the initial and internal letter positions was maximized. Even so, pairs of items with differences in initial letter positions were rare. This procedure yielded 75 pairs of "close" items. The computer was then programmed to select two lists of 16 pairs each according to the following criteria: (1) The frequencies of

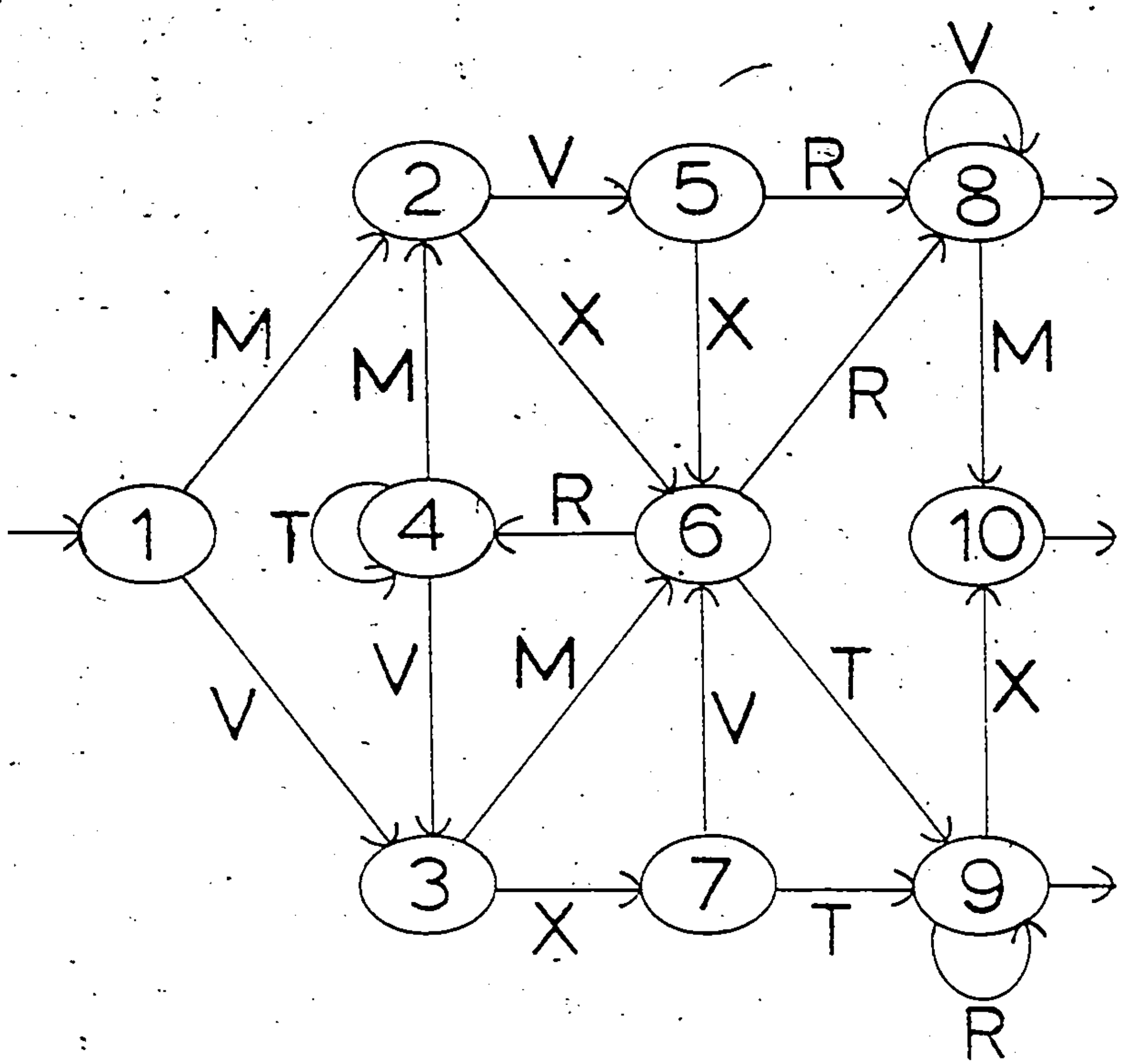

Figure 7.1: The artificial grammar used in Experiments 4, 5, and 6.

pairs with changes in the initial, middle and terminal letter positions were to be balanced as completely as possible across the two lists. (2) Differences in the frequency with which the 24 node to node transitions of the grammar were used were to be minimized both across the pairs within a given list (i.e., across training and transfer items) and across lists. This was an attempt to ensure that the two lists would be equally representative of the grammar, and that the transfer items would not differ systematically from the training items. (3) Each item selected as a transfer item had to be at least two letters in position different than all of the training items in both lists save for its matched training item, and each item chosen as a training stimulus for a given training list had to be at least two letters in position different from each of the remaining training items in the list. No approximation to this criterion was allowed. This criterion was to ensure that each transfer item, when used as a "close" transfer item, would have one and only one close analogy in the training list, and would have no close analogy when used as a "far" transfer item. (4) Differences in the frequency of items of different lengths were to be minimized across the two lists.

Because of the 10^{30} possible pairs of lists that could be selected,[7.2] the computer constructed the lists at random and then tested the lists for their ability to meet the design criteria. If the present lists exceeded some prior lists with respect to at least one of the design criteria, the present lists were stored and the prior set discarded. In this way the sets of lists that up to that point best satisfied each of the criteria, each pair of criteria, and so on were retained.

The selection process continued for a period of about 100 hours, and over 5000 different sets of lists were assessed. As it turned out, the set of two lists that best satisfied all of the design criteria simultaneously, was also the set that, with the exception of criterion (4) best satisfied each of the criteria independently. On the one exception, it just missed being the best set. Throughout the selection procedure, this was almost always the case. A set of lists that at a given point best satisfied one of the criteria almost inevitably best satisfied all of the remaining criteria. This suggests that the properties of the lists that are reflected in these criteria are extensively inter-correlated. Consequently, conclusions drawn with respect to only one of these properties should be viewed with some caution since it is unlikely to be independent of the many other properties of the lists of items. For example, sets of items that differ in the grammatical sub-rules used to produce the items also tend to differ in the length of the items they contain. Thus, finding that subjects respond differently to sets of items of different grammatical structure would not necessarily indicate that the effective variable was the grammatical structure of the item sets.

Statistically, the two resulting lists of 16 item pairs each do not differ from each other. Nor do the items designated as transfer items differ significantly from the training items with respect to any of the relevant selection criteria.[7.3] Thus, at least as assessed by the above criteria, the two lists of training items are equally representative of the grammar. Similarly, with the exception of being "close" to one item of its appropriate training list, the grammatical transfer items associated with one training list are equally similar to the training items of both lists. This is a more extreme degree of matching across training lists than was accomplished in the first three experiments, and, hence, provides a more extreme test of the specific similarity manipulation. The non-grammatical transfer items were produced from the grammatical transfer items according to the non-grammatical substitution operations shown in Table 7.1. The two lists of training items and the grammatical and non-grammatical transfer items are

Table 7.1
Production of Non-grammatical Items

The non-grammatical substitution operations used to produce the non-grammatical transfer items for Experiments 4, 5 and 6.

Where the grammatical transfer stimulus differed from its closest training stimulus by a:	The non-grammatical transfer stimulus differed in the same letter position by a:
M	R
R	M
V	T
X	T
T	X

listed in Table 7.2.

Each of the training lists was divided into four sub-lists of four items each, and 4 random orders of the items in each sub-list were generated. The 64 transfer items were randomly divided into four, 16 item lists, once for the initial test and again for the re-test. Thus, there were 8, 16 item transfer lists, and each item occurred twice.

One-half of the transfer items were selected to serve as distractor items for the recognition test. These were chosen so as to maintain the orthogonal structure of the transfer test. Thus, there were eight items from each of the combinations of grammaticality and similarity, for a total of 32 distractor items. Two recognition tests were constructed with these items, one containing the 16 list 1 training items and the other containing the 16 list 2 training items. Unlike the recognition tests used in the previous three experiments, the distractor items were the same for both recognition tests. The 48 items of each recognition test were randomly divided into three lists of 16 items each.

A final interview questionnaire was constructed consisting of 8 statements of possible strategies that the subjects may have used to fulfill the requirements of the categorical transfer task. These are listed in Appendix B. One-half of the questions referred to rules, while the remainder referred to item similarity. Orthogonal to this, one-half of the statements were general and the remainder specific. Finally, for each of the statement type (rule or similarity) by specificity (general or specific) combinations there was one statement that referred to the acceptance of items as obeying the rules and another statement that referred to the rejection of items.

Subjects

Thirty-two McMaster University undergraduates served as subjects in each of

Table 7.2
Training and Transfer Items

The training and transfer items used in Experiments 4, 5 and 6. One-half of the subjects were trained with training list one, and the remainder were trained with training list two. All subjects were tested, twice, with all 64 of the transfer items. Transfer items in the same row as a given training item are the "close" items for a subject trained with that item. Transfer items associated with the alternate training list are the "far" items for a subject trained with a given list.

Training Items		Transfer Items	
		Gram	Non-gram
List 1	MXRVXT	MXRMXT	MXRRXT
	VMTRRRR	VMTRRRX	VMTRRRT
	MXTRRR	VXTRRR	TXTRRR
	VXVRMXT	VXVRVXT	VXVRTXT
	VXVRVM	VXVRVV	VXVRVT
	VMRVVVV	VMRVVVM	VMRVVVR
	MXRTMVR	MXRTMXR	MXRTMTR
	VMRMXTR	VMRVXTR	VMRTXTR
	MXR	MVR	MTR
	VMRVXVR	VMRVXVT	VMRVXVX
	MVRVM	MXRVM	MTRVM
	VMRMVRV	VMRMXRV	VMRMTRV
	VMRMVXR	VMRMVXT	VMRMVXX
	MXRTVXT	MXRTMXT	MXRTRXT
	MXRMVXR	MXRMVXT	MXRMVXX
	MVXTR	MVXTX	MVXTT
List 2	MVXRM	MVXRV	MVXRT
	MVXRMVR	MVXRMXR	MVXRMTR
	VXVT	VXVR	VXVM
	MXRMXRV	MXRMVRV	MXRMTRV
	MXRTVMT	MXRTVMR	MXRTVMM
	VMT	VMR	VMM
	MVXRVMR	MVXRVMT	MVXRVMX
	MVXRVVV	MVXRVVM	MVXRVVR
	MXRMXRM	MXRMVRM	MXRMTRM
	VMRVMT	VMRVMR	VMRVMM
	MXRVVVM	MVRVVVM	MTRVVVM
	VXVRMXR	VXVRMVR	VXVRMTR
	MXTRRRX	VXTRRRX	TXTRRRX
	VXVTRRX	VXVTRRR	VXVTRRM
	MXTRRX	VXTRRX	TXTRRX
	VMRTMXT	VMRTVXT	VMRTTXT

the two transfer conditions, for a total of 64 subjects. Most participated in exchange for a course credit in introductory psychology. The remainder, assigned at random to the two conditions, were paid $3.00 each to participate. From four to 20 subjects were tested at a time in a classroom setting, and the two transfer conditions were deliberately mixed within a session.

Procedure

All of the materials and instructions were presented to the subjects in booklets. As in the first three experiments, the pages of the booklets during the training phase were back-printed to prevent the subjects from viewing the contents of the previous and the subsequent pages.

For training, all 64 subjects were instructed to study carefully the four items on each page and then to recall as many of the four items as they could remember on a following blank page. One-half of the subjects were presented with list 1 training items, while the remainder received list 2 training items. Order of the 4 sub-lists within each training list was systematically rotated across subjects within each of the two transfer conditions. Subjects received all four random orders of a given sub-list before proceeding to the next sub-list. Between sub-lists, the training instructions were repeated, and the subjects were informed that they would be receiving a list of four new items. Rate of presentation was subject-paced. At no time during the training phase were the subjects informed about the number of training sub-lists that they would receive or that their task consisted of anything more than the training phase.

Following training, all subjects were presented with one of the two recognition tests that was appropriate to the training list that they had received. Two orders of the three, 16 item recognition lists were used, and these were counter-balanced across subjects. Each 16 item recognition list was presented on a separate page, and each item was preceded by a blank in which the subjects were to indicate their judgement of each item. The subjects were instructed to use a six-valued response scale, where the numbers 1 to 6 represented the judgements "sure old", "fairly sure old", "guess old", "guess new", "fairly sure new" and "sure new", respectively. This scale was repeated for the subjects on the over-leaf of each preceding page such that the subjects always had the response scale in view while making their recognition judgements.

After the recognition test, all subjects were presented with the transfer test. Each transfer item was presented twice, for a total of 128 separate judgements, although the subjects were not informed of this fact. Each 16 item transfer list was presented on a separate page, and each item was preceded by a blank in which the subjects were to indicate their judgement of each item. Subjects in the Categorical Transfer condition read instructions to the effect that all of the items that they had studied during training were constructed according to a complex set of rules and that their task was to indicate which of the new transfer items were probably constructed according to the same set of rules. As in the previous experiments, this was the first mention of the rules that underlay the letter strings. The subjects were instructed to use a six-valued response scale similar to the one used in recognition. In this case, the numbers 1 to 6 represented the judgements "sure obeys the rules", "fairly sure obeys the rules", "guess obeys the rules", "guess does not obey the rules", "fairly sure does not obey the rules", and "sure does not obey the rules", respectively.

Subjects in the Similarity Transfer condition read instructions to the effect that some of the items in the lists of new items that they would receive were constructed to be very similar to at least one of the training items, while the remainder were constructed to be very dissimilar to any of the training items. The subjects were asked to indicate for each transfer item whether they thought the item was similar or dissimilar to the items they had been asked to learn. As in the Categorical Transfer condition, the subjects were instructed to indicate their judgement for each item using a six-valued response scale. Here, the numbers 1 to 6 represented the judgements "sure similar to at least one training item", "fairly sure similar to at least one training item", "guess similar to at least one training item", "guess not similar to any training item", "fairly sure not similar to any training item", and "sure not similar to any training item", respectively. For both transfer conditions, the response scale was repeated on the over-leaf of each preceding page.

One-half of the subjects in each transfer condition received the lists of the transfer test and re-test in one order, while the remainder received them in the opposite order. The sequence of items within each transfer test and re-test was fixed across subjects. At no time during either the recognition test or the transfer test were the subjects informed about the frequency of the various alternatives.

Only subjects in the Categorical Transfer condition received the interview questionnaire. Each accept-reject pair of the four statement types was printed on a separate page with a blank preceding each statement in which the subjects were to indicate their response to the statement. One random order of the four pages was used for all subjects. The subjects were instructed to assign a number between zero and 100 to each statement to indicate their estimate of the percentage of items for which they had used that particular strategy. They were told that since they may have used more than one strategy for any given item, and may have used strategies not mentioned in the interview, there was no reason to make their estimated percentages sum to 100.

Results and Discussion

Training

The training phase of the booklets was scored for the number of items correctly recalled. These data were subjected to a three-way analysis of variance with subjects nested within transfer condition. The combination of four training trials per each of four sub-lists resulted in 16 cells per subject. The mean percentages of items correctly recalled per trial per sub-list for each of the two transfer conditions are shown in Figure 7.2.

As would be expected since the subjects in the two transfer conditions were treated identically during training, transfer condition was not significant either as a main effect or in interaction with with any of the remaining variables in the training phase of the experiment. The 64 subjects as a whole correctly recalled an average of 53% of the items per trial per sub-list. There was a significant effect trials. The subjects correctly recalled significantly more items on later than earlier trials of each sub-list (trial 1 = 42%; trial 4 = 60%). In addition, there was a significant increase in the mean number of items correctly recalled as a function of sub-list. More items were correctly recalled on later sub-lists (sub-list 4 = 55%) than on earlier sub-lists (sub-list 1 = 47%). Thus, the subjects demonstrated both intra- and inter-list transfer. Finally, there was no trials by sub-list interaction. The increment due to trials on earlier sub-lists was essentially the same as that on later sub-lists.

Recognition

For both the recognition and categorical transfer tasks two different methods of scoring the data were used: (1) the number of "old" responses for recognition or the number of "grammatical" (or "similar") responses for categorical transfer (i.e., <= 3 on the 6-point response scale) for each item type, or (2) the mean response (average response on the 6-point response scale) for each item type. The results of

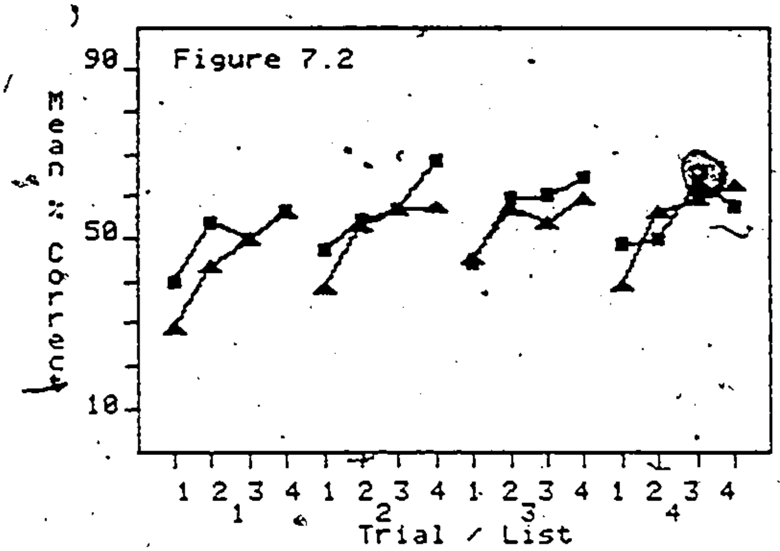

Figure 7.2: Training Phase. Mean percentage of items correctly recalled as a function of trials and sub-list for Categorical (squares) and Similarity (triangles) transfer conditions.

this experiment and those that follow were analyzed for each of these response measures. With a few minor exceptions (listed in Appendix C), the results for both measures were virtually identical. Consequently, to maintain compatibility with the previous experiments the results are reported in terms of the frequency of responding "old" or "grammatical" (or "similar") to the various item types. The analysis of variance summary tables for both analyses may be found in Appendix C.

The major results of the recognition phase of the experiment are shown in Table 7.3. As in the previous experiments, hits were analyzed separately from false-positive responses. The mean percentage of training items recognized as "old" for the 64 subjects as a whole was 72%, and, as would be expected, the two transfer conditions did not differ significatly. For false-positive responses, the mean percentage of items incorrecly labelled as "old" was 29% and, again, the two transfer conditions did not differ significantly. As in the previous experiments, there were significant effects of both specific similarity and grammatical status, with grammaticality being the significantly larger of the two effects ($z = 2.62$). The fact that both similarity distance and grammatical status were significant in the recognition phase of this experiment in which recognition preceded categorical transfer suggests that the concern over these results in the previous experiments as having been a consequence of the subjects' experience during categorical transfer was unfounded. Both specific similarity and grammaticality exert their effects in recognition apparently regardless of whether recognition occurs before or after categorical transfer.

Treated as a categorical transfer task, the subjects' mean percentage correct in discriminating grammatical from non-grammatical items was 59%, although, as usual, their mean pseudo-Pc was higher at 63%. This result is important since it indicates that subjects may evidence what looks like categorical transfer even if they are not

Table 7.3
Recognition Phase

The mean percentage of items labelled "old" for each transfer condition as a function of the specific similarity and the grammatical status of the items. The first eight columns are the results of the recognition phase analysed as a categorical transfer task, and their interpretation is the same as that for categorical transfer. The remaining three columns refer to the mean percentage of recognition hits, the mean percentage of recognition false-positives, and the mean percentage of recognition items correctly labelled, respectively.

	Close		Far				Actual	Pseudo	Recognition		
	Gram	Non-G	Gram	Non-G	G-NG	C-F	Pc	Pc	Hits	Fa	Pc
Categorical Transfer	42.6	25.4	35.2	17.2	17.6	7.8	58.8	62.7	73.4	30.1	71.1
Similarity Transfer	39.5	23.4	34.4	13.7	18.4	7.4	59.2	62.9	70.7	27.3	71.8
	41.0	24.4	34.8	15.4	18.0	7.6	59.0	62.8	72.1	29.9	71.4

aware that there exists a category to be generalized. Remember that up to this point in the experiment, no mention of the underlying grammar nor the categorical transfer task had been made to the subjects. Other demonstrations of categorical transfer effects during recognition (e.g., Neumann, 1974; Omohundro, 1981) typically have occurred in investigations in which training consisted of category discrimination, and in which, consequently, subjects were aware of a category to be generalized. The present finding, then, suggests that not only may subjects evidence categorical transfer if they are unaware of the category during training (as demonstrated in the previous experiments), they also may do so when they are unaware of the category during a test of transfer.

Categorical Transfer

The major results of the categorical transfer phase, summarized as in the previous experiments, are shown in Table 7.4. The frequencies of responding "grammatical"or "similar" (scale responses <= 3) were subjected to a five-way analysis of variance with subjects nested within transfer task. The factorial combination of two levels for each of the independent variables of grammaticality, specific similarity, pass, and old-new (present or absent on the preceding recognition task) resulted in 16 cells per subject with 8 responses per cell. This analysis, as well as that based upon mean response, may be found in Appendix C.

One particularly remarkable result of this phase of the experiment was that the two transfer conditions did not differ significantly from one another on overall responding. Subjects in the Categorical Transfer condition labelled an average of 48% of the transfer stimuli as "grammatical" while Similarity Transfer subjects labelled an average of 46% as "similar". In fact, the only major difference between the two transfer conditions is that Similarity Transfer subjects, not surprisingly given the task set for them, were significantly more sensitive to the specific

Table 7.4
Categorical Transfer Phase

The mean percentage of items labelled "grammatical" or "similar" for each of the transfer conditions as a function of the specific similarity and the grammatical status of the items. Column G-NG refers to the mean percentage difference of positive responses between grammatical and non-grammatical items. Column C-F refers to the mean percentage difference of positive responses between close and far items. The column labelled Actual Pc refers to the mean percentage of transfer stimuli correctly labelled. Pseudo Pc refers to the mean percentage of close grammatical and far non-grammatical items correctly labelled. Relative Frequency refers to the mean percentage of items receiving positive responses.

	Close		Far		G-NG	C-F	Actual	Pseudo	Rel.
	Gram	Non-G	Gram	Non-G			Pc	Pc	Freq.
Categorical Transfer	58.3	44.8	49.1	38.9	11.9	7.6	55.9	59.7	47.8
Similarity Transfer	59.9	46.3	45.6	30.9	14.2	14.8	57.1	64.5	45.7
	59.1	45.6	47.4	34.9	13.0	11.2	56.5	62.1	46.7

similarity of the transfer stimuli than were Categorical Transfer subjects, as is shown (column C-F) in Table 7.4.

Subjects in both conditions labelled as "grammatical" or "similar" significantly more "close" items than "far" items, and significantly more grammatical items than non-grammatical items. As in the recognition phase, grammaticality was the larger of the two effects, although, in this case, not significantly ($z = 0.72$). Thus, the initial objectives of the experiment appear to have been met. Overall responding (an average of 46.7% of the items were labelled as "grammatical" or "similar") was greater than that in previous experiments, and a large effect of categorical status (i.e., grammaticality), at least as large as that of specific similarity, was obtained. This latter result is consistent with the results and notions of Homa et al. (1981) that larger category sizes lead to greater response rates, improved performance and, in their view, increased abstraction. But note, subjects in both transfer conditions evidenced this result, and did so to statistically identical degrees (in fact, if anything, the effect of grammaticality was larger for subjects in the Similarity Transfer condition - see column G-NG in Table 7.4). Since subjects asked to judge the similarity between transfer and training stimuli judged grammatical items to be more similar to training stimuli than they judged non-grammatical items to be, it seems reasonable to conclude that grammatical items are more like remembered training stimuli than are non-grammatical items. The consequences of this conclusion are three-fold. First, as was discussed in previous chapters, "matching" transfer items for their similarity to training stimuli is a difficult, if not impossible task. Despite the efforts in the present experiment to produce grammatical and non-grammatical items that were equally similar to the training stimuli by controlling such formal item characteristics as letters in position (which, as indexed by the specific similarity effect for the Similarity

Transfer condition, <u>is</u> a component of inter-item similarity), grammatical items still appear to the subjects to be more similar to training stimuli than do non-grammatical items. If it is considered that most of the evidence for implicit abstraction of structure has occurred in experimentsin which the similarity between training and transfer items was left uncontrolled, the weakness of the case for implicit abstraction in these tasks becomes apparent. Second, the results for the Similarity Transfer condition confirm the notion that categorical judgements based upon the similarity between training and transfer stimuli can effectively discriminate categorical members from non-members. Even with a large component of training-transfer inter-item similarity controlled for by the orthogonal manipulation of the specific similarity variable, subjects in the Similarity Transfer condition discriminated grammatical from non-grammatical items with an accuracy of 57%, which rises to 65% (pseudo Pc in Table 7.4) if specific similarity is confounded with grammaticality. Third, actively attempting to discriminate categorical members from non-members, as subjects were instructed to do in the Categorical Transfer condition, does not improve this discrimination beyond that obtained by Similarity Transfer subjects who were not even aware that such a distinction existed. As is shown in Table 7.4, subjects in the Categorical Transfer condition discriminated grammatical from non-grammatical items with an accuracy of 56% and a pseudo-accuracy of 60%. As mentioned, precisely the same results were obtained with false-positive responding in recognition when subjects in neither condition were aware of the categorical distinction. Taken in total, these results confirm the notions suggested in the discussions of the previous experiments that (a) the variable of grammaticality in these experiments represents a measure of training-transfer, inter-item similarity that is not captured by the specific-similarity manipulation, and (b) that categorical and recognition judgements in these experiments are based primarily on the subjects' judgements of the similarity between transfer items and specific training stimuli.

The results of the analyses of position of "violation" and the interview data serve to buttress the above conclusions. Before discussing these results, however, a few of the remaining effects of interest in the transfer phase of the experiment will be presented. As usual, the effect of pass was significant, and affected both transfer conditions equally. The 64 subjects as a whole judged more of the items on the first pass (48%) to be "grammatical" or similar to the training stimuli than they did for those on the second pass (45%). Since the two passes represent repeated judgements to the same items, it is clear that, as suggested in the previous experiments, the effect of pass is to reduce the apparent similarity between training and transfer stimuli. Given this interpretation, it seems reasonable to conclude that the interaction of pass with specific-similarity in the previous experiments represents a reduction in the subjects' ability to discriminate among the items on the basis of their specific similarity to the (memory of) training stimuli. Unfortunately, while in the appropriate direction, this interaction was not significant in the present experiment. Across experiments, however, the combined results provide support for the suggestion of Medin and Schaeffer (1978) and Hintzman and Ludlam (1980) that the effect of increasing time between training and test is to reduce the difference in training-transfer inter-item similarity between items that initially were "close" to training stimuli and those that were not.

The only other effect of interest during the transfer phase is that of whether or not having been on the prior recognition test affected the likelihood that a transfer item would be judged "grammatical" or more similar to the training stimuli. As a main effect, this variable was not significant, nor was there any suggestion that items that were on the recognition test were more likely to be labelled "grammatical" or judged to be more similar than items not on the recognition test. However, this variable was involved in the two remaining significant effects in the transfer phase of the experiment. One of these was a three-way interaction including the variables of transfer condition and grammaticality, while the other was a three-way interaction including grammaticality and specific-similarity. Neither of these interactions involved cross-overs of any sort. Hence, they imply simply changes in slope (size of effect) of the lower-order effects, rather than changes in direction of effect. Consequently, the generality of these lower-order effects is not compromised. Generally speaking, these significant interactions suggest that the effect of having been on the recognition test is to reduce the effect of grammaticality to a greater extent than the effect of specific similarity, with the effect on the magnitude of the grammaticality effect differing (slightly) for the two transfer conditions. Thus, while there are effects attributable to items having been or not been on the prior recognition test, these effects were those of moderating, but not qualitatively altering, the major effects of interest. Note, however, that any effect of the recognition variable during categorical transfer implies that the subjects' transfer decisions were affected by single, prior experiences.

The results of the analysis of position of change are shown in Table 7.5. Because of the limited number of transfer items whose position of change from the "closest" training stimulus was in the initial letter position, it was not possible to manipulate this factor orthogonally to all of the remaining within-subject factors simultaneously. It was possible, however, to distribute the items with changes in initial, middle, and terminal letters across the levels of the various factors such that, collapsed across the remaining variables, position of change was independent of each within-subject factor taken one at a time. Since transfer condition is a between-subjects variable, position of change was manipulated orthogonally to

Table 7.5
Position of Change

Mean percentage of transfer items labelled "grammatical" for each transfer condition as a function of grammatical status and position of change from the "closest" training item.

Transfer Condition	Item Status	Letter Position of Change Initial	Middle	Terminal
Categorical Transfer	Gram	43.8	68.5	49.4
	Non-gram	26.8	58.4	37.0
Similarity Transfer	Gram	49.8	58.1	48.4
	Non-gram	31.3	42.4	36.5

transfer condition. For the analysis, the data were collapsed across all variables except grammaticality and transfer condition, and the frequencies of initial, middle, and terminal letter position changes were weighted to equate the a priori frequencies of the three position changes. The analysis of the data transformed in this way may be found in Appendix C.

Concentrating first on the pattern of responding to non-grammatical items, it can be seen in Table 7.5 that the results are a virtual replication of the position of violation finding reported by Reber and Allen (1978) and Reber et al. (1980) for these artificial materials, and are consistent with the detection of spelling errors found with real words (e.g., Erlich & Rayner, 1981; Haber & Schindler, 1981). Subjects are more likely to detect correctly non-grammatical transfer items whose "violations" are in the initial or terminal letter positions than those whose violation occurs in a deep, internal letter position. Note, however, that this effect was equally true for the subjects' responses to grammatical items. That is, while the effect of position of change was significant as a main effect, it did not interact significantly with the grammatical status of the items, as shown in Table 7.5. Since these changes in the grammatical transfer stimuli are, by definition, not "violations" of the grammar, the results suggest that what the subjects are more readily detecting in the initial and terminal letter positions are simply changes or differences from the "closest" training items rather than "violations" of the underlying grammar. Once again, if another measure of specific similarity is unconfounded with grammaticality it can be seen that an effect that was originally attributed to grammatical knowledge is actually a function of the similarity between transfer items and particular training stimuli. The conclusion is made all the more clear if it is recognized that subjects in both transfer conditions demonstrated the effect in exactly the same way. There was no effect at all of transfer condition.

Subjects in the Similarity Transfer condition were just as likely as Categorical Transfer subjects to detect "violations" in the initial and terminal positions of non-grammatical transfer items, whereas Categorical Transfer subjects were no less likely than Similarity Transfer subjects to reject incorrectly grammatical transfer stimuli that differed from their "closest" training stimulus in the initial or terminal letter positions.

I must admit to being at bit surprised at the results for position of change, particularly the lack of difference between grammatical and non-grammatical items with respect to changes in the initial letter position. However, further discussion of these results, and in particular discussion of how they relate to what the subjects may know explicitly about the grammar, will be deferred to the discussion of the subsequent experiments in which the grammatical status of the transfer items does appear to modify the effect of position of change.

The lack of effect of transfer condition in this experiment is remarkable. In fact, with the exception of the Similarity Transfer subjects being more sensitive to the specific similarity of the transfer items, the two conditions seem virtually interchangable. The only conclusion that seems reasonable under these circumstances is that subjects asked to sort the transfer stimuli on the basis of grammaticality do so on the basis of similarity to remembered training stimuli. As is discussed next, the similarity to remembered training stimuli is what the subjects in the Categorical Transfer condition reported as the primary basis of their decisions.

Interview

The results of the interview phase of the experiment are shown in Table 7.6, which presents the Categorical Transfer subjects' mean estimated percentage of items for which each of the eight strategies were used. The factorial combination of the three factors of similarity versus rules, general or specific, and acceptance or

Table 7.6
Interview Phase

Categorical Transfer subjects' mean estimated percentage of transfer decisions for which each of the eight strategies were used.

		Rules	Similarity	Mean
Specific	Accept	28.0	68.2	44.1
	Reject	40.2	52.5	46.3
General	Accept	41.6	48.0	44.8
	Reject	44.4	61.3	52.8
Mean		38.5	55.5	47.0

rejection resulted in eight judgements per subject. The subjects' estimated percentages were subjected to a three-way analysis of variance. The primary result is that subjects estimated that they used similarity to training stimuli significantly more often than rules in making their categorical transfer decisions. By their reckoning, they used similarity for an average of 55% of their decisions and rules for an average of 39% of their transfer judgements. This result is consistent with the conclusions drawn above about the relative importance of similarity during categorical transfer, and suggests further that, to a certain extent, the subjects are aware of the judgement processes that they use, if not the specific information upon which the judgements are based. This suggestion raises the possibility that some aspects of the subjects' knowledge about the bases of their judgements are less implicit that previously suggested. Failure to acknowledge the use of rules and the like as the basis of their judgements only implies that the subjects' knowledge is implicit if the subjects actually use the rules attributed to them. Clearly, if the subjects are not using abstract knowledge when making their judgements, their failure to attribute their behaviour to knowledge of the abstract structure of the category is hardly surprising.

The only other significant effect during the interview was an interaction involving all three of the factors of the interview phase. As is shown in Table 7.6, this interaction arose because, with the exception of statements about the use of specific similarity, subjects estimated that they used both rule and similarity strategies more often to <u>reject</u> than to accept items as "grammatical". Specific similarity was the only strategy that subjects acknowledged as being used more often to accept items as "grammatical" than to reject them. This result appears implies that the subjects are aware that they use the strategy that I have been suggesting they use: reject all items that are generally unfamiliar and accept only those items that are highly similar to <u>specific</u>, remembered training stimuli. Note that, according to the subjects, general similarity to the training stimuli (which, for example, could be indexed by an item's similarity to the prototype of a given training list) is <u>not</u> the primary basis for the acceptance as grammatical of the transfer stimuli. Also note that, according to the subjects, neither specific rules nor general properties of the training items play an overwhelming role in identifying transfer items as "grammatical". Thus, since specific similarity provides a consistent account of the subjects' categorical transfer judgements for both transfer conditions, and since the subjects in the Catgorical Transfer condition acknowledge specific similarity as the primary basis for their acceptance of transfer items as "grammatical", I see little reason not to take them at their word.

Chapter 8

EXPERIMENT 5

Up to this point in the series of experiments, similarity between training and transfer stimuli has been determined experimentally by the physical or formal properties of the items. An item has been considered to be a "close" match of a training stimulus if it differs from the training stimulus by no more than a single letter in position. Similarity, however, is not strictly or even primarily a function of common formal characteristics between events, except in so far as the formal characteristics are correlated with the informal characteristics between one event and another. That is, it is the correspondance between the encoding of events that determines the extent to which two events will be judged to be similar to each other. Thus, counting "features" in common between events, even assuming that Tversky's (1977; Gati & Tversky, 1982) model of similarity as a feature-matching process is valid, only provides a measure of similarity as assessed by the subjects to the extent that these "features", and no others, are those encoded and represented by the observer for both of the events in question. As has been argued throughout this thesis, the failure to meet this criterion may underlie many of the purported demonstrations of transfer on the basis of abstracted structure.

In the first three experiments, the effect of encoding (training) condition during recognition and categorical transfer was attributed to variations in the degree to which the encoding of transfer stimuli "matched" the encoding of specific training stimuli. It was argued that for the transfer stimuli presented to the subjects, an encoding condition that resulted in too few or too many "matches" among the transfer items would hurt performance by providing little basis for discriminating among the transfer stimuli.

In the next two experiments, a different procedure that allows for direct experimental control over subjects' encoding of specific items was used. This procedure, the heart of the paradigm for demonstrating encoding specificity or context effects in recognition, is similar to one used in a series of experiments by Light and Carter-Sobell (1970; see also Hunt & Ellis, 1974; Donaldson, 1981). In Light and Carter-Sobell's (1970, Experiment III) paradigm, the interpretations of homograph nouns in adjective-noun phrases during the study or training phase are biased toward one of their meanings by the presence of the adjectives. Thus, for example, the meaning of the homograph noun "jam" is biased toward one of its possible meanings by the context provided by the adjective in the phrase "strawberry jam". During the recognition test the homograph noun is presented either in a similar context to that encountered during training (e.g., "raspberry jam") or a new context that biases a different interpretation of the noun (e.g., "traffic jam"). The subjects' task is to indicate which of the nouns (not phrases) presented during recognition were encountered during training. The basic results, which have been replicated several times (e.g., Hunt & Ellis, 1974), are that homograph nouns presented in similar contexts across training and test are recognized better than homograph nouns whose contexts are changed to bias a dissimilar interpretation from training to test. Similar results to these are obtained if the familiarity of test contexts is unconfounded with the similar-changed context manipulation (Donaldson, 1981; e.g., pre-exposure of the context adjective "traffic" in the above example). Thus, although recognition is further enhanced if the adjective-noun pairing is literally identical across training and test (Light & Carter-Sobell, 1970) and even though the familiarity of the test context appears to moderate the effect of repetition in these tasks (Donaldson, 1981), the context effect on recognition does appear to occur even if all aspects of variations in the literal familiarity (i.e., formal characteristics) of the context–noun pairings are eliminated or controlled. The context effect, then, appears to be a function of changes in the informal (non–literal) properties of the stimulus events. As such, it is a clear demonstration of encoding specificity.

The next two experiments were designed to extend the encoding specificity effect to categorical transfer. In this thesis, analogy to individual instances has been described as an extension of a process similar to misrecognition. It seems reasonable to suppose that manipulation of the encoding context across training and test in the transfer paradigm should influence the extent to which this misrecognition occurs in much the same way as recognition was affected in Light and Carter–Sobell's (1970) experiment. Specifically, more positive responses should be produced to transfer items in contexts similar rather than dissimilar to that of their "closest" training items. For example, imagine that two of the items and associated contexts that subjects are asked to memorize during training are MVX – MOOSE and VMT – TIGER, and that during transfer (recognition or categorical) they encounter, also in associated contexts, the "close" items MVR – DEER, MVT – LION, VMR – ELK, and VMX – LEOPARD. By analogy to the results of Light and Carter–Sobell's (1970) experiments, subjects should preferentially respond to MVR – DEER and VMX – LEOPARD, despite the fact that each transfer item differs from its "closest" training item by only a single letter and they all occur in equally "familiar" contexts. Since this relation between an item and its encoding context may be manipulated independently of categorical status of the item itself, such an effect would demonstrate that transfer was a function of the <u>specific</u> <u>encoding</u> of the items.

The context manipulation in the present experiment was implemented exactly as outlined in the above example. As in the Label training condition in Experiment 2, each training item was associated with a unique animal name. Unlike Experiment 2, however, each transfer item during recognition and categorical transfer was also associated with an animal name, some of which referred to animals similar to that referred to by the animal name associated with the "closest" training stimulus, with the remainder referring to animals similar to that associated with some training stimulus other than the "closest".

Method

Materials

The materials were the same as those used used for the Categorical Transfer condition in Experiment 4 with the exception of the modifications required to present the items with labels during training, recognition and categorical transfer. Sixteen common animal names were chosen as the training item labels. The names were randomly assigned to the 16 items in each training list. Thus, the labels used for training list 1 were the same as those used for training list 2. For each of these training labels, the names of two similar animals were produced as labels for the transfer items. For example, for the training label DEER, the animal names ELK and MOOSE were chosen, and for the training label MACAW, the names PARROT and TOUCAN were used.

The transfer labels were assigned to the transfer items so that one-half of the grammatical and one-half of the non-grammatical items would have the name of the animal that was similar to the name of the animal assigned to the closest training item. Hence, these items were "appropriately" labelled. The remaining transfer items were assigned labels by swapping what would have been the appropriate labels, across pairs of items. For example, if the transfer items VMRMVXX and MXRTMXT would have been appropriately labelled had they been assigned the names MUSKRAT and PARROT, respectively, they were inappropriately labelled by assigning PARROT to VMRMVXX and MUSKRAT to MXRTMXT. Thus, these items were "inappropriately" labelled. The complete list of label to item assignments is given in Appendix A. Note that across the 64 transfer items each transfer label occurred twice, once with the "close" items for subjects trained with a given list, and once with the "far" items. Thus, by themselves, the transfer item labels provide no basis for discriminating among the items with respect to either grammaticality or specific similarity. Also note that while one-half the transfer items were "appropriate" items, this distinction is only meaningful for the "close" transfer items of subjects trained with a given list (i.e., since subjects trained with a given list never see the label to training item pairings of the alternate training list, all "far" items are, in a sense, "inappropriately" labelled). However, the distinction for "far" items will be retained for analysis since it provides a direct control for potential a priori item, label, and/or item-label pairing differences. Consequently, any effect of the appropriateness of the labelling should occur only on "close" items.

Subjects

Thirty-two McMaster University undergraduates served as subjects. Most participated in exchange for course credit in introductory psychology. The remainder were paid $3.00 each to participate. From 4 to 14 subjects were tested at a time in a classroom setting.

Procedure

The training phase was similar to that used in the Label condition of Experiment 2. The subjects were instructed to study carefully the items and associated names on each page and then to recall the items and their associated animal names on a following blank page. As in all of the preceding experiments, one-half of the subjects received training list 1, while the remainder received training list 2, and the subjects received all four sub-lists of one training list before proceeding to the next sub-list. Rate of presentation was subject-paced.

Following training, the subjects received the recognition and categorical transfer tests as outlined in Experiment 4. For both tasks, they were informed that while they may find the labels associated with each item to be helpful, their task was to judge the letter strings. As before, the six-valued response scales were explained to the subjects, and were repeated on the over-leaf of the pages of each test so that they were in view throughout both tasks.

After completion of the categorical transfer task, the subjects received the interview questionnaire as outlined in Experiment 4.

Results and Discussion

Training

The training data were analyzed as in Experiment 4. The mean percentages of items correctly recalled per trial per sub-list are shown in Figure 8.1. Like their label-trained counterparts in Experiment 2, the 32 subjects in this experiment demonstrated both intra-list and inter-list transfer. That is, there were significant effects of both trials and sub-lists, with no significant interaction of these two variables.

Recognition

The major results of the recognition phase of the experiment are shown in Table 8.1. The subjects' recognition accuracy was quite good, averaging 81.6% correct, although at least part of this successful recognition of the items is probably attributable to recognition of the labels of the "old" items. Surprisingly, analysing the recognition false-positive responses as a categorical transfer task revealed only one significant effect, grammaticality (column G-NG in Table 8.1). As discussed in the previous chapter, this effect may be the result of inter-item similarity being assessed at a higher level of similarity than that captured by the specific similarity manipulation, although it also may represent the use of categorical information during recognition. These interpretations of the effect of

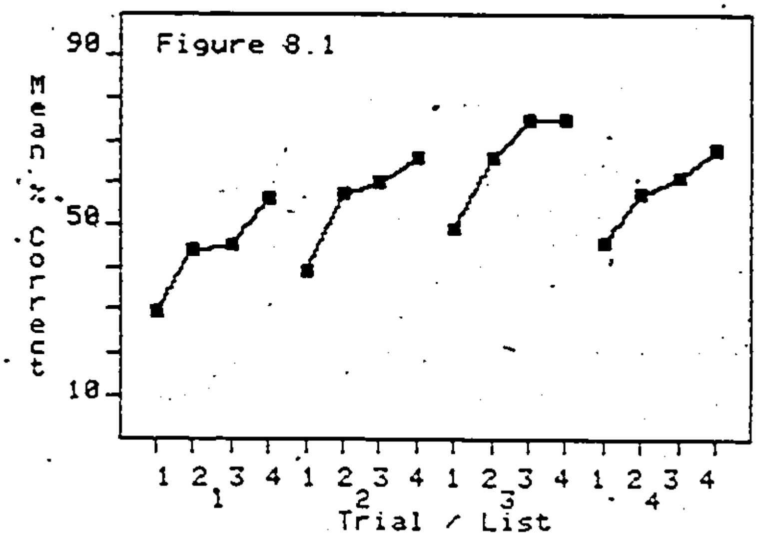

Figure 8.1: Training Phase. Mean percentage of items correctly recalled as a function of trials and sub-lists for Animal Names training.

Table 8.1
Recognition Phase

The mean percentage of items labelled "old" as a function of the specific similarity and the grammatical status of the items. The first eight columns are the results of the recognition phase analysed as a categorical transfer task, and their interpretation is the same as that for categorical transfer. The remaining three columns refer to the mean percentage of recognition hits, the mean percentage of recognition false-positives, and the mean percentage of recognition items correctly labelled, respectively.

	Close		Far				Actual	Pseudo	Recognition		
	Gram	Non-G	Gram	Non-G	G-NG	C-F	Pc	Pc	Hits	Fa	Pc
Animal Names	21.5	13.7	24.2	8.2	11.9	1.4	56.0	56.6	78.5	16.9	81.6

grammaticality during recognition are discussed in Chapter 10. Neither specific similarity nor the appropriateness of the item labels had any significant effect on false-positive responding, and the interaction of these two factors, while in the appropriate direction (see Figure 8.2), was also not significant. As is discussed below, this latter effect also failed to emerge during categorical transfer.

Categorical Transfer

The major results of the categorical transfer phase of the experiment are shown in Table 8.2. Both the specific similarity and the grammaticality of the transfer items were significant as main effects, with there being no significant difference in the magnitude of the two effects ($z = 0.15$).[8.1] The effect of grammaticality for the label-trained subjects in this experiment contrasts with results obtained with the Label trained subjects in Experiment 2 in which, it will be remembered, no significant effect of grammaticality was found. Whether the difference is due to the sheer presence of the labels during transfer or the greater number of training items during training for the present experiment is not known. However, in Experiment 4 in which the items were presented without labels, the effect of grammaticality was similarly enhanced relative to that of the earlier experiments – an effect that was attributed to the increased breadth of transfer arising from doubling the number of training items. Thus, it seems probable that the source of the increased effect of grammaticality in the present experiment relative Label trained subjects in Experiment 2 represents a similar "category size" effect.

The presence of the labels during transfer did produce one clear effect. Unlike Experiment 4 in which no main effect of items having been or not been on the prior recognition test was found, this variable was significant as a main effect in the present experiment. Subjects labelled as "grammatical" significantly more "old" items (those on the prior recognition test: mean of 56.1% labelled "grammatical")

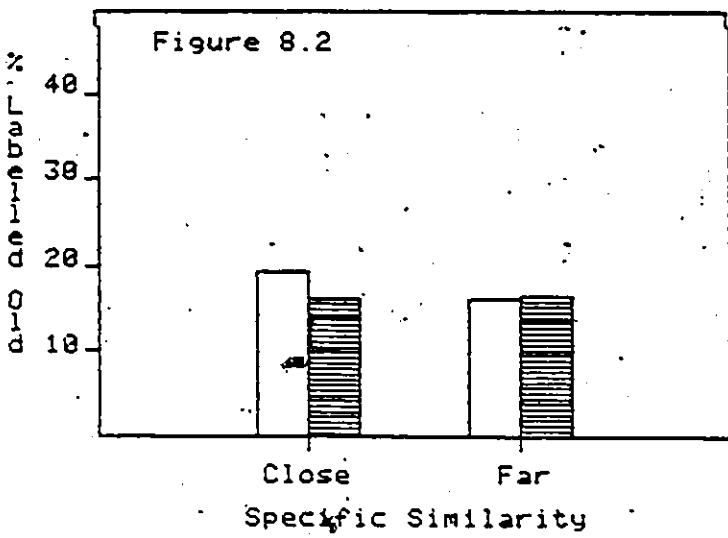

Figure 8.2: Recognition Phase. Mean percentage of recognition false-positive ("old") responses as a function of the specific similarity of the items and the appropriateness of the associated animal names. (Appropriate names: open bars; inappropriate names: hatched bars)

Table 8.2
Categorical Transfer Phase

The mean percentage of items labelled "grammatical" as a function of the specific similarity and the grammatical status of the items. Column G-NG refers to the mean percentage difference of positive responses between grammatical and non-grammatical items. Column C-F refers to the mean percentage difference of positive responses between close and far items. The column labelled Actual Pc refers to the mean percentage of transfer stimuli correctly labelled. Pseudo Pc refers to the mean percentage of close grammatical and far non-grammatical items correctly labelled. Relative Frequency refers to the mean percentage of items receiving positive responses.

	Close		Far		G-NG	C-F	Actual	Pseudo	Rel.
	Gram	Non-G	Gram	Non-G			Pc	Pc	Freq.
Animal Names	62.5	50.7	52.4	44.7	9.8	8.0	54.9	58.6	52.6

than "new" items (mean of 49.1% labelled "grammatical"). However, because of the design of the experiment, this effect probably represents little more than subjects responding positively to familiar labels. The orthogonal manipulation of the appropriateness of the labels during recognition resulted in particular transfer labels being confounded with the "old"-"new" distinction during categorical transfer. While the effect suggests that the subjects were attending to the labels while making their transfer decisions, it does not necessarily indicate that they were responding to the training-transfer relation between the items and the labels, a necessary condition for demonstrating the effect of the appropriatness or encoding specificity manipulation.[8.2] In fact, there was no significant effect of the appropriateness variable, either as a main effect or in the expected interaction with specific similarity. This result suggests that the items and their labels were treated relatively independently of each other by the subjects. It should be noted though that, as in recognition, the interaction between specific similarity and the appropriateness of the labels, while not significant, was in the predicted direction (see Figure 8.3).[8.3] Thus, it is possible that the reason for the failure of the present experiment to evidence the expected encoding specificity effects is more a function of a weak manipulation possibly because of the arbitrary association between the items and their labels than a true lack of effect. This possibility is investigated in the next experiment.

Ignoring for the moment, then, the original concerns of the experiment, it can be seen that the results of the transfer phase are generally in accord with those of previous experiments. In particular, the results of the position of change and interview analyses generally replicate those of Experiment 4. Presented in Table 8.3 are the mean percentages of items labelled "grammatical" as a function of grammatical status and the letter position of change from the "closest" training item. As in

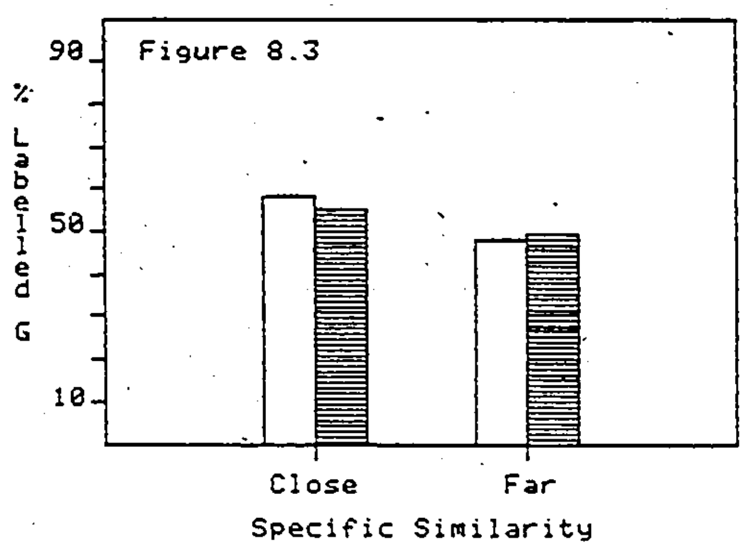

Figure 8.3: Transfer Phase. Mean percentage of items labelled "grammatical" as a function of the specific similarity of the items and the appropriateness of the associated animal names. (Appropriate names: open bars; inappropriate names: hatched bars)

Table 8.3
Position of Change

Mean percentage of transfer items labelled "grammatical" as a function of grammatical status and position of change from the "closest" training item.

	Letter Position of Change		
Item Status	Initial	Middle	Terminal
Grammatical	54.7	62.8	53.8
Non-grammatical	35.9	52.8	45.3

Experiment 4, there was a significant main effect of the position of change. Items with changes from their "closest" training item in initial or terminal letters recruited significantly fewer "grammatical" responses than did items with changes in internal letter positions. In addition, although the effect of letter position of change was more pronounced for non-grammatical items, particularly non-grammatical items with changes in the initial letter position, the interaction of grammatical status and position of change was only marginally significant ($p = 0.05$). Hence, as in Experiment 4, the effect of position of change was about the same for both grammatical and non-grammatical items. However, if it is assumed that the interaction effect is reliable for this experiment, the effect appears to be one of subjects accepting, relative to the other letter positions for each of the two item types, fewer non-grammatical items than grammatical items with changes in the initial position. Such a result is not all that surprising if it is considered that all grammatical items are produced with either an M or a V in the initial letter position. To produce a non-grammatical item with its position of change from its "closest" training (i.e., grammatical) item in the initial position required that some letter other than M or V be used. Hence, non-grammatical items with changes in the initial letter position differed from <u>all</u> other items in this rather salient letter position, whereas items, grammatical and non-grammatical alike, with changes in the other letter positions did not. That subjects detected such a radical difference more readily than the M for V alteration on grammatical items than the other letter position changes for both grammatical and non-grammatical items is consistent with their responding on the basis of inter-item similarity. However, it is also possible that the effect is due to a number of the subjects having recognized that the initial letter of the training items was always either an M or a V, leading them to reject (correctly) any items that violated this simple rule. Given the small size of the effect and the fact that subjects in the previous experiment apparently did not discover or at least did not use such information in their transfer judgements, the learning or use of even such a simple rule as this one is not all that common. Be that as it may, it is clear that the effect of position of change for the remaining two letter positions provides no support for the claim of Reber and his colleagues that it is violations of the underlying grammar that subjects more readily detect in terminal letter positions.

Interview

The results of the interview phase are presented in Table 8.4. As in the previous experiment, subjects estimated that they used similarity to the training items significantly more often than rules in making their transfer judgements. As is shown in Table 8.4, the subjects estimated that they used similarity for an average of 41.5% of their transfer judgements and rules for an average of only 25.4%. They also estimated that they used similarity significantly more often to accept than to reject items as "grammatical" while using rules relatively equally for both of these decisions. That is, while the subjects estimated that similarity played a larger role in all transfer decisions than did rules, they estimated that its major role was in the acceptance of items as grammatical. Moreover, as indicated by the significant interaction of all three of the factors of the interview phase, the particular type of similarity that the subjects estimated was important for the acceptance of items was the specific similarity between training and transfer items. As can be seen in Table 8.4, while subjects estimated that they used similarity more often than rules in making their transfer judgements, specific similarity was estimated to be used most often to accept items while rules, and specific rules in particular, were estimated to be used much less frequently for this purpose. Thus, like the subjects in Experiment 4, the subjects in the present experiment estimated that they used

Table 8.4
Interview Phase

Subjects' mean estimated percentage of transfer decisions for which each of the eight strategies were used.

		Rules	Similarity	Mean
Specific	Accept	19.5	50.6	35.1
	Reject	27.0	32.5	29.8
General	Accept	30.1	44.8	37.5
	Reject	25.9	38.1	31.6
Mean		25.4	41.5	33.5

similarity to specific training items as the primary basis of their acceptance of transfer items as "grammatical".

The results from the interview phases of Experiment 4 and the present experiment are consistent in suggesting that the the subjects' knowledge of the bases of their responses in these tasks is far less implicit than previously suggested. They are also consistent, however, in suggesting that the experimentally determined and subject attributed basis of responding during categorical transfer is not one of abstracted rules. Both the transfer results and the subjects' attributions implicate similarity to the memory of specific training stimuli and not abstracted rules as the primary basis for the acceptance of items as "grammatical". The experiment reported in the next chapter extends the specificity of the subjects' knowledge even further.

Chapter 9

EXPERIMENT 6

The encoding context manipulation in Experiment 5 was designed as a direct analogue of Light and Carter-Sobell's (1970) paradigm. However, it differed in several fundamental ways from their experiment, any or all of which may have been responsible for the failure to demonstrate the expected effects of encoding specificity. For example, for both the recognition and categorical transfer tests of Experiment 5, the encoding manipulation occurred on items that were merely similar ("close") rather than identical to the training items. It is possible that the effect of encoding specificity occurs only for stimuli that are literally identical from training to test, or at least far more similar than the items used in Experiment 5. A more likely possibility, however, is that the key difference between Experiment 5 and Light and Carter-Sobell's (1970) paradigm has to do with the relationship between the items and their contexts. In Experiment 5, the relationship between any given item and its associated context was strictly arbitrary, whereas such was clearly not the case for the homograph nouns and their associated contexts in Light and Carter-Sobell's (1970) experiments. That is, the contexts of the items in Light and Carter-Sobell's (1970) experiments were far more specific to their associated nouns than were the animal names to their associated letter strings in Experiment 5. Consider the likely consequences of repeating Light and Carter-Sobell's (1970) experiments using adjective-noun pairs in which the relationship between the nouns and their adjectives were as arbitrary as that of the items and labels in Experiment 5. Although it is quite difficult to produce adjective-noun pairings in which subjects can derive no meaningful interpretation of the resulting phrases, it does seem likely that arbitrarily pairing adjectives and nouns within Light and

Carter-Sobell's (1970) paradigm would have a detrimental impact on the magnitude of the encoding specificity effects they observed.

To mimic more closely Light and Carter-Sobell's (1970) paradigm requires that that the relationship between the items and their associated contexts be less arbitrary, that is, more specific to the particular items with which they are associated, than was the case for the materials in Experiment 5. In Experiment 2, the key factor apparently leading to the difference between the two encoding conditions was the specificity of the unique mnemonics to their associated acronym items. The unique mnemonics in that experiment apparently led to more constrained or more item specific encoding of the items than did the arbitrary animal name associates. Thus, the arbitrary animal name associates of the items used in Experiment 2 and Experiment 5 may be less effective in biasing a particular interpretation of the items than are the unique mnemonics, and for that reason may be less effective in promoting the expected effects of encoding specificity on recognition and categorical transfer. Since unique mnemonics apparently are more effective in biasing the encoding of the items, unique mnemonic phrases rather than arbitrary labels were used as the biasing contexts of the items in the present experiment. In all other respects, the procedure of the present experiment was identical to that of Experiment 5.

<u>Method</u>

<u>Materials</u>

The materials were the same as those used in Experiment 5. The only difference in the present experiment was that unique, mnemonic phrases were substituted for the animal names used in the previous experiment. Thus, the present experiment is an exact analogue of Experiment 5 with mnemonic phrases serving in the role of item labels. Sixteen unique themes were generated to be used as mnemonic associates of the 16 items in each of the two training lists, and the "appropriately" and "inappropriately" labelled transfer items were assigned mnemonic themes in the same way as the animal names were assigned to the transfer items in Experiment 5. In fact, each item retained its experimental status across experiments. For each training list, each of the 16 themes was associated with a training item and a related and an unrelated transfer item such that for one-half of the transfer items the related item was grammatical and the unrelated item non-grammatical, while for the remaining items the reverse was true. For the resulting six items of a given thematic set, the theme was expressed for each item using key words unique to that set. As in the previous experiments using mnemonic associates, each item was an acronym of the key words of its associated phrase. The complete list of mnemonic phrase to item assignments is given in Appendix A.

Subjects

Thirty-two McMaster University undergraduates served as subjects. Most participated in exchange for a double course credit in introductory psychology. The remainder received a single course credit and were paid $3.00 each to participate. From 2 to 14 subjects were tested at a time in a classroom setting, although a few were tested individually.

Procedure

The procedure for all phases of the experiment was the same as that used in Experiment 5.

Results and Discussion

Training

The mean percentages of items correctly recalled per trial per sub-list are shown in Figure 9.1. The 32 subjects in this experiment correctly recalled an average of 68.3% of the items per trial per list. As with Unique Mnemonic training

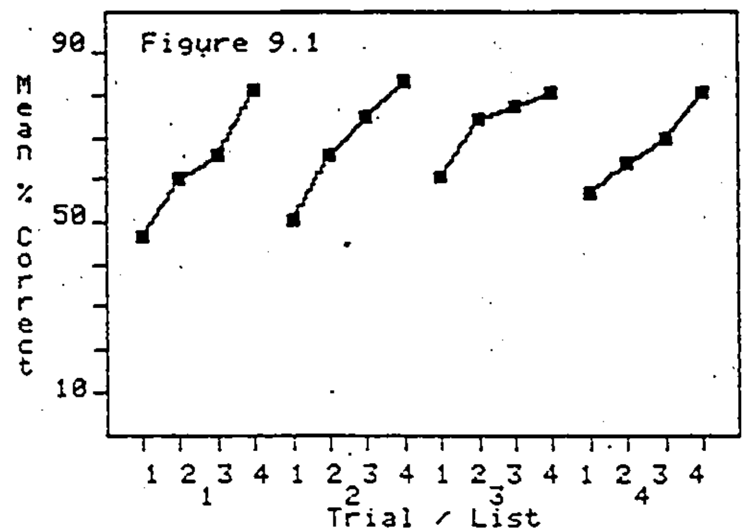

Figure 9.1: Training Phase. Mean percentage of items correctly recalled as a function of trials and sub-list for Mnemonic Phrases training.

in Experiment 2, there was a significant effect of trials. Subjects recalled significantly more items on the later trials with a given sub-list (trial 4 = 81.3%) than on the earlier trials (trial 1 = 53.9%). Unlike the subjects receiving Unique Mnemonic training in Experiment 2, however, the subjects in the present experiment showed no significant effect of sub-lists either as a main effect or in interaction with trials, as is shown in Figure 9.1. Thus, the subjects in the present experiment demonstrated intra-list transfer, but no inter-list transfer.

The Grammatical Mnemonics condition in Experiment 3 was the only other training condition in this series of experiments that failed to demonstrate inter-list transfer during training. In discussing those results, it was argued that the lack of inter-list transfer arose as a consequence of too much transfer from one training item to another brought on by the extended similarity of the associated mnemonics. Other results from that experiment, such as the limited intra-list transfer during training and the relatively poor recognition performance of the Grammatical Mnemonics subjects, were held to buttress this conclusion. On the other hand, the lack of inter-list transfer for Reber and Allen's (1978) paired-associate condition was attributed to reduced breadth of transfer, presumably arising from the individuation requirements of the paired-associate task. The lack of inter-list transfer for the training phase of the present experiment also probably arose from reduced breadth of transfer, representing a more extreme version of the Unique Mnemonic training in Experiment 2. This conclusion is suggested by both the relatively high level of intra-list transfer during the training phase of the present experiment and by the remarkable degree of recognition accuracy obtained by these subjects, discussed below. In fact, from the results of the training phase of the present experiment, it appears that the subjects did not generalize what they learned from their experience with a particular training item beyond the immediate item.

Under the more typical circumstances of Experiments 1 – 4, the extremely limited breadth of transfer suggested by these results would be taken to suggest that the subjects should demonstrate little transfer to the recognition and categorical transfer phases of the experiment. However, in these previous experiments, the items during recognition and categorical transfer were presented without labels or mnemonic phrases, potentially increasing encoding variability relative to the encoding of the training items and thereby reducing transfer to all items. In the present experiment, each transfer item was associated with a mnemonic phrase that was designed to bias the encoding of the item either in the direction of that of a particular training item for "close" appropriate items, or away from the encoding of any of the training items of a given list for all of the remaining transfer items. Thus, for "far" transfer items and inapropriately labelled "close" items, little contact should be made with the memory of particular training items, and transfer should be at a minimum. However, the similarity of the mnemonic themes between training and appropriately labelled "close" items, coupled with the near identity of the items themselves, suggests that whatever transfer does occur should do so preferentially to the appropriately labelled "close" items. That is, the specificity of the mnemonic phrases should actually enhance contact with the "closest" training experiences of these items, and, hence, increase transfer for these items relative to all other transfer items. For the training items themselves, the literal repetition of the associated mnemonics across training and the recognition test should result in subjects correctly detecting virtually every training stimulus. As is discussed below, this is precisely what was observed.

Recognition

Shown in Table 9.1 are the major results of the recognition phase of the experiment. As can be seen (recognition Pc in Table 9.1), the subjects were

Table 9.1
Recognition Phase

The mean percentage of items labelled "old" as a function of the specific similarity and the grammatical status of the items. The first eight columns are the results of the recognition phase analysed as a categorical transfer task; and their interpretation is the same as that for categorical transfer. The remaining three columns refer to the mean percentage of recognition hits, the mean percentage of recognition false-positives, and the mean percentage of recognition items correctly labelled, respectively.

	Close		Far				Actual	Pseudo	Recognition		
	Gram	Non-G	Gram	Non-G	G-NG	C-F	Pc	Pc	Hits	Fa	Pc
Mnemonic Phrases	13.7	8.2	7.4	3.5	5.9	5.5	52.3	55.1	96.3	8.2	93.3

remarkably accurate in recognizing the items. While at least part of this almost perfect performance is probably attributable to the literal recognition of the mnemonic associates of the training items rather than the items themselves, the majority is probably a function of the extreme item-specific transfer that these mnemonic associates engender. It is clear from the extreme disparity between the hit and false-positive rates shown in Table 9.1 that the subjects rarely considered anything but the training stimuli (items and associated phrases) to be "old", and rarely missed recognizing them. Apparently, as suggested above, the effect of including the mnemonic phrases during a test of recognition is to enhance contact with the specific training experiences for training items while, to a large degree, preventing this contact, or at least preventing misrecognition on the basis of such contact, for "new" items considered as a whole.

Treated as a categorical transfer task, the results of the recognition phase are typical of those of previous experiments. Both the grammatical status and the specific similarity of the transfer items were significant as main effects, and, although specific similarity was the larger of the two effects, this difference was not significant ($z = 0.34$), as is shown (G–NG and C–F) in Table 9.1. Importantly, however, the appropriateness of the item–mnemonic phrase pairings was also significant. Subjects produced almost twice as many false-positive responses to items associated with appropriate mnemonics (mean of 10.6%) than to items associated with inappropriate mnemonics (mean of 5.9%). As expected, this effect interacted significantly with the specific similarity of the items. In fact, as is shown in Figure 9.2, which presents the mean percentage of transfer items labelled "old" as a function of the specific similarity and the appropriateness of the mnemonic associates, the main effects of both of these variables arose entirely as a function of their interaction. As indicated by Fisher's least significant difference (LSD)

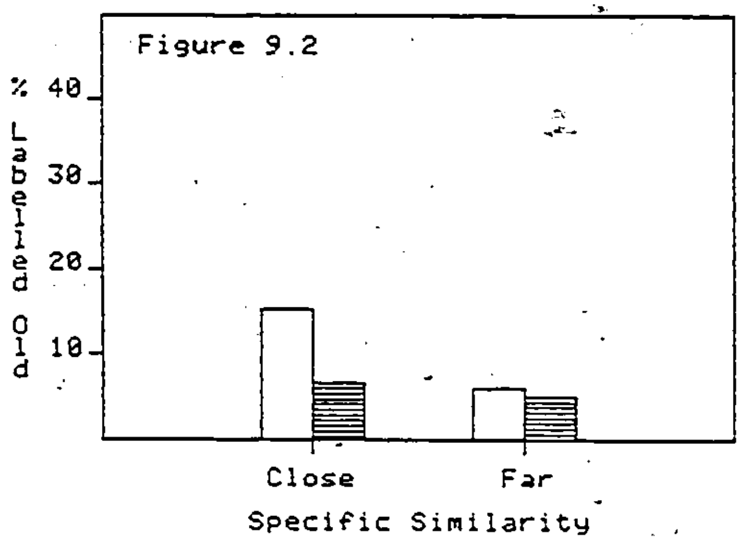

Figure 9.2: Recognition Phase. Mean percentage of recognition false-positive ("old") responses as a function of the specific similarity of the items and the appropriateness of the mnemonic associates. (Appropriate mnemonics: open bars; inappropriate mnemonics: hatched bars)

test, the significant main effect of specific similarity (or, equivalently, the appropriateness of the item–mnemonic pairing) was due to "close"-appropriate items recruiting significantly more false-positive responses than either "close"-inappropriate items or either type of "far" items, which did not differ significantly among themselves (LSD = 6.7%). Thus, being "close" was of no benefit in the recruitment of recognition false-positives unless the encoding context was appropriate. If the context was inappropriate to that of the "closest" training item, "close" items were treated no differently by the subjects than were "far" items, despite the difference in the formal similarity of these two types of items to the training stimuli.

The results of the recognition phase of this experiment extend the effects of encoding specificity on recognition reported by Light and Carter-Sobell (1970) and others by demonstrating that it is the breadth of transfer of the learning of individual items that is reflected in encoding specifity effects. Importantly, the present effect was demonstrated on recognition errors. That is, the same process that typically is cited to enhance recognition "hits" (e.g., Light & Carter-Sobell, 1970), and was clearly instrumental in the enhanced recognition performance of subjects in the present experiment, is also the process responsible for the significant variation in false-positive responses. Subjects can be led to accept as "old" distractor items that are similar, but not identical, to training items by varying the extent to which the encoding of the items is biased in the direction of specific training experiences. If the resulting encoding falls within the breadth of transfer of specific training experiences, the item is accepted as "old" by the subjects, regardless of whether the item actually is one of those previously experienced. Importantly, however, to be accepted as "old" by the subjects an item has to be both formally similar to some training experience and encoded as such.

Being formally similar is of no consequence unless the item is encoded so as to reflect the similarity. Aside from highlighting the points advanced in Chapter 3 about not attributing variations in recognition "hits" and "false-positive" responses to different processes, the results of the present experiment present a clear case for considering the typical recognition test as a transfer paradigm that measures the extent to which experiences gained in one task may be generalized to experiences in another task. Failure to recognize some items and the correct rejection of others may represent nothing more in both cases than the failure to contact traces in memory that are similar enough to justify a response of "old". Whether or not a particular item is truly "old" may be irrelevant to this process except in so far as as its prior presentation (or presentations) increases the probability that some present experience with the item will find a close match in memory. But, as the present results suggest, the important factor appears to be whether or not encoding is such that a close match is found rather than whether or not the event is nominally identical to some specified prior event. As is discussed next, the same appears to be true for categorical transfer.

<u>Categorical Transfer</u>

The major results of the categorical transfer phase are shown in Table 9.2. As can be seen (G-NG and C-F in Table 9.2) both the grammatical status and the specific similarity of the items were significant as main effects. As in previous experiments, specific similarity was the significantly larger of the two effects (z = 2.79) with a relative effect size, as indicated by the ratio of the estimated variance components, almost 23 times that of grammaticality. In fact, the 20.3% difference in responding to "close" and "far" items in the present experiment is the largest effect of specific similarity found for any encoding condition in the previous experiments.

Table 9.2
Categorical Transfer Phase

The mean percentage of items labelled "grammatical" as a function of the specific similarity and the grammatical status of the items. Column G-NG refers to the mean percentage difference of positive responses between grammatical and non-grammatical items. Column C-F refers to the mean percentage difference of positive responses between close and far items. The column labelled Actual Pc refers to the mean percentage of transfer stimuli correctly labelled. Pseudo Pc refers to the mean percentage of close grammatical and far non-grammatical items correctly labelled. Relative Frequency refers to the mean percentage of items receiving positive responses.

	Close		Far		G-NG	C-F	Actual	Pseudo	Rel.
	Gram	Non-G	Gram	Non-G			Pc	Pc	Freq.
Mnemonic Phrases	67.8	61.6	45.9	42.9	4.6	20.3	52.3	62.5	54.6

Depicted in Figure 9.3 are the mean percentage of items labelled "grammatical" as a function of the specific similarity of the items and the appropriateness of the associated mnemonic phrases. There was a large and significant main effect of the appropriateness manipulation. However, as in recognition, the main effects of both specific similarity and the appropriateness of the mnemonic associates were entirely a function of the significant interaction of these two variables. Fisher's LSD test indicates that subjects responded positively significantly more often to "close" appropriate items than to either "close" inappropriate items or either type of "far" items, with there being no significant differences among these latter three item types (LSD = 20.2%). Thus, the large effect of specific similarity for the categorical transfer phase in the present experiment was entirely a function of elevated responding to "close" appropriate items rather than increased sensitivity to the specific similarity of the items generally. As indicated by comparison with items associated with inappropriate contexts, there was no significant effect of the specific similarity of the items <u>unless</u>, as with the distribution of false-positive responses during recognition, the encoding contexts of the items were appropriate.

The effect of specific similarity in the present experiment is attributable to differences across encoding conditions in the subjects' perception of the similarity between training and transfer items <u>as encoded</u>. By varying the consistency of encoding between specific training items and specific transfer items within a single training condition, it was possible in the present experiment to eliminate the effect of specific similarity on one-half of the transfer items (the mean C-F for inappropriate items = 3.1%), while producing the largest effect of specific similarity obtained in any of the experiments on the other half of the items (the mean C-F for appropriate items = 34.4%), despite the fact that these two sets of

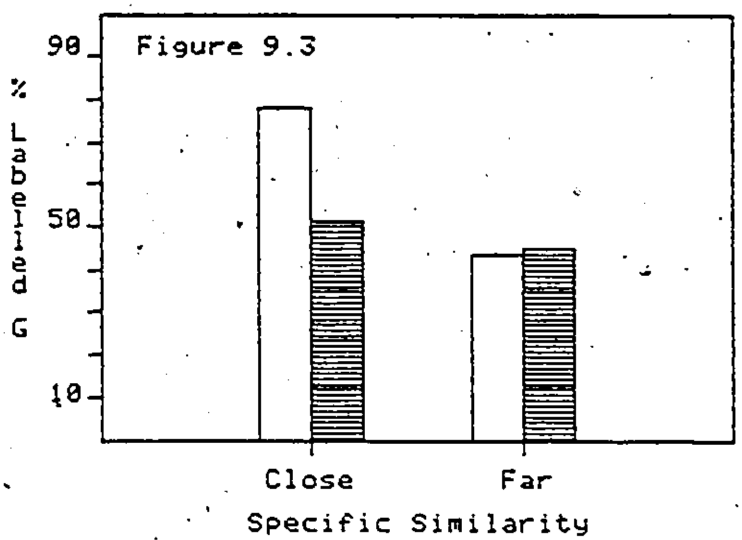

Figure 9.3: Transfer Phase: Mean percentage of items labelled "grammatical" as a function of the specific similarity of the items and the appropriateness of the associated mnemonics. (Appropriate mnemonics: open bars; inappropriate mnemonics: hatched bars)

items were matched for their formal similarity to the training items.

It is informative to view the results of the transfer phase of the experiment in terms of the accuracy of the subjects' categorical transfer responses. As is shown (Actual Pc = 52.3%) in Table 9.2, the subjects' accuracy in discriminating grammatical from non-grammatical items was about what would be expected on the basis of random responding. As usual, by confounding specific similarity and grammaticality (Pseudo Pc in Table 9.2), the subjects' accuracy climbs to a 62.5% correct. However, if the accuracy of the subjects' responses is calculated for those items for which encoded similarity rather than just formal similarity is confounded with grammaticality ("close" appropriate grammatical items and "far" inappropriate non-grammatical items), the accuracy of the subjects' transfer decisions rises to almost 70% correct. For this subset of transfer items, the subjects' performance is similar to that of subjects in Reber and Allen's (1978) observation condition. The present subjects demonstrated, as did Reber and Allen's (1978) subjects, a pronounced tendency to label all of these items as "grammatical" (61.6% of the items in the subset were labelled "grammatical"), and their accuracy was concentrated primarily in the correct acceptance of the grammatical items in this subset (81.1% correct) rather than in the correct rejection of the non-grammatical items in this subset (57.81% correct).

The present experiment demonstrates that it is possible to produce effects attributable to encoding differences within subjects that, in previous experiments, were demonstrated only across subjects in different encoding conditions. Thus, while it might be possible to argue that the encoding effects in the first three experiments were a consequence of qualitative changes in the processes subjects used to sort the transfer stimuli, as Reber and Allen (1978) argued for the differences between their observation and paired-associate conditions, it is not possible to argue the same for the encoding effects in the present experiment. Yet, in the present experiment, the effects were essentially the same as those associated with the encoding manipulations in the previous experiments. In this sense, for example, the subjects in the present experiment responded as Observation trained subjects for items associated with appropriate mnemonics, while responding as extreme Unique Mnemonic subjects for items associated with inappropriate mnemonics. In this case, however, the encoding effect was not mediated by variation in the breadth of transfer of specific training items, but rather by affecting the extent to which the encoding of certain transfer items would fall within the limited breadth of transfer of the training experiences resulting from the Unique Mnemonic training the subjects were given. The conclusion seems clear. The effect of encoding on categorical transfer (and recognition), whether manipulated globally as in Experiments 1 – 3 or locally as in the present experiment, occurs as a function of the extent to which the encoding of the training and transfer stimuli leads the subjects to perceive the transfer stimuli to be virtual twins of specific training stimuli.

The results of the present experiment question the abstraction interpretation of the results not only of those experiments in which the formal similarity between individual training and transfer items was left uncontrolled (e.g. Reber & Allen, 1978; and the majority of experiments discussed in Chapter 2), but also of those experiments that controlled at least some aspects of the formal similarity between training and transfer items (e.g., Posner & Keele, 1970; Omohundro, 1980; Homa <u>et al.</u>, 1981). Since no attempt was made to control the encoded similarity in these latter experiments, there is no assurance that the results they obtained were not entirely a function of a confounding between categorical status and the encoded similarity between transfer items and specific training items, even if it is assumed that formal similarity was completely controlled. As was demonstrated for the "close" transfer items in the present experiment, even if the items are matched for their formal, specific similarity to the training items, how this similarity is encoded by the subjects can have large, and significant, effects on categorical transfer. In light of these results, the demonstrations of transfer on the basis of implicit, abstract knowledge by reason of the experimenter's claim that the transfer items were "matched" for their similarity to the training stimuli should be viewed as interesting, but hardly conclusive. For those experiments in which the formal similarity between training and transfer items was not experimentally controlled (e.g., Hayes-Roth & Hayes-Roth, 1977; Reed, 1972, 1978), implicit abstraction of structure is one possibility, but so is transfer on the basis of the encoded similarity between transfer items and specific training experiences.

To this point in the discussion, specific similarity and the appropriateness manipulation have been discussed as if they were measures of different aspects of the training stimuli. From the persepective of the design of the experiment, such a distinction seems reasonable. Specific similarity is a measure of the formal similarity between specific training and specific transfer items, while the appropriateness variable is a measure of the informal or encoded similarity subjects should produce between transfer items and specific training items. However, as indicated by the inter-dependence of these two variables in producing effects on categorical transfer and recognition, it probably makes little sense in these experiments to consider the formal similarity between items independently of how the items are encoded by the subjects. The present results confirm the notion advanced for the results of previous experiments that what is important is not that the transfer items are or are not similar to the training stimuli, but that they are encoded as such by the subjects.

The remaining effects of interest during the transfer phase of the present experiment were generally in accord with those of previous experiments. As in Experiment 4, there was no significant main effect on categorical transfer responses of whether or not a particular item (and associated phrase) had been experienced on the prior recognition test. Also as in Experiment 4 (and Experiment 5), however, prior exposure of particular transfer stimuli interacted significantly with a number of the remaining variables of the transfer phase, although these effects generally were those of moderating rather than disordinally affecting the effects of these other variables.[9.1] Any effect of the prior recognition exposure on categorical transfer highlights the relative importance of specific, prior experience in the subjects' transfer judgements that is independent of any abstract knowledge of the grammar that the subjects may have induced from their training experiences. Such effects also are independent of any transfer from training to the categorical transfer task. Rather, they represent effects of transfer from non-training (i.e., recognition) experiences to particular transfer experiences, presumably arising from the heightened literal familiarity of the items, the associated mnemonics, or both. That subjects' transfer judgements were influenced, but not directly determined, by the literal familiarity of the transfer stimuli suggests that at least some "grammatical" responses resulted from subjects responding on the basis of the overall familiarity of the items rather than on the basis of analogy to specfic training experiences, or at least modifying their decision on the basis of familiarity information. Whatever the exact locus of the effect, it is clear that single, prior experiences can mediate transfer decisions, contrary to the general contention common to most researchers in the area.

To this point in the discussion of the subjects' categorical transfer responses, the results suggest that the transfer judgements were to a large degree a function of item specific knowledge. However, the analysis of the position of change from training to transfer items suggests that the subjects did learn something about the general properties of grammatical items. Shown in Table 9.3 are the mean percentages of transfer items labelled "grammatical" as a function of grammatical status and the position of change from the "closest" training item. As in previous experiments, there was a significant effect of the position of change of the items. Subjects were more likely to accept items with changes in internal letter positions than items with changes in either initial or terminal positions. However, this effect occurred primarily on non-grammatical items, replicating once again the position of change effects observed by Reber and his associates. Unlike Experiment 4 in which the position of change affected responding to grammatical and non-grammatical items in the same way, the effect of position of change on responding to grammatical items was significantly different than on responding to non-grammatical items. Subjects were more likely to accept grammatical items with changes in the initial letter position rather than, as in the case of non-grammatical items, being more likely to reject them. This result may be a function of the subjects having noticed that all the training items began with either an M or a V, yielding a simple rule that, for changes in this position, effectively discriminates between grammatical and non-grammatical items. If so, then it could be claimed that at least some aspects of the subjects' transfer judgements were mediated by abstract knowledge of the underlying grammar. As indicated by the non-differential effect of position of change across grammatical and non-grammatical items for the remaining two letter positions, and the major effects of specific similarity and mnemonic appropriateness discussed above, the use of this knowledge was limited to the few items in which these changes occurred and even there was probably not the major determinant of responding.

While it is tempting to attribute the difference in responding between

Table 9.3
Position of Change

Mean percentage of transfer items labelled "grammatical" as a function of grammatical status and position of change from the "closest" training item.

Item Status	Letter Position of Change Initial	Middle	Terminal
Grammatical	62.5	57.9	54.7
Non-grammatical	33.3	57.1	51.5

grammatical and non-grammatical items with changes in the initial letter position to abstract grammatical knowledge, there is a problem. Specifically, the effect emerged in an encoding condition in which subjects would be expected to be least likely to notice that all training items began with either an M or a V, but failed to emerge in Experiment 4 in which, because the items were presented without distinguishing mnemonics, the consistency among first letters of the training items would be expected to be more obvious. This anomaly suggests that the differential effect of position of change occurred as a function of the mnemonic associates that, for changes in the initial letter position, affect the subjects' encoding of grammatical items differently than that of non-grammatical items. Although the precise locus of this effect is not clear, it probably involves the salience of the initial letter and, for the associated mnemonics, the initial word in subjects' assessments of the similarity between training and transfer experiences. It is possible that the uncommon first letter of the non-grammatical items with changes in this position, further emphasized by the associated word change in the mnemonics, resulted in too marked a difference from any or all training items for subjects to accept these items as "grammatical". Since this would not be as true for grammatical items with changes in this letter position, responding on the basis of such differences in similarity effectively would sort the items with respect to their grammaticality. Thus, while the position of change results for the initial letter position are consistent with the subjects detecting violations of the underlying grammar, the results do not point unequivocally to abstract knowledge as the source of his ability.

Interview

The results of the interview phase of the experiment are presented in Table 9.4. There was only one significant effect. As in Experiments 4 and 5, the subjects estimated that they used similarity to the training items significantly more often

Table 9.4
Interview Phase

Subjects' mean estimated percentage of transfer decisions for which each of the eight strategies were used.

		Rules	Similarity	Mean
Specific	Accept	29.0	53.6	41.3
	Reject	31.4	41.2	36.3
General	Accept	32.3	42.2	38.8
	Reject	26.3	42.9	34.6
Mean		29.7	45.7	37.7

than rules when making their categorical transfer judgements. Although there was a suggestion that, as in the previous experiments, the subjects estimated that they used primarily specific similarity to accept items as "grammatical" while using general similarity primarily to reject them, this effect was only marginally significant ($p = .05$). In general, though, the subjects' assessments of the bases of their categorical transfer decisions are in accord with the objective analyses discussed above. As in the previous experiments, the subjects apparently were aware of the basis of their judgements which, by their own reckoning, has less to do with abstract rules than the similarity between training and transfer items.

Chapter 10

GENERAL DISCUSSION

The major portion of the work in this thesis was intended to re-examine Reber and Allen's (1978) evidence for rapid learning of abstract structure. More generally, the work was intended to challenge the widely held notion that much of our knowledge consists of implicit, abstract structures such as prototypes, grammars, scripts and frames. The research reported in this thesis provides no support for the notion of implicit abstraction of structure. Instead, the results from the series of six experiments are consistent in suggesting that the appearance that subjects have learned and are using implicit abstractions is illusory, arising from the memorial distribution of individual experiences that subjects use to accomplish the tasks set for them. In the first section of this chapter, ~~these~~ research results are summarized and their bearing on an instance model of structural learning in terms of encoding episodes is discussed. Subsequent sections discuss the extension of these ideas to the areas of education and artificial intelligence.

Summary and Discussion of the Experiments

The results of the six experiments present a complex array of effects. They all, however, may be accounted for in terms of their expected effects on the subjects' perception of the encoded similarity between events. This conclusion is most clearly evident for the effect of the specific similarity of the items used in the tests of transfer of structural knowledge, and is discussed first.

Specific Similarity and Categorical Status

The core of the basic paradigm used in each of the experiments was the orthogonal manipulation of the categorical status of the transfer items and their specific similarity to the training items. The intent of this manipulation was two-fold. First, it was to demonstrate that, in the absence of a confounding with categorical status, subjects' transfer decisions would occur more as a function of the similarity between training and transfer items than as a function of the categorical status of the transfer stimuli. Second, the specificity of the similarity manipulation was to demonstrate that the basis of this effect was the similarity to specific training items rather than transfer on the basis of some general property of the training items. In all six experiments, this result was confirmed. In each experiment, the specific similarity of the transfer items proved to be more important in determining subjects' judgements of the categorical status of the transfer items than did the categorical status of the test items themselves. In most cases, the effect of the categorical status of the items on judgements of categorical status was very much smaller than that of specific similarity.

The orthogonal manipulation of categorical status and specific similarity in these experiments demonstrated that subjects' judgements of the categorical status of test items in these tasks were determined in the main by the similarity between transfer items and specific training items. However, the categorical status of the items was not ineffective. In most experiments it exerted small, but significant effects on categorical transfer judgements, and, in every experiment, was found to affect subjects' false-positive responses on tests of item recognition. Thus, it could be argued that, while categorical status did not play a major role, the subjects had abstracted a partial representation of the underlying grammatical rules determining the categorical status of the items that they used in conjunction with specific similarity information for at least some of their categorical transfer and recognition judgements. To make such an argument, however, requires that it be assumed that the specific similarity manipulation captured <u>all</u> the similarity between test items and particular transfer items, and that, consequently, the categorical status variable represents a manipulation of the categorical status of the items unconfounded with with any similarity between training and transfer items. Such an assumption is unlikely to be true.

In addition, there is the evidence from Experiment 4 which compared the categorical transfer judgements of one group of subjects with the judgements from a second group of subjects of the similarity between transfer and training items. Despite the fact that subjects in the similarity transfer group were not informed about the categorical distinction, they discriminated the categorical status of the items on the basis of similarity judgements virtually equivalently to subjects asked to discriminate the items on the basis of categorical status. In fact, with the exception of the similarity transfer subjects being more sensitive to the specific similarity of the items than were the categorical transfer subjects, the patterns of responses of the two groups were indistinguishable. In sum, on the basis of the relative effects of the two variables of categorical status and specific similarity, there is strong evidence that the subjects' transfer judgements were based on the similarity between test items and specific training items.

The Effect of Encoding

One of the major concerns of the experiments was to investigate the effects on judgements of categorical status of how the subjects were asked to encode the training items. According to Reber and Allen (1978), transfer on the basis of analogy to specific training experiences is a result peculiar to training tasks requiring that the subjects individuate the items from one another, preventing the implicit abstraction process from abstracting the regularities among the items. In their view, if subjects are exposed to the training items under more natural circumstances, implicit abstraction of the underlying structure inherent in the items rather than the specific similarity between training and transfer experiences provides the primary basis for subjects' transfer judgements. The results of Reber and Allen's (1978) experiment were consistent with this notion.

In contrast to Reber and Allen's implicit abstraction hypothesis, the breadth of transfer notion provides an account of encoding effects in terms of quantitative variation in only a single domain. According to this notion, the superior transfer performance following observation training is due to a greater sensitivity to the specific similarity between training and transfer stimuli, resulting from a less restricted breadth of transfer of the memories of the individual training experiences. On the assumption that specific similarity is confounded with the categorical status of the transfer items, increased sensitivity to the similarity between training and transfer stimuli, up to some point, would result in superior transfer performance.

The predictions arising from the two approaches for the situation if specific similarity and categorical status are set orthogonal to one another are clear and differential. According to the implicit abstraction hypothesis, observation training should lead to greater sensitivity to the categorical status of the items while paired-associate training, or any training condition in which subjects are required to individuate the training items from one another to meet the demands of the training task, should lead to increased sensitivity to the similarity between training and transfer items. In contrast, according to the breadth of transfer notion, observation training should lead to greater sensitivity to the similarity between training and transfer items than should training procedures requiring greater item individuation, and sensitivity to the categorical status of the transfer items should not vary as a function of training procedure.

The results of Experiment 1 supported the breadth of transfer explanation. Subjects' sensitivity to the specific similarity of the transfer stimuli decreased as a function of the expected individuation requirements of the training tasks, and there was no significant variation in subjects' sensitivity to the categorical status of the items as a function of training procedure. Similarly, in Experiment 2 in which the training conditions were constructed to differ in breadth of transfer only, subjects' sensitivity to the specific similarity of the transfer items varied significantly in the expected direction as a function of training condition with there being no significant variation in subjects' sensitivity to categorical status. In Experiment 3, however, subjects' sensitivity to the categorical status of the items was found to be significantly related to how the subjects were asked to learn the training items, and training procedure was not significantly related to the subjects' sensitivity to the specific similarity of the transfer items. In this case, however, the results with respect to categorical status were the opposite of those found in Reber and Allen's (1978) experiment, and contrary to the predictions of the implicit abstraction hypothesis. For the conditions of this experiment, observation trained subjects demonstrated no sensitivity to the categorical status of the transfer items while subjects required to individuate the items during training evidenced a large degree of sensitivity to the categorical status of the items. Although the results also were not consistent with the predictions of the breadth of transfer notion as outlined above, the inconsistency was suggested to be a consequence of the subjects in the latter group transferring item-specific information on the basis of higher-order similarity other than that captured by the specific similarity manipulation. Since the similarity between training and transfer items assessed at this level would not be orthogonal to the categorical status of the transfer items, the effect rather than the intent of this training procedure may have been to re-introduce a confounding between categorical status and the similarity between training and transfer items that allowed subjects trained in this way to to sort the transfer items with respect to categorical status on the basis of the similarity between training and transfer stimuli.

In the breadth of transfer terms used in this thesis, tasks such as recall and recognition that nominally assess subjects' learning of the individual items, and tasks such as categorical transfer that assess subjects' ability to transfer this knowledge to other items are all considered to be measures of the breadth of transfer of prior learning experiences to some present task or situation. The important point is that performance in these different tasks may be affected differently by breadth of transfer. Recognition and recall performances provide measures of the extent to which transfer from prior experiences is limited to each individual training event on some clearly identified test of memory for specific events. Categorical transfer, on the other hand, provides a measure of the extent to which the breadth of transfer is broad enough to encompass similar, but non-identical events on a test in which the subjects' task is to generalize beyond the original learning experiences. Thus, over some fairly broad range of breadth of transfer, subjects' performance on tests of item learning and tests of categorical transfer should be inversely related since the limited breadth of transfer underlying good recognition and recall performance will probably be too limited for optimal categorical transfer performance. According to most theorists, however, as far as instance-theory is concerned, categorical transfer should be a monotonically increasing function of how well the subjects have learned the training items. In fact, several researchers (e.g., Reed, 1978; Homa et al., 1981; Reber & Allen, 1978) have used this as a key component of a their arguments against individual instances as an explanation of categorical transfer since, if anything, the relationship appears to be just the opposite (Reber & Allen, 1978; Homa et al., 1981; Omohundro, 1981).

Experiments 1 and 2 provided support for the breadth of transfer interpretation of the relationship between performance on the different tasks. Variations in how the subjects were asked to learn the training stimuli, designed to induce different degrees of breadth of transfer, produced inverse correlations between performance on tests of subjects' memory for specific items and their performance on categorical transfer (as assessed by Pseudo-Pc). Accuracy on some task, is not a preferred measure of breadth of transfer, since breadth of transfer is only indirectly reflected through normative standards of "correctness". More direct measures are provided by inter-list transfer during training, the rate and distribution of false-positive responses during recognition, and the rate and distribution of positive responses during categorical transfer. All of these measures, unlike the accuracy measures of performance on these tasks, should be positively related to breadth of transfer and, hence, positively related to each other. Cast in these terms, the results of Experiments 1 and 2 in particular, and the results across all six experiments more generally, support the breadth of transfer interpretation of the relationship between the tests of item "learning" and categorical transfer.

The importance of considering the nature of the conditions of the transfer task when assessing the relationship between subjects' learning of the training items and subsequent transfer is highlighted by the results of Experiment 6. This experiment demonstrated the results anticipated on the basis of an assumed positive relationship between item learning and categorical transfer. The subjects demonstrated both the best learning of the training items, as indicated by their superior training and recognition performances, and the best categorical transfer, as indicated by their sensitivity to the specific similarity of the transfer items, of subjects in any of the experiments. Yet, the results are in accord with the breadth of transfer notion outlined above. The high rate of categorical transfer achieved was not a consequence of superior memory for the individual training items. In fact, as indicated by the lack of inter-list transfer during training, the low rate of false-positive responses during recognition, and, most important, the lack of sensitivity to the specific similarity of one-half of the transfer items in this experiment, the subjects showed virtually no tendency to generalize across items strictly on the basis of their learning the individual training items. A major reason for the high-rate of categorical transfer relative to that obtained in the other experiments was the fact that one-quarter of the transfer items were associated with mnemonic phrases that biased the encoding of these items in the direction of their "closest" training items, while the encoding of the remaining items was biased in the direction of training items which were poor analogies of the transfer items. Thus, by increasing the encoded similarity between training experiences and a sub-set of the transfer items, subjects in this experiment were able to demonstrate a high rate of transfer to this sub-set of items despite their apparently limited breadth of transfer considered more generally. But, as in the previous experiments, transfer was a function of the encoded similarity between transfer experiences and specific training experiences. Thus, transfer, whether limited to judgements of categorical status or broadened to include inter-item transfer during training and recognition judgements, and whether considered as a function of global or local encoding effects, appears to be a function of the encoded similarity between specific events.

A Spatial Metaphor of Encoding Effects

Inter-item transfer may be conceived to be a function of two, basic factors: (1) the manner in which, as a function of encoding, the original experiences with some item or items are distributed in memory, and (2) the extent to which the encoding and subsequent representation of transfer events (items, or retrieval cues more generally) affects the likehood that the memory of some prior experience will be contacted. The global encoding operations used during training in these experiments appear to exert their effects primarily on the first factor. The effect of unique mnemonic training, for example, relative to that of arbitrary label training may be represented as a reduction in the overlap in encoding across different training items, resulting from both an increase in the <u>mean</u> difference in the encoding of different training items and a reduction in the <u>variance</u> of encoding across repeated experiences with the same item. Since transfer is conceived as a reflection of the likelihood of memorial contact, transfer to a given encoding of a particular transfer item following unique mnemonic training should occur with a lower probability than that following arbitrary label training because of the relatively more sparse coverage of the range of possible transfer experiences in the former condition. Thus, global encoding conditions such as observation training and no label training result in superior levels of transfer because their experiences with the training items more completely covers the range of possible transfer experiences.

Local encoding effects during transfer appear to affect the second factor. Thus, for a given memorial distribution of training experiences, directing the subjects' encoding of transfer items toward the encoding of particular training items increases the probability of contact and, hence, transfer. In the experiments reported in this thesis, there appeared to be several variables that affected transfer by this second route, with specific similarity in all of the experiments and the appropriateness of the mnemonic associates in Experiment 6 being the most notable examples. However, the variables of alteration type (i.e., whether the transfer items were produced by adding or substituting letters in the training items) and pass (i.e., first or second sort of the transfer items) also may have produced their occasional effects by affecting the probability that the transfer items would be encoded to match the remembered encoding of the training items.

The Transfer of How to Encode

Although it is convenient to represent the effects of the global encoding operations solely in terms of their effects on the distribution of training experiences in memory, the effect of how the subjects are requested to encode the training items is unlikely to be limited to this one factor. In fact, this variable also would be expected to influence the second factor discussed above, since subjects are unlikely to limit the manner in which they encode the training items to the training task itself. Rather, it seems probable that subjects transfer more than just their memory for particular training experiences to the transfer task. They also probably transfer the particular way of encoding or representing the items that they were trained to use in the original training task. Thus, for example, observation-trained subjects may give the transfer items little more than a cursory glance before making their transfer decisions, while mnemonically-trained subjects may attempt to generate mnemonics for each item and base their judgements about the items on the relative success or failure of this process. A number of unanticipated results of the experiments are consistent with this notion. Thus, for example, although no attempt was made to quantify the relationship, there were marked differences in the time taken by subjects in the different conditions to complete the transfer task. Subjects given observation training, as in Reber and Allen's (1978) experiment, found the transfer task relatively effortless, typically completing the transfer phase of the experiment in less than 10 minutes. Subjects in the mnemonics conditions apparently agonized over their transfer decisions, sometimes volunteering that none of the items looked like the training items, and often took more than twice as long to complete the task. At least part of this difference is probably due to differences in the perceived similarity between training and transfer items arising because of differences in the memorial distribution of the training experiences.

However, the similarity between the effort the subjects in the different conditions devoted to the transfer task and that requested of them during training, suggests that at least some part of the differences during transfer were due to the subjects in the different training conditions attempting to apply not only what they learned about the training items but also how they learned the training items to the items in the transfer task.

At least one consequence of subjects transfering not only what they learned about each of the individual training items but also how they learned it to the transfer task is that the encoded similarity between training and transfer experiences should be increased beyond that expected on the basis of the subjects' memorial distribution of the training experiences. That is, it is unlikely that the subjects in different training conditions would encode the transfer items in exactly the same way. Transferring the manner in which the training items are encoded to the items in the transfer task would have the effect of reducing the magnitude of encoding differences on the measures of categorical transfer since it would increase generally the encoded similarity between training and transfer items. The reduction in performance differences arising from the transfer of how to encode may account for Reber and Allen's (1978) finding that the effects of the different training conditions were generally much larger for subjects' experience of the transfer task than for their performance on the transfer task. Be that as it may, the possibility that the manner in which subjects are requested to encode the training items may affect the manner in which the transfer items are encoded played a large role in the interpretation of the results of Experiment 3.

In Experiment 3, one group of subjects was trained with mnemonic associates of the training items that emphasized the grammatical commonalities across the items. There were two major effects of this training procedure. First, it appeared to reduce the subjects' ability to discriminate adequately among the items during training, resulting in the poorest training and recognition performance of any of the memory-training conditions in the other experiments. Second, relative to a group of observation-trained subjects in the same experiment, Grammatical Mnemonic training enhanced the subjects' ability to discriminate among the transfer items on the basis of the categorical status of the items - the only experiment in the series in which such an effect was found. This result is explicable in terms of the meta-memory perspective outlined above if it is assumed that the grammatical mnemonic training affected how the subjects encoded the transfer items relative to that of the observation-trained subjects in the same experiment and subjects in the training conditions in the other experiments. The effect of the grammatical mnemonic training on subjects' ability to discriminate among the training items suggests that the grammatical mnemonics induce subjects to encode the training items in a manner that <u>reduces</u> the formal differences between the items. As discussed in Chapter 6, at least one way of doing this is to represent each item in terms of higher-order features, groups of letters or patterns, rather than in terms of the individual letters making up each item. Since the items are more alike one another at this higher-order level, representing each item in this way would increase the encoded similarity between one training item and another. If the subjects then attempted to encode the transfer items in much the same way, the encoded similarity between the transfer and training items also would be increased. But since these higher-order features are primarily a function of the <u>grammaticality</u> of the items, this increase in encoded similarity would extend almost exclusively to the grammatical transfer items, effectively discriminating the categorical status of the items.

<u>Categorical Transfer and Recognition</u>

Some researchers have used subjects' recognition judgements to assess the learning of categorical structure. The general conclusion from the results of these experiments has been that subjects abstract the categorical structure underlying the training items they are asked to learn and use this information when making their recognition judgements. Thus, the general view within the classification learning literature is that the abstraction and use of categorical structure is ubiquitous, applying to subjects' recognition judgements and to judgements of categorical status. The inclusion of a recognition task in each of the experiments in this thesis was to make a similar point, but in this case it was to demonstrate that while recognition and categorical transfer are undoubtedly related, the relation is probably through subjects' dependence upon the memory for individual encoding experiences in each task rather than some abstract representation of the categorical structure of the items.

In each of the six experiments, subjects' false-positive responses during recognition were distributed in much the same way as their positive responses during categorical transfer. This was most clearly demonstrated in Experiment 6 in which in addition to the usual effects of categorical status and specific similarity on subjects' responses in both tasks, subjects' responses in both recognition and categorical transfer showed similar effects of the local encoding manipulation. In both cases, positive responses varied as a function of the encoded similarity between transfer items and specific training stimuli. Thus, while it seems that subjects' responses in the two tasks depend upon similar processes and information, this dependence probably has little to do with the use of some abstraction of categorical structure.

Demonstrating similar effects of an encoding specificity manipulation in both categorical transfer and recognition is important for several reasons. First, the similarity of effects provides a common ground for the interpretation of encoding effects in both tasks. Since recognition tasks have provided one main source for the demonstrations of the effects of encoding specificity (see Craik, 1979, 1981, for reviews), the demonstration of similar effects of encoding specificity on categorical transfer suggests that much of this work on recognition may be profitably extended to the investigation of the learning of categorical structure. Second, the common ground suggests reciprocal benefits for the investigation of recognition processes. In particular, viewing recognition as a categorical transfer task may provide promising leads into the nature of the myriad of variables that have been found to influence recognition performance, especially through their effects on subjects' false-positive responses.

This is not to say that categorical transfer and recognition are interchangable tasks. Across the six experiments in this thesis, subjects' categorical transfer responses and their false-positive responses during recognition differed in at least two related ways. First, the effects of the independent variables common to both tasks, particularly that of specific similarity, were generally larger for categorical transfer than for recognition. Second, relative to the effect of specific similarity in both tasks, the effect of the categorical status of the items was generally larger for recognition false-positive responses than for positive responses during categorical transfer. The difference in the magnitude of the effects across the two tasks probably represents a difference in the subjects' criterion for acceptance between the two tasks. That is, the subjects probably demanded a higher degree of match in terms of the encoded similarity between transfer items and specific training experiences if making recognition judgements for which the task was to indicate only previously experienced items than if making categorical transfer judgements for which the task was to generalize beyond the training experiences. Such a result is important since it suggests that the categorical transfer in these experiments was more than the literal "misrecognition" of the transfer stimuli, but iincluded a component of generalizing beyond some perceived identity of training and transfer experiences.

The relatively larger effect of grammaticality on recognition false-positive responses than positive responses during categorical transfer is consistent with the interpretation of subjects imposing a more restrictive acceptance criterion for recognition judgements than categorical transfer judgements. Since the effect of specific similarity was larger than grammaticality in the transfer phase of each of the experiments, the relatively greater effect of grammaticality during recognition was primarily a function of a reduction in the magnitude of the specific similarity effect between categorical transfer and recognition. This reduction in the magnitude of the effect of specific similarity across tasks is similar to the effect of the encoding manipulations discussed earlier, and is consistent with the breadth of transfer interpretation of encoding effects. Similar variation in breadth of transfer as a function of task variation also has been found in a series of concept learning experiments by Wittlesea (Note 3). In the present case of recognition judgements, the reduction is probably due to the subjects intentionally restricting the breadth of transfer around their memory for specific training experiences in order to avoid producing false-positive responses to highly similar transfer items. However, since the encoding differences emanating from the different training procedures affected the production of false-positive responses and not hits during recognition, breadth of transfer is not entirely a function of a subject-exercised decision criterion, although variations in the breadth of transfer appear to produce similar results whether they are a function of subjects varying their criterion of acceptance across tasks or a function of the original encoding of the training items.

Position of Change

One of the concerns of the final three experiments was the investigation of

Reber and Allen's (1978) and Reber, et al.'s (1980) finding that the subjects' ability to detect correctly the non-grammatical transfer items varied as a function of the letter position in which the violation of the generative grammar occurred. Reber and his co-workers found that non-grammatical items with grammatical violations in either the middle of terminal letter positions were more readily detected by the subjects than were items with violations in internal letter positions. The effect of position of change on subjects' categorical transfer decisions is intriguing since it parallels results obtained in investigations of subjects' detections of spelling errors with real words (e.g., Haber & Schindler, 1981).

Although there are a number of theories about the source of the effect of letter position of violation and other "word-superiority" effects with real words (see Baron, 1978, for review), many of them implicate subjects' use of abstract orthographic knowledge coupled with the relative perceptual salience of the different letter positions within the words. The general idea here is that initial and terminal letter positions are more salient (or are processed earlier) than internal letter positions so that the abstract knowledge about the orthography of words may more readily be brought to bear on violations in these letter positions.

Reber and Allen's (1978) representation of the similar effects in a categorical transfer task as detections of violations of the underlying grammar suggests that they interpret these effects to be more than superficially similar to those associated with real words. Consistent with the thesis of implicit abstraction, the implication from Reber and Allen's (1978) presentation of the effects of position of violation is that they are a function of subjects' use of abstract knowledge about legal constructions. The analogy is indeed striking. Not only are the effects similar across the letter-strings used in Reber's experiments and the real words used in investigations of proof-reading, the artificial grammar itself is perhaps better viewed as an artificial orthography than a grammar since it produces ordered sequences of letters rather than sentances. Other parallels, such as the predominantly implicit nature of the abstract knowledge in both domains, further strengthen the analogy. However, to the extent that parallels with other domains may be held to buttress Reber and his co-workers' conclusions about implicit abstraction in the learning of artificial grammars, demonstrations that the effects within the learning of artificial grammars that give rise to the parallels are <u>not</u> due to implicit abstraction question the strength of the evidence for this claim in other domains. Thus, finding that Reber and Allen's (1978) effect of position of violation is unlikely to be due implicit abstraction has implications for domains beyond that of the learning of artificial grammars.

The results of Experiment 4 question an implicit abstraction interpretation of the effect of letter position of change on categorical transfer judgements. Beyond the fact that none of the other major effects of the experiment supported the implicit abstraction hypothesis, the effect of position of change in the experiment appeared to be a function of its relation to the similarity between transfer and specific training items. Unlike Reber and his co-workers' investigation of position of change, which applied to non-grammatical items only, Experiment 4 orthogonally manipulated position of change across the categorical status of the transfer items. If the effect of position of change is a function of subjects' detections of grammatical violations then the effect should be limited to non-grammatical items since, by definition, only these items contain grammatical violations. In contrast, if the effect is a function of the similarity between training and transfer items, then grammatical and non-grammatical items alike should evidence the effect of position of change. The results were consistent with the similarity interpretation. Subjects were more likely to label as "non-grammatical", transfer items with changes from their "closest" training item in the initial or terminal letters than items with changes in internal letter positions, and this effect was virtually independent of the categorical status of the items. Moreover, the effect was virtually unchanged for another group of subjects' who were asked to ~~sort the~~ transfer items on the basis of their similarity to the training items, and who had not even been informed about the existence of the underlying grammatical rules. Thus, the effect of position of change appears to be a function of the similarity between training and transfer items rather than implicit abstraction.

Experiments 5 and 6 confirmed these results for the middle and terminal letter positions of the items, but produced relative differences across grammatical and non-grammatical items for changes in the initial letter position. Although these effects may be interpreted within an inter-item similarity framework, it was suggested that they also may have occurred as a function of the subjects' explicit knowledge about the grammar. This possibility is discussed below in the context of a general discussion about subjects' explicit reflections on the bases of their categorical transfer decisions.

Introspective Reports

An interesting aspect of Reber and Allen's (1978) experiment was the relationship between the objective analyses of the bases of their subjects' categorical transfer judgements and the subjects' reports of how they accomplished the task. Since the majority of the subjects' transfer decisions were unjustified, the data from the introspective reports provide support for Reber and Allen's (1978) claim of implicitness, but provide less support for their parallel claim of abstraction. However, the introspective reports collected by Reber and Allen (1978) and the interview data obtained from the subjects in my experiments suggest that if subjects are given similarity to training instances as a possible account of categorical responses, subjects will often attribute at least some of their decisions to this process. That the subjects do so does not mean, necessarily, that the subjects actually used analogy to prior experiences when making their judgements. However, it does suggest that the failure in other experiments to recruit such attributions from their subjects is poor evidence that the subjects were not using analogy. Being aware that such an attribution is an acceptable response may be an important component of the subjects' perception of the demands of the attribution task, and, consequently, the attributions they are likely to produce.

In Experiments 4, 5, and 6, following the categorical transfer task, subjects were presented with a choice of eight strategies that they were informed they might have used when producing their categorical transfer decisions. They were asked to estimate the proportion of their transfer decisions for which they used each strategy. In an attempt to prevent the subjects from viewing the task as a forced-choice, they were informed that they may have used any or all of the strategies, even for the same item, and may have used strategies not included in the interview. Thus, the subjects had a considerable degree of freedom in the estimates they chose to associate with each strategy. However, in all three experiments, the subjects estimated that they used similarity to the training items rather than rules as the primary basis of their transfer decisions, and, in two of the experiments, attributed the <u>acceptance</u> of transfer items primarily to the similarity between any given transfer item and some <u>specific</u> training item. Thus, the subjects in these experiments affirmed the conclusions advanced for their categorical transfer performance on the basis of the objective analysis of their behaviour.

In Reber and Allen's (1978) experiment, analogy to the memory of particular training instances was not the predominant attribution produced. Rather, the majority of justified decisions involved statements about relatively simple aspects of the items. These statements generally referred to the rejection of transfer items, typically on the basis of some suspected violation of the initial or terminal letters of the items. For the most part, these statements were correct with respect to the rules of the grammar. Thus, to a limited extent, the subjects could actually sort the items correctly (or at least correctly reject some of the non-grammatical items) on the basis of the relatively simple rules contained in their explicit justifications. Note the consistency of this result with the position of change effect for this experiment discussed above. Although Reber and Allen (1978) do not emphasize the point, this consistency suggests that if the effect of position of change is to be attributed to abstract knowledege, the knowledge is probably best represented as explicit rather than implicit abstraction. More generally, the finding that the majority of the subjects' explicit justifications were correct with respect to the underlying grammar, and the finding that the majority of of transfer decisions were left unjustified, suggest that if a mixture of processes did occur, as Reber and Allen (1978) suggested, such a mixture could be characterized as a mixture of only the two processes of analogy to prior experiences and explicit abstraction, rather than the three-way mixture with implicit abstraction that Reber and Allen (1978) found congenial to their view.

The explicit, abstract knowledge that Reber and Allen's (1978) subjects' reported was limited to relatively simple aspects of the letter-strings. Even at this, however, the amount of explicit knowledege that Reber and Allen's subjects reported exceeded that demonstrated in the transfer performance of my subjects. At most, the results of the transfer phase of the last two of my experiments could be used to support the contention that some of the subjects had learned a simple rule about permissible first letters of the items – a reduced version of the type of explicit knowledge expressed by Reber and Allen's (1978) subjects. Although there are numerous differences between Reber and Allen's (1978) experiment and those reported in this thesis, the fact that Reber and Allen's subjects gained their explicit, abstract knowledge in an experiment in which the transfer task included the additional requirement that the subjects attempt to justify each decision suggests that this requirement may be important in the development of explicit knowledge from what, immediately following training, probably consisted of little more than than the memory for individual training items. The relationship between the post-training task and the form of knowledge likely to arise as a consequence of the task is discussed next.

Relationship To Other Tasks and Knowledge

The central issue addressed in this thesis has been the question of what information people apprehend and use in their identification of instances as members of categories. On the basis of experiments reported in this thesis, it has been argued that much of the knowledge that subjects use in making their categorical judgements is at the level of memory for individual experiences. As mentioned earlier, however, retaining relatively complete memories of individual events endows the learner with a great flexibility in the uses to which this knowledge may be put. That is, the item-specific information that the subjects apparently gained in the experiments may be used for far more than judgements about the categorical status or prior occurrence of some item or set of items. It is this flexibility of memory for individual instances that makes individual instances such an effective way of representing our knowledge about an often unpredictable world.

In this section, the issue of the flexibility of memory for individual experiences is approached in three, related ways. First, I expand on an argument made earlier that abstraction only appears more economical than memory for individual events because of an over-emphasis on the learning and representation of particular abstractions considered one at a time. When we recognize how many descriptions (abstract labels) may be applied to any given event and the relation between events, retaining each particular event that we encounter can be made to appear more economical. Second, I discuss the related issue of the variety of tasks for which instance knowledge may be used, and how these tasks may produce knowledge in the form of explicit abstractions. Both of these issues are related to the third issue which concerns the mix between explicit, abstract knowledge and implicit knowledge in the form of memory for individual events that may be important for an understanding of expert performance. The discussion of this third issue raises several considerations for the field of education, and concludes with a discussion of some recent research within the artificial grammar paradigm that has been suggested to have practical application in the training of sonar operators.

The Economy of Individual Instances

From the perspective of any single task, abstracting and retaining only the essential invariances across events is more economical than retaining a separate memory of each relevant event encountered, particularly if the set of relevant events is large and the process of abstraction relatively simple and effortless. Thus, on the assumption that processing and memory should favor the more economical of processes when a choice is available, abstraction would appear to be the best process for most tasks. However, there are at least two reasons why abstraction is probably far less ubiquitous than suggested by this argument. First, with the exception of the relatively simple abstractions used in investigations of explicit rule induction, few natural domains allow "simple" abstractions, at least for explicit learning. And the claim that what may be difficult for explicit processing may be relatively simple for a process of implicit abstract, while possibly true, is of little consequence in the absence of any explanation as to why this may be the case. As far as I am aware, no such explanation other than the claim of "automaticity" has been offerred. Second, and perhaps more important, most of our knowledge about our world is acquired incidentally; we are often not aware of the tasks to which the knowledge we have gained from particular experiences will be used. Often, we are unaware that we have gained the knowledge relevant for some task until confronted with the task itself. For this knowledge to be in the form of abstractions, implicit or otherwise, would require that the myriad of possible abstractions be obtained simultaneously during our original experiences with the events in question – a highly unlikely possibility that Brooks (1978) used to great advantage in his investigation of the implicit learning of artificial grammars. To my knowledge, no theorist has explicitly suggested such a possibility. But, the claim is implicit in many models of learning, particularly those that posit an automatic and abstract representation of incidentally acquired frequency information. Since the representation of frequency information represents a currently contested area in cognitive psychology and since memory for individual experiences has been suggested as an alternative to implicit abstraction in accounts of people's knowledge for event frequency in these tasks, I will use it as a demonstration of the relative efficiency of individual instances.

The subjects in my and Reber's experiments were asked to do little more with their learning of the training items than discriminate training items from transfer items on a test of item recognition, and discriminate grammatical items from non-grammatical items on a test of categorical transfer. Considered in isolation from other tasks the subjects could have been given, implicit abstraction does not appear to be an unlikely possibility. However, we can imagine requiring that the subjects engage in numerous other tasks following their original training experience.

For example, the subjects could have been asked to estimate the relative frequency of occurrence of particular training experiences. As has been demonstrated in other investigations, subjects have little difficulty providing accurate estimates even if, as in the present experiments, knowledge about event frequency was incidental to the demands of the original training task (e.g., Hintzman & Stern, 1978; Zacks, Hasher, & Sanft, 1982). Again, considered by itself, implicitly abstracted representations of event frequency in the form of automatic frequency accumulators, propositions, associative "strengths", and the like do not appear as implausible, and even appear as economical since a single representation of each item along with its accumulated frequency would require far less storage than that associated with storing every experience with each item. However, even if we limit the questions to those about frequency, the subjects still may be asked numerous questions of an "abstract" nature about their training experiences. For example, they could be asked to estimate the relative frequency of items that began with the letter "M", or items whose second letter was "V", or items whose first two letters were "MV", and so on. In fact, we could ask them to report on each of these frequencies, and it seems likely that they would perform relatively well on each of them. To argue that their answers to each of these virtually limitless number of questions arises from the independent abstraction of each type of event frequency during the original training task when the subjects were not even aware that knowledge about the frequency of <u>any</u> of the different event types would be required seems highly implausible. Even if it is assumed, despite its implausibility, that this is what subjects do, such a representation of frequency information is hardly economical since the number of abstractions necessary to answer all of the potential questions will have vastly exceeded the number of original training experiences. In contrast, as is discussed next, <u>indirect</u> retention of frequency information in the form of the memory for each of the individual experiences now appears to be a far more "economical" approach to the multitude of tasks to which the subjects' knowledge could be put.

Abstraction At Test

In Reber and Allen's (1978) experiment, subjects were requested to introspect on the basis of each categorical transfer decision that they made. As noted earlier, the subjects reported several simple rules that were consistent with the grammatical rules underlying the items. The question addressed here is that of the source of this knowledge. Reber and Allen (1978) implied that the subjects obtained the simple rules that they reported during the training phase of the experiment. However, it is possible that these rules were computed from the the subjects' memory for the individual training experiences only when the subjects were confronted with the surprise transfer task. That is, when the subjects discovered, following training, that they would be required to sort the transfer items on the basis of grammatical rules and that they would have to provide justification for their transfer decisions, they may have recalled a representative sample of training items and computed the simple rules that they may have used to determine the categorical status of some of the items. Since the training items, themselves, were chosen to be a representative sample of grammatical items, rules derived in this way would tend to be veridical with respect to the grammar. That the rules that the subjects reported were relatively simple suggests that, for the most part, the underlying grammatical rules were too difficult to compute from a small sample of remembered training items in the limited time available.

The possibility that Reber and Allen's (1978) subjects computed the few rules that they did report only after the training task suggests that the limited degree of abstraction that they demonstrated arose from the demands of the transfer task rather than as an automatic by-product of original learning. In this regard, there is no unequivocal evidence in Reber and Allen's (1978) results to suggest that their subjects abstracted anything, either implicity or explicitly, during training.

Post-training computation of "abstract" properties of the training experiences has not received much attention in concept learning research. If subjects are able to engage in some "abstract" task (e.g., categorical transfer, production of some previously unseen prototype, statement of "rules", etc.) following training, the abstraction typically has been attributed to the training rather than the post-training phase of the experiment. While this is not implausible when the subjects are aware of the type of transfer task during training, it does appear unlikely for subjects who have no knowledge of any task subsequent to the learning of the items. For many of these post-training tasks, there is no reason to believe that the subjects have knowledge in an abstract form until they have been asked to produce some abstraction which they may compute, where possible, from their memory for the original training experiences. Two consequences follow from this view. First, following the computation process necessary to meet the demands of the post-training task, the subjects now have abstract knowledge. Second, the abstract knowledge is <u>explicit</u>. That is, if the subjects are asked the same post-training question at a later date, they could simply report their memory for the results of the prior computation.

A recent series of experiments by Hintzman, Nozawa, and Irmscher (1982) reported results consistent the view that subjects' explicit, abstract knowledge in many tasks may be computed from their memory for individual prior experiences rather than abstractly represented directly during the original training task. Hintzman <u>et al.</u> (1982) were concerned with demonstrating that the incidental learning of event frequency was not represented in the form of abstract propositions immediately following training, but rather was held indirectly in the form of multiple traces of the original training experiences. In one experiment, they compared the paired-associate recall of digits that had been paired with pictures for two groups of subjects. One group recalled the digits before being asked to estimate the relative frequency of occurrence of the different training picture-digit pairs whereas the other group recalled the digits after having produced their frequency estimates. Both groups were quite accurate in their frequency estimates, but the training frequency of the digits intruded upon digit recall only for those subjects who were asked to recall <u>after</u> producing their frequency estimates. Hintzman <u>et al.</u> (1982) interpreted these results to suggest that the frequency information was represented in a propositional (abstract) form only after the subjects had been asked to estimate the frequency of the items such that it interfered with the recall of the digits associated with the pictures. In a related experiment, Hintzman <u>et al.</u> (1982) found that if the subjects were given the frequency information in a propositional form during training, item frequency (which was now reported quite poorly) interfered with digit recall regardless of whether digit recall occurred before or after subjects' statements about item frequency. Thus, as Hintzman <u>et al.</u> (1982) concluded, the subjects' representation of event frequency in the former experiment was in a non-propositional form immediately following training which was then rendered into an explicit, abstract form via computation from subjects' memory for the individual experiences when the subjects were confronted with the unexpected frequency-estimation task.

Unlike simple event-frequency, and perhaps initial and terminal components of letter-strings, most of the abstract structure underlying the training items in an artificial grammar experiment is not computable from the individual training stimuli, at least not explicitly (Brooks, 1978; Reber, 1976). Thus, for the most part, the explicit rules and abstract summaries that subjects produce should be limited to the relatively simple aspects of the letter-strings that are explicitly computable. But, as Reber and Allen (1978) noted about the performance and explicit statements of their subjects, the explicit rules that the subjects do produce do not even approximately account for the subjects' performance in the categorical transfer task. A large part of the subjects' performance must arise directly from their memory for the individual training experiences. As is discussed next, the relationship between memory for individual experiences and relatively simple, explicit abstractions may be an integral part of performance in many tasks and domains.

Relation to Performance and Expertise

It is a sad, but unfortunately true comment on human performance that skilled performance in most complex domains does not follow directly from explicit knowledge of the rules. As any teacher can testify, being able to cite the rules for such things as reading and calculus is a long way from being able to read or to integrate some function. On the other hand, skilled performance in many domains seems inevitably to be associated with explicit knowledge of, at best, only rudimentary rules or "rules of thumb". Thus, explicit knowledge of the "rules" is neither sufficient for nor the primary basis of skilled performance in most complex domains. Yet, despite this, we often attribute skilled performance to knowledge of the rules.

Most teachers are aware of the loose connection between explicit knowledge of the rules or principles underlying some domain and skilled performance within that domain. As a consequence, the standard pedagogical remedy is to emphasize "practice in applying the rule", with the resulting successful performance attributed to the student having "finally learned the rule." It is also possible, however, that what the student gains from the practice in applying the rules of some domain is a collection of specific experiences that may be used, by analogy, to enhance application of the rules of the domain to some new instance or problem. In fact, if the analogy is close enough, the student can dispense with the rule altogether and simply reproduce the previously computed answer or operations. Thus, to a large extent, examples and practise problems may not serve simply to exemplify the abstract principle of interest, but may actually be the knowledge responsible for skilled performance.

The same dependence upon analogy to specific, prior experiences may apply to differences between educated novices and experts in complex domains (e.g., medicine, chess, etc.). Typically, the expert's explicit knowledge is not markedly superior to that of the well-read novice. In fact, in many cases, the expert and educated novice may have read the same books and performed the same rudimentary computations on cases and situations they have experienced. As with the above-described learning of artificial grammars, however, most of the underlying rules of the domain are probably too complex to be determined on a post-hoc, case by case analysis. And for the minimal rules that are readily amenable to explicit analysis of this sort, the amount of experience necessary to obtain them is probably quite small. Thus, the explicit abstractions of the novice and the expert may differ little because both have taken that form of analysis about as far as it can go in the absence of an intensive research program. Hence, the performance of the expert may exceed that of the novice not because the expert's extensive experience necessarily has equipped him or her with better rules, but because the vast collection of individual experiences that the expert has accumulated provides a close analogy to almost every new experience likely to be encountered within the domain.

Cast in terms of individual instances, the road to expertise appears to follow the adage that there is no substitute for experience. This may be true, but not necessarily in the manner in which this adage is usually interpreted. The encoding and breadth of transfer effects discussed earlier suggest that what is important is not that the individual acquire prior experiences that are literally similar to those within the domain of interest, but rather that the prior and subsequent events be encoded such that they may be perceived as similar. Thus, it may be possible to simulate the relevant experiences and thereby provide the learner with the appropriate instance-based knowledge to function adequately within a previously unexperienced domain.

Recent research by Howard and Ballas (1980; 1982) suggests an interesting application of this principle and, as well, extends the artificial grammar technology to a domain beyond that of the classification of letter-strings. Impressed with Reber's (e.g., 1969) demonstrations of the implicit learning of artificial grammars, Howard and Ballas essentially repeated a number of Reber's experiments, but substituted sounds for letters in the sequences from the artificial grammar that their subjects were to learn. The results were virtually equivalent to those of Reber's. Most interesting for our present concerns, however, are the results of an experimental treatment that was based upon Reber's (1969) demonstration that subjects could transfer their prior learning of grammatical items to new items from the same grammar even though the letters had been changed. In this case, Howard and Ballas (1982) associated easily described, environmental sounds (e.g., the hiss of a radiator, the clank of a hammer on metal) with particular node to node transitions of the grammar. Subjects were trained with verbal descriptions of the training sound sequences and then transfered to the actual sound sequences themselves. Despite the formal difference in their experience with the training sequences, subjects trained in this way performed as well on a test of categorical transfer as did subjects who received their training stimuli in the same format as the transfer stimuli. Thus, transfer between training and test need not be based on the literal similarity between events.

Although Howard and Ballas (1980; 1982) interpreted their results in terms of the abstraction of grammatical rules, their suggestions for the potential application of the results are still valid if the results are interpreted in terms of memory for specific training experiences. Howard and Ballas (1982) viewed their results in terms of the training of sonar operators, and suggested the results of the symbolic training could be used to design training programs to acquaint sonar trainees with the background (ship-board) transient noise sequences they would have to disambiguate from relevant signals. Since the training is symbolic, the trainees could acquire the relevant experience without having to undergo the actual experiences themselves, resulting in an efficient, land-based training program.

Relationship to Artificial Intelligence

The individual instances approach to classification and recognition is simple, potentially very accurate, and will work for virtually every stimulus domain. It would seem, then, to be a natural candidate for machine pattern recognition. Yet, as with its application to human performance, it has received only sporadic attention as a machine pattern recognition process. Hunt (1975), for example, suggests that one probable reason for the lack of attention to this approach in artificial intelligence circles is that "... it appears too simple to be interesting!" (p. 86) This may be true, but it hardly is justification for for the rejection of a potentially useful pattern recognition process.

In the the area of artificial intelligence, the individual instances approach to classification and recognition goes under the rubric of "nearest neighbor" (Hunt, 1975; Watanabe, 1969), and is considered to be but one of many proximity, or distance, algorithms available for machine pattern recognition. As mentioned earlier (see Chapter 2), classification of any given new instance according to the nearest neighbor algorithm is a function of the $\underline{n}$-closest ($\underline{n}$ >= 1) instances in some multi-dimensional, memorial hyper-space. According to Hunt (1975) and Watanabe (1969) the distance metric for this algorithm is Euclidean, although no where in my admittedly less than complete survey of the relevant literature could I find justification for this restriction of the nearest neighbor algorithm to a Euclidean distance metric in particular or distance metrics (i.e., dimensional logic) in general. As I argued earlier with respect to human performance based on this approach, the rejection as inadequate of an individual instances process in machine pattern recognition may be a consequence of unquestioned adherence to this one similarity metric when assessing the algorithm rather than to a failure of the approach considered more generally.

Although the nearest neighbor model has not played a major role in artificial intelligence applications, it is generally acknowledged to have one very major advantage over all other pattern recognition, proximity algorithms. Specifically, it does not require that the categories of instances be linearly separable for successful classification - a property, as mentioned earlier, used to great advantage in Medin and Schaffer's (1978; Medin & Scwanenflugel, 1981; Medin & Smith, 1981) tests of their context-cue model of classification learning. In fact, according to Hunt (1975), there is only one major problem with the nearest neighbor algorithm relative to the other methods considered in artificial intelligence work: the memory requirement. As he points out, efficient use of the nearest neighbor method requires both storage of the entire sample of (learning) instances and an information retrieval process that provides access to the nearest neighbor of any new instance quickly. Other methods, he notes, typically require storage of the dimensions upon which the instances vary, which, typically, will be far less than the number of instances, and where, consequently, access time is trivial. Hunt (1975) concludes, however, that "...the nearest neighbor model gives surprisingly good results considering its simplicity." (p. 88) And that "The major drawbacks to the nearest neighbor model are not lack of accuracy, but rather its large memory requirements and the fact that it, like all proximity techniques, depends very much upon Euclidean distance assumptions." (p. 88) Thus, the concerns for machine application of an individual instances approach are similar to those voiced by researchers concerned with modeling human classification performance. And, for the most part, these concerns apparently have provided the justification for consideration of alternative approaches and models in both areas.

As much of the discussion in this thesis has been an extended argument for individual instances as the basis of a large part of human classification behaviour, I would like to apply the same arguments in favor of reconsidering an individual instances approach to machine pattern recognition. First, it is clear that an individual instances approach is not tied to Euclidean distance assumptions. Adapting, say, Tversky's feature matching approach to inter-item similarity to machine similarity assessments is, as Hintzman and Ludlam (1980) have demonstrated, a relatively simple programming exercise that frees the judgements based upon individual instances from the problems associated with Euclidean distance metrics. Second, and perhaps more important, the concerns about the memory overhead associated with the storage (and retrieval) of individual instances are not as relevant for today's machines as they were even as little as ten years ago. With the advent of inexpensive, reliable and fast RAM chips in recent years, the ability to store and quickly access 64 kbytes of information has become a feature common even to most personal computers, with mega-bytes of even faster memory promised for the very near future. Besides, as with human classification, the arguments about the lack of economy of an individual instances approach to machine pattern recognition are suspect. The savings associated with storing just the features (or dimensions) relevant for a particular discrimination only represents an economy if the particular discrimination is the only one required and if the attainment of the knowledge about what is relevant for the discrimination is relatively cost-free. If the machine is required to make many different kinds of classification decisions (e.g., at different levels of abstraction) or deal with highly variable "features" or dimensions of variation that are only occassionally present (e.g., as in the reading of hand-written text), the storage, access, and processing requirements associated with abstractive algorithms may vastly exceed those associated with individual instances.

Whatever the merit of these theoretical considerations, Jim Dowe (1982) has developed an instance-based pattern recognition process for the Apple II micro-computer that appears to have succeeded where the more abstractive algorithms on much larger machines have failed. Dowe (1982) refers to the process as "artificial intuition" to emphasize that it is not analytical in the usual sense of artificial intelligence. As he describes the process: "It compares an incoming pattern with other patterns that it has in its memory, and then says, 'Of all the things I have ever seen, which of those things is most like this new pattern that I have.' And then it does what it should based on that match." (p. 30) Remarkably, Dowe (1982) reports that that the algorithm itself requires only about 4 kbytes of computer memory (and is implemented on a small, 64 kbyte computer), and that, since the algorithm functions independently of the source of the information in memory, it will work for speech recognition, the extraction of meaning from text, etc. with no modification in the algorithm itself. He further emphasizes the flexibility and adaptability of the process by noting that "...if a dyslexic individual uses the machine, then the machine will communicate like a dyslexic." (p. 31). While a dyslexic computer may not meet with everyone's idea of an advancement in artificial intelligence technology, such a result does emphasize the virtually limitless possibilities open to the application of this one, relatively simple process of analogy to individual, prior experiences.

Concluding Comments

On the basis of the evidence in the classification learning literature and the results of the work reported in this thesis, the case for the implicit abstraction of structure is a weak one. This raises the question as to why within cognitive psychology and more generally there exists such a widely held, and for the most part, unquestioned belief in implicit abstraction and complex, unconscious processes of incredible sophistication. At least part of this belief is probably attributable to a general failure to consider less abstractive alternatives and to the general tendency to describe effects in terms of the information reflected rather than in terms of how such information is represented in memory and acquired in the first place. But the attribution of complex, unconscious processes that vastly exceed what we are capable of conciously also may represent a form of reverse snobbery; that much of what we accomplish in perception and memory has little to do with our much prized possession of conscious intelligence. This may be true, but on the basis of the evidence it appears that much of what passes for implicit abstraction does so on the basis of a relatively mundane and unintelligent process that happens to work well, rather than on the basis of incredible abstractions that we can at best only hope to approximate in our conscious reflections.

Reference Notes

1. Brooks, L. R. On the Insufficiency of Analysis. Manuscript in preparation.

2. Hock, H. S., Tromley, C., Marshall, L., and Webb, E. Configuration Information in Visual Concept Formation. Paper presented at the annual meeting of the Psychonomic Society, Phoenix, Arizona, November, 1979.

3. Wittlesea, B. W. A. The Representation of Concepts: Abstraction and Instance Perspectives. Unpublished Doctoral dissertation, McMaster University, 1983.

References

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

Allen, R. and Reber, A. S. Very long term memory for tacit knowledge. Cognition, 1980, 8, 175-185.

Armstrong, S. L., Gleitman, L. R., and Gleitman, H. What some concepts might not be. Cognition, in press.

Baron, J. The Word Superiority Effect: Perceptual Learning From Reading. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 6). Linguistic functions in cognitive theory. Hillsdale, N. J.: Erlbaum, 1978.

Barresi, J., Robbins, D. and Shain, K. Role of distinctive features in the abstraction of related concepts. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104, 360-368.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. Discrete multivariate analysis: Theory and practice. Cambridge, Ma: MIT Press, 1975.

Bourne, Jr., L. E. An Inference Model for Conceptual Rule Learning. In Solso, R. L. (Ed.), Theories in cognitive psychology: The Loyola symposium. Potomac, Maryland: Erlbaum, 1974.

Bourne, Jr., L. E. Typicality effects in logically defined categories. Memory & Cognition, 1982, 10, 3-9.

Bourne, Jr., L. E., Ekstrand, B. R., Lovallo, W. R., Kellogg, R. T. and Hiew, C. C. Frequency analysis of attribute identification. Journal of Experimental Psychology: General, 1976, 105, 294-312.

Bransford, J. D. and Franks, J. J. Abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 351-350.

Brooks, L. R. Non-analytic concept formation and memory for instances. In E. Rosch and B. Lloyd (Eds.), Cognition and concepts. Hillsdale, N. J.: Erlbaum, 1978.

Bruner, J., Goodnow, J., and Austin, G. A study of thinking. New York: John Wiley, 1956.

Chumbley, J. I., Sala, L. S. and Bourne, Jr., L. E. Bases of acceptability ratings in quasinaturalistic concept tasks. Memory & Cognition, 1978, 6, 217-226.

Clark, H. H. The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and

Verbal Behavior, 1973, 12, 335-359.

Craik, F. I. M. Human memory. Annual Review of Psychology, 1979, 30, 63-102.

Craik, F. I. M. Encoding and retrieval effects in human memory: A partial review. In A. D. Baddeley and J. Long (Eds.) Attention and Performance IX. Hillsdale, N.J.: Erlbaum, 1981.

Dowe, J. A Personal Computing interview with Jim Dowe. Personal Computing, 1982, 6, 30.

Donaldson, W. Context and repetition effects in recognition memory. Memory and Cognition, 1981, 9, 308-316.

Erlich, S. F., and Rayner, K. Contextual effects on word perception and eye movements during reading. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 113-132.

Fienberg, S. E. The analysis of cross-classified categorical data. Cambridge, Ma.: MIT Press, 1977.

Flexser, A. J. Homogenizing the 2 x 2 contingency table: A method for removing dependencies due to subject and item differences. Psychological Review, 1981, 88, 327-339.

Franks, J. J. and Bransford, J. D. Abstraction of visual patterns. Journal of Experimental Psychology, 1971, 90, 65-74.

Fried and Holyoak, Cognitive Science Technical Report #38, March, 1982.

Gati, I. and Tversky, A. Representations of qualitative and quantitative dimensions. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 325-340.

Glenberg, A. M. Influences of retrieval processes on the spacing effect in free recall. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3,282-294.

Goldman, D. and Homa, D. Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 375-385.

Haber, R. N., and Schindler, R. M. Error in proof reading: Evidence of syntactic control over letter processing? Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 573-579.

Hayes-Roth, B. and Hayes-Roth, F. Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 321-338.

Hintzman, D. L. Repetition and Memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 10). New York: Academic Press, 1976.

Hintzman, D. L. Simpson's paradox and the analysis of memory retrieval. Psychological Review, 1980, 87, 398-410.

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

Hintzman, D. L. and Ludlam, G. Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. Memory and Cognition, 1980, 8, 378-382.

Hintzman, D. L., Nozawa, G. and Irmscher, M. Frequency as a nonpropositional attribute of memory. Journal of Verbal Learning and Verbal Behavior, 1982, 21, 127-141.

Hintzman, D. L. and Stern, L. D. Contextual variability and memory for frequency. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 539-549.

Homa, D. Abstraction of ill-defined form. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 407-416.

Homa, D. and Chambliss, D. The relative contributions of common and distinctive information on the abstraction from ill-defined categories. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1044, 351-359.

Homa, D., Cross, J., Cornell, D., Goldman, D. and Shwartz, S. Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. Journal of Experimental Experimental Psychology, 1973, 101, 116-112.

Homa, D. and Hibbs, B. Prototype abstraction and the rejection of extraneous patterns. Bulletin of the Psychonomic Society, 1978, 11, 1-4.

Homa, D., Sterling, S. and Trepel, L. Limitations of exemplar-based generalization and the abstraction of categorical information. Journal of Experimental Psychology: Human Learning and Memory, 1981, 7, 418-439.

Homa, D. and Vosburgh, R. Category breadth and the abstraction of prototypical information. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 322-330.

Howard, J. H. and Ballas, J. A. Acquisition of acoustic pattern categories by exemplar observation. Organizational Behavior and Human Performance, 1982, 30, 157-173.

Howard, J. H. and Ballas, J. A. Syntactic and semantic factors in the classification of nonspeech transient patterns. Perception & Psychophysics, 1980, 28, 431-439.

Hunt, E. B. Artificial Intelligence New York, New York: Academic Press, 1975.

Hunt, R. R. and Ellis, H. C. Recognition memory and degree of semantic contextual change. Journal of Experimental Psychology, 1974, 103, 1153-1159.

Jacoby, L. L., Craik, F. I. M., and Begg, I. The effect of decision difficulty on recognition and recall. Journal of Verbal Learning and Verbal Behaviour, 1979, 18, 585-600.

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

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

Kellogg, R. T. Feature frequency and hypothesis testing in the acquisition of rule-governed concepts. Memory and Cognition, 1980, 8, 297-303.

Kellogg, R. T. Feature frequency in concept learning: What is counted? Memory and Cognition, 1981, 9, 157-163.

Kolers, P. A. A Pattern-Analyzing Basis of Memory. In L. S. Cermak and F. I. M. Craik (Eds.), Levels of Processing in Human Memory. Hillsdale, N. J.: Erlbaum, 1979.

Lasky, R. E. and Kallio, K. D. Transformation rules in concept learning. Memory and Cognition, 1978, 6, 491-495.

Light, L. L. and Carter-Sobell, L. Effects of changed semantic context on recognition memory. Journal of Verbal Learning and Verbal Behavior, 1979, 9, 1-11.

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

McNemar, Q. Psychological statistics. New York: Wiley, 1969.

Medin, D. L. and Schaffer, M. M. Context theory of classification learning. Psychological Review, 1978, 85, 207-238.

Medin, D. L. and Schwanenflugel, P. Linear separability in classification learning. Journal of Experimental Psychology: Human Learning and Memory, 1981, 7, 355-368.

Medin, D. L. and Smith, E. E. Strategies and classification learning. Journal of Experimental Psychology: Human Learning and Memory, 1981, 7, 241-253.

Murphy, G. L. Cue validity and levels of categorization. Psychological Bulletin, 1982, 1, 174-177.

Neisser, U. Cognitive psychology. New York, New York: Appleton-Century-Crofts, 1967.

Neumann, P. G. An attribute frequency model for the abstraction of prototypes. Memory and Cognition, 1974, 2, 241-248.

Neumann, P. G. Visual prototype formation with discontinuous representation of dimensions of variability. Memory and Cognition, 1977, 5, 187-197.

Omohundro, J. Recognition vs. classification of ill-defined category exemplars. Memory and Cognition, 1981, 9, 324-331.

Osherson, D. N. and Smith, E. E. On the adequacy of prototype theory as a theory of concepts. Cognition, 1981, 9, 35-58.

Peterson, M. J., Meagher, Jr., R. B., Chait, H. and Gillie, S. The abstraction and generalization of dot patterns. Cognitive Psychology, 1973, 4, 378-398.

Posner, M. I., Goldsmith, R. and Welton, R. D. Perceived distance and the classification of distorted patterns. Journal of Experimental Psychology, 1967, 73, 28-38.

Posner, M. I. and Keele, S. W. On the genesis of abstract ideas. Journal of Experiemntal Psychology, 1968, 77, 353-363.

Posner, M. I. and Keele, S. W. Retention of abstract ideas. Journal of Experimental Psychology, 1970, 83, 304-308.

Reber, A. S. Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behaviour, 1967, 6, 855-863.

Reber, A. S. Transfer of syntactic structure in synthetic languages. Journal of Experimental Psychology, 1969, 81, 115-119.

Reber, A. S. Implicit learning of artificial grammars: The role of instructional set. Journal of Experimental Psychology: Human Memory and Learning, 1976, 2, 88-94.

Reber, A. S. and Allen, R. Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. Cognition, 1978, 6, 193-221.

Reber, A. S., Kassin, S. M., Lewis, S., and Cantor, G. On the relationship between implicit and explicit modes in the learning of a complex rule structure. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 492-502.

Reber, A. S. and Lewis, S. Implicit learning: An analysis of the form and structure of a body of tacit knowledge. Cognition, 1977, 5, 333-361.

Reddy, B. G. and Bellezza, F. S. Encoding specificity in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1983, 9, 167-174.

Reed, S. K. Pattern recognition and categorization. Cognitive Psychology, 1972, 3, 304-308.

Reed, S. K. Category vs. item learning: Implications for categorization models. Memory and Cognition, 1978, 6, 612-621.

Reed, S. K. and Friedman, M. P. Perceptual vs conceptual categorization. Memory and Cognition, 1973, 1, 157-163.

Reitman, J. S. and Bower, G. H. Storage and later recognition of exemplars of concepts. Cognitive Psychology, 1973, 4, 194-206.

Robbins, D., Barresi, J., Compton, P., Furst, A., Russo, M., and Smith, M. A. The genesis and use of exemplar vs. prototype knowledge in abstract category learning. Memory & Cognition, 1978, 6, 473-480.

Rosch, E. Natural categories. Cognitive Psychology, 1973, 4, 328-350.

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

Rosch, E. The nature of mental codes for color categories. Journal of Experimental Psychology: Human Perception and Performance, 1975b, 1, 303-322.

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

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

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

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

Russell, B. Aristotle's Logic. A history of western philosophy, London: Allen and Unwin, 1946.

Shepard, R. N. Recognition memory for words, sentances, and pictures. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 156-163.

Simpson, E. H. The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society, Series B, 1951, 13, 238-241.

Standing, L. Learning 10,000 pictures. Quarterly Journal of Experimental Psychology, 1973, 25, 207-222.

Standing, L., Conezio, J. and Haber, R. N. Perception and memory for pictures: Single trial learning of 2560 visual stimuli. Psychonomic Science, 1970, 19, 73-74.

Strange, W., Keeney, T., Kessel, F. S. and Jenkins, J. J. Abstraction over time of prototypes from distortions of random dot patterns: A replication. Journal of Experimental Psychology, 1970, 83, 508-510.

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

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

Tversky, A. and Kahneman, D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 1973, 5, 207-232.

Watanabe, S. Knowing and Guessing New York, New York: John Wiley & Sons, Inc., 1969.

Young, D. R. and Bellezza, F. S. Encoding variability, memory organization, and the repetition effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982, 8,545-559.

Zacks, R. T., Hasher, L., and Sanft, H. Automatic encoding of event frequency: Further findings. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982, 8, 106-116.

NOTES

1.1 As discussed by Russell (e.g., 1946), the failure to distinguish between statements of the form "Socrates is a man" and statements of the form "All Greeks are men" led to 2000 years of, as he puts it, "bad metaphysics". The problem is that the grammatical subject "All Greeks" is not a logical subject in the same sense as we typically consider "Socrates" to be; it is necessary to distinguish between the general (Greeks) and the particular (Socrates). For example, the statement "Socrates is a a man", if true, is simply a statement that there is at least one man, Socrates. In contrast, the statement "All Greeks are men", if true, does not mean that there are, necessarily, men who are Greeks any more than the statement "All unicorns are pleasant", if true, implies that that some (i.e., at least one) pleasant things are unicorns. The statement "All unicorns are pleasant" can be true without there being unicorns to instantiate it. One way of viewing this distinction is in terms of entities (individuals) and properties (or predicates). One can be or not be the entity "Socrates", but one either has or does not have the property of being Greek.

In some versions of modern logic, as opposed to traditional logic, all classes are treated as predicates such that the statement "All Greeks are men" is treated as the (logical) linking together of properties or predicates. Thus, a present-day interpretation of the statement is "If something is a Greek then it is also a man". This prevents some logical contradictions that were inherent in the classical treatment (i.e., a number of Aristotle's "valid" syllogisms become invalid) and also prevents one from equating a class of only one member with that one member, which, over the years, had produced no end of confusion and paradox in philosophy and mathematics (see Russell, 1946).

The approach toward concepts or classes taken in this thesis maintains the logical distinction between classes and individual exemplars, but meets this objective by treating all labels, class or individual, as predicates. ~~Thus, for~~ example, "Socrates" may be treated as a class label or predicate in the same way as the term "Greek". The interpretation of the statement "Socrates is a man" in this case would be "If an event may be labelled as 'Socrates' then it also may be labelled as a man". One advantage of this approach is that it suggests that the recognition of individuals as members of of a class and the recognition of them qua individuals need not be considered as fundementally different acts. In both cases, one is assigning a class label, albeit at different levels, to an event.

1.2 The results of the first three experiments, reported in Chapters 4, 5, and 6, have been prepared for publication in Vokey, J. R. and Brooks, L. R. Taming the clever unconscious: Analogic and abstraction strategies in artificial grammar learning. Cognition, in press.

2.1 Rosch, Mervis, Gray, Johnson & Boyes-Braem (1976) equate the maximization of cue-validity with "basic-level" categories. That is, they claim that the basic-level category in a hierarchical structure of categories is that level that maximizes cue-validity, where cue-validity for a given level is defined as the sum of the individual cue-validities of all features associated with the category. As Murphy (1982) notes, however, this conception is fallacious since the more inclusive (i.e., higher) category levels necessarily will have the greater category cue-validity, at least in so far as the definition of cue-validity is given as the conditional probability that given the cue, some object or event is a member of the category. While this is the definition that Rosch et al. (1976; see also Rosch, 1978) use, the common conception in the literature appears to be that cue-validity is the same as category diagnosticity. That is, the more valid the cue, the more likely it is that the cue predicts category membership and the more likely it is that category membership predicts the cue. With this definition, basic-level categories do appear to maximize cue-validity. Consider the basic-level concept "bird" and the cue "has wings". The superordinate level, "animal", is well predicted by the cue (i.e., has high cue-validity by the original definition), but poorly predicts the cue. Conversely, the subordinate level, "robin", is poorly predicted by the cue, but predicts the cue virtually perfectly. The basic-level, "bird", predicts the cue well and also is well-predicted by cue. Thus, by equating cue-validity with category diagnosticity, Rosch's notions about basic-levels and, possibly, the reasons for their apparent predominance may be retained.

2.2 The original stimuli were not random-dot patterns; they were distortions of common, geometrical shapes and letters. Moreover, Peterson, Meagher, Chait, and Gillie (1973) have evidence to suggest that distortions of known shapes are treated differently by the subjects than are random-dot patterns.

2.3 In Rosch, Simpson and Miller's (1976) experiment, "typicality" is defined with respect to the category to which a given exemplar belongs and not the set of training experiences presented to the subjects. Thus, while a particular exemplar may have been typical of its category, it may not have been representative of the set of training experiences. In fact, in a second experiment, Rosch, Simpson, and Miller (1976) inversely related the presentation frequency and "typicality" of the training items, but still obtained the "typicality" effects discussed in the text.

2.4 Chumbley, Sala and Bourne (1978) reported an experiment that they claim provides evidence inconsistent with a prototype or "averaging" model and, thereby, in their view, consistent with a feature-frequency model. The heart of their design was the comparison of the learning of the ill-defined category of public acceptability of uniforms for two different training conditions that differed in whether the the stimulus variation was quantitative (thereby allowing the calculation of the "average" or prototypical acceptable uniform) or qualitative (where "averaging" is not possible; i.e., what is the average of nylon and leather?). They found no difference between the two conditions.

That is, in both cases, the subjects treated the stimulus dimensions in a fashion consistent with feature-frequency and inconsistent with abstraction of a prototype. Although this result would appear to provide strong evidence for a feature-frequency over a prototype model of classification learning, the experiment is flawed. Specifically, all of the stimuli were presented to the subjects as verbal descriptions. The effect of this would be to render all dimensions of the uniforms, "quantitative" or otherwise, as consisting of discrete values. Thus, it is not surprising that the results were similar for the two conditions. But there is no reason to believe that such would be the case if the "quantitative" dimensions were presented in a fashion that stressed their continuous dimensionality (e.g., if pictures rather than descriptions of the uniforms were used). While I do not present this experiment as representative of those in the area, it should be noted that Chumbley et al. (1978) presented their experiment as a serious test of the two models and as providing strong disconfirmation of a prototype model. Whatever the quality of their evidence, however, I do agree with part of their conclusion: "In fact, there is no evidence that subjects have anything in memory before they begin rating the stimuli except the memories of some of the exemplars thay have experienced. It is entirely possible that subjects do not normally judge test stimuli against a prototype or schema, frequency or central tendency, but against the raw memory events, as with the average distance model." (p. 226). Oddly, after making this claim, they add "A frequency model which assumes every act of remembering is a conceptual inference based on independent episodic traces will handle the data adequately." (p. 226) Maybe so, but given their premise, a "frequency model" seems like an unnecessary complication.

2.5 This result may be a consequence of the fact that, while the "features" in these latter stimuli may have been frequent, their combinations may not have been as frequent. Unfortunately, Neumann does not report the training stimuli he used, so this notion can not be evaluated, nor can the notion that, as a consequence of being composed of infrequent combinations, these stimuli may have been less similar to the individual training stimuli than was the case for the "average" or prototypical stimulus.

2.6 Clearly, features may be salient because they are diagnostic of the category, particularly if the subject is using some fairly reliable "rule of thumb" based upon one or two of the features (see Medin & Smith, 1981). The point in using salience in the model is to cover those many cases in which features are emphasized for reasons other than cue-validity.

2.7 Reed (1978) reported a series of experiments in which he claimed to have demonstrated that "category learning" occurs independently of, and more rapidly than, "item learning". He interpreted the results to suggest that "category learning" involves the abstraction and use of categorical prototypes rather than exemplar-specific knowledge. He also suggested that his experimental paradigm provides a useful technique for identifying the various types of knowledge that subjects acquire and use in fulfilling categorical task objectives. There is no reason that I can see to accept these claims.

In a computer simulation of Reed's experiments, I was able to replicate his two primary findings using only similarity to prior instances as the basis of both item and category responses. The simulation was implemented as follows. Reed's (1978) face stimuli were converted into numerical analogues. Each face could vary at three levels along four dimensions, where the levels within a dimension were ordered. Similarity between any two faces was defined as the sum of the absolute values of differences across the four dimensions. Identity, then, would yield a similarity score of zero, while maximal dissimilarity would yield a score of eight. Memory was treated as a push-down stack (LIFO structure) such that the memory search was from the most recent to the least recent experience, with the search terminating on a perfect match (similarity equal to zero) or when all traces had been searched. Traces were added to memory only when at least one of the two responses (item or category) was correct, and only correct responses were stored with the faces in memory. Responding was on the basis of "best-match"; the item and category responses generated for any given face were those associated with the most recent face in memory that was most similar to the present face. Where one of the two responses was unavailable (it had been incorrect when the best-match face had been stored in memory) a response was generated at random (i.e., a guess was generated).

The simulated experiment followed Reed's (1978) procedure for his Experiment 3. Forty simulated subjects were given 10 learning trials, where each learning trial consisted of a random ordering of the ten faces. As in Reed's (1978) experiment, there was a significant effect of response-type; $F(1/39) = 531.27$, $p = .0000$. The simulated subjects were significantly more accurate in providing the category labels of the faces (83% correct) than the item labels (37% correct). Also as in Reed's (1978) experiment, accuracy increased significantly over trials [$F(9/351) = 160.01$, $p = .0000$], and category labelling accuracy increased at a significantly faster rate over trials than did item labelling accuracy; $F(9/351) = 11.77$, $p = .0000$. Probably most importantly, there was no relationship at all between "item learning" and "category learning". As would be expected from the implementation of the simulation, "learning" the correct item label and "learning" the correct category label were virtually independent as in Reed's (1978) Experiment 3, even though all "learning" and all responding was instance-based. As discussed in the text, contrary to convential wisdom, poor "item learning" coupled with good "category learning" <u>is not</u> inconsistent with categorical responding on the basis of individual instances (nor, for that matter, is good "item learning" coupled with poor "category learning" inconsistent with individual instances). This was graphically demonstrated in the simulation. Even by the tenth trial, my 40 simulated subjects were averaging only 59.5% correct with their item-label responses, while averaging 94.5% correct with their category-label responses.

While I do not take my simulation as a serious model of either memory or concept formation, the simulation does make it clear that

Reed's (1978) a priori rejection of exemplar models as being unable to account for the results of his experiments is simply incorrect. Even a simplistic approach as the one taken in the simulation can mimic the results of his experiments. Furthermore, the simulation demonstrates that equating some measure (e.g., "item learning", recognition, recall) of the degree of subjects' memory for the individual items with the usefulness of the subjects' prior experience with the individual items for categorization is in error. There is no reason to expect that there will be a simple, monotonic relation between "item learning" (however assessed) and categorization (however assessed) even for instance models. In fact, as discussed in the text, there is good reason to expect this relationship to be complicated and variable across tasks and measures. Finally, the simulation demonstrates that Reed's (1978) claim that his paradigm provides a convenient method of distinguishing among the various types of information and processes that subjects use in concept learning tasks is mistaken. Only in the situation where one accepts Reed's (1978) assumptions about what it is that "item learning" and "category learning" measure could his paradigm conceivably provide information not readily obtainable from the more typical concept learning tasks.

2.8 Correctly selecting the training stimulus in a triplet does not mean that, therefore, the a priori best analogy was contacted. Thus, using correct responses to estimate the number of training items in memory, as Omohundro (1981) attempts to do, is a questionable procedure for this, and many other reasons.

2.9 Reber & Allen (1978) used a within-subjects design. One-half of the subjects received the observation training and test before the paired-associate training and test, while the remaining subjects received these conditions in the reverse order. Performance on these tasks converged from first to second training condition. That is, most of the differences between observation and paired-associate training discussed by Reber & Allen occurred during first task performance. As would be expected from the my interpretation of these results, the reduction in differences from first to second task performance was accompanied by a reduction in differences in response tendencies between the two conditions (see Table 2.1).

3.1 In general, in all experimental designs, a distinction may be drawn between two different <u>internal validity</u> problems. The first of these, the <u>reliability</u> of any given effect, is what is tested by the chosen statistic. An effect is considered reliable if it may be concluded that it was unlikely to have arisen as a consequence of measurement or sampling error. On this basis, I have no quibble with the basic finding in the classification learning literature that the categorical status of the items was correlated with the dependent measures in the experiments. It is with the second type of internal validity that a problem arises: <u>extraneous covariation</u> with the independent variable (e.g., grammatical status in Reber's experiments). The extraneous covariation with the independent variable may be limited to a single extraneous variable, or may reflect the conglomeration of many variables. In either case, the effect may be modeled with a single extraneous variable.

Suppose that for a given experiment we have three sources of systematic variance (unsystematic or "error" variance is associated with internal validity of the first type): Grammatical status (G), the independent variable of interest, some response measure (R), the dependent variable, and some source (or sources) of systematic extraneous variation (E) that has not or could not have been recorded. The effect of interest, the correlation between G and R in the absence of variation due to E is given by the standard formula for partial correlation (e.g., Kerlinger & Pedhazur, 1973):

$$r_{GR.E} = \frac{r_{GR} - r_{GE} \cdot r_{RE}}{\sqrt{1-r_{GE}^2} \cdot \sqrt{1-r_{RE}^2}} \qquad (1)$$

Of course, the partial correlation is the conclusion that the experimenter wishes to infer, given that the simple correlation r_{GR} is reliable. Solving equation (1), then, for r_{GR}

$$r_{GR} = r_{GR.E} \cdot \sqrt{(1-r_{GE}^2) \cdot (1-r_{RE}^2)} + r_{GE} \cdot r_{RE} \qquad (2)$$

yields the equation for the observed correlation of interest. Note first that if the extraneous variation does not covary with G then equation (2) reduces to

$$r_{GR} = r_{GR.E} \cdot \sqrt{1-r_{RE}^2} \qquad (3)$$

which is equivalent to the semi-partial correlation ($r_{R(G.E)}$) between G and R. The extraneous variation in this case simply reduces the the reliability of the effect due to the unreliability of the measurement of R induced by E. Given that r_{GR} is reliable and that equation (3) holds, the experimenter is justified in concluding that G and R are directly related in the particular experiment, although the size of the effect will be underestimated as a function of r_{RE}. Also note that the observation of a reliable r_{GR} if G and E are correlated but R and E are not also indicates of a direct relationship between G and R. In this case, the simple correlation between G and R is equivalent to the semi-partial correlation $r_{G(R.E)}$, and is underestimated to the extent that the measurement of G is unreliable due to covariation with E. In general, as long as at least one of the two variables in question (i.e., G and R in this case) is independent of E, the observed correlation, if reliable, may be taken as indicative of the effect.

This general conclusion is perhaps more easily appreciated by setting the partial correlation in equation (2), $r_{GR.E}$, to zero:

$$r_{GR} = r_{GE} \cdot r_{RE} \qquad (4)$$

Here it may be seen that in the absence of any direct correlation between G and R it is still possible to observe a reliable, but spurious, r_{GR}, as long as both G and R are correlated with E. The observed correlation will be negative if one of the two correlations with E is negative, and positive if both of the correlations with E are positive or both negative. Thus, in those cases in which the extraneous variation is correlated with both variables of interest, neither the existence of a direct relationship nor the sign of that relationship may be inferred, regardless of the reliability of the observed r_{GR}.

My main contention is that the categorical status of any given item in Reber's experiments and the majority of experiments discussed in Chapter 2 is confounded with the similarity (the "extraneous" variation in Reber's view) between that item and some one or more members of a given category; that grammatical items are more similar to other grammatical items than are non-grammatical items. Assume, for the experiments reported in this thesis, that the direct relationship between categorical status and the response measure is approximately given by equation (4); that the partial correlation between categorical status and the response measure is small to non-existent. Assume also that the correlation between the response measure (R) and item-specific similarity (S) is reliable. If this is the case, then the observed relationship between grammaticality and responding should vary as a function of r_{GS} (where S corresponds to E in equation (4)), all other things being equal. The observed relationship should be positive if r_{GS} is positive ("close" grammatical items and "far" non-grammatical items), negative if r_{GS} is negative ("far" grammatical items and "close" non-grammatical items), and near zero if r_{GS} is zero (all items). As is demonstrated in subsequent chapters, this is what is found.

3.2 As is explained later, each transfer item serves as both a "close" and a "far" item across subjects in the basic paradigm. Thus, the terms "close" and "far" are enclosed in quotes to emphasize that, unlike categorical status, the similarity manipulation is not a property of the items, but rather a property of the relationship between the transfer and training items for subjects trained with a given list of training items. Thus, there are no close or far items as such. The expressions "close" and "far" should be taken as shorthand for the more cumbersome statement of the training-transfer similarity relationship of the items across subjects.

3.3 This manipulation does more than control for a confounding with categorical status. All potential covariation with specific similarity across the different items has been eliminated by manipulating specific similarity within items. Thus, for the specific similarity manipulation, concerns such as those raised by Clark (1973) about the internal validity of effects produced by using different sets of items may be safely ignored. These concerns may be raised with respect to the "manipulation" of categorical status, but in this case, they represent the exact issue confronted in the experiments; that the effect of categorical status is due to a confounding with other variables. Rather than attempt statistical solutions to the problem, the present research attempts to resolve the problem directly by unconfounding the suspected covariates.

3.4 The training procedure used in these experiments differs in at least two ways from the procedure of category discrimination training used in most concept learning research. First, only a single category of items is used. The subjects do not encounter the alternate category to be discriminated until the categorical transfer task. Second, the subjects are not asked to learn the items as members of a category. Rather, their task is to learn the items as individuals. Thus, whatever categorical information the subjects obtain is incidental to the original training task.

Although the training task may appear to be a strange one for the investigation of concept learning, there are several arguments that can be advanced to justify the approach. In particular, the task appears to have higher ecological validity than discrimination training. As model of how we typically acquire conceptual knowledge, discrimination training implies that concept learning is normally a result of categorical contrast at the point of learning; that we typically acquire our knowledge about a particular category in contexts that require the discrimination of members of that category from members of some other. However, discrimination training is not <u>necessary</u> for successful subsequent discrimination. Both Reber (e.g., 1976) and Brooks (1978) have shown that subjects can successfully discriminate exemplars of different categories even if the categorical distinctions were not apparent during training and if one of the categories had not even occurred during training. Moreover, prior experience with the alternate category for the discrimination is insufficient for many world tasks, as well as unnecessary. The features that are relevant for one categorical discrimination (e.g., dogs vs. cats) may not be those that are relevant or most effective for another (e.g., dogs vs. tables). If all one "knows" about dogs is how they differ from cats, one can hardly be said to have the concept "dog". Finally, the results of Brooks' (1978) experiment on the learning of artificial grammars suggests that the importance of discrimination training for concept learning with these materials is minimal. In one of the conditions of Brooks' (1978) experiment, the apparent categorical contrast at the point of learning was orthogonal to the subsequent discrimination required of the subjects. Had the subjects concentrated on the contrasting information provided during training their performance should have been adversely affected. However, the subjects performed just as well as other subjects given discrimination training consistent with the discrimination required of them during transfer.

4.1 An error in the algorithm used to generate the addition transfer items resulted in two of the "far" addition items for List 2 training (VXRM and MVRXRR) having to be reclassified as "close" addition items for both the transfer and recognition phases of the experiment, Similarly, one of the "far" addition items for List 1 training (MTTTVT) had to be reclassified as a "close" addition item for the categorical transfer phase of the experiment (MTTTVT was not used in the recognition test of List 1 training).

4.2 In each of the six experiments to be reported, the reported number of subjects refers to the number actually used for analysis. In every experiment, the data from a few of the subjects tested (less than five) were not used, either because the subject failed to respond to any of the items or because the subject produced the same response to every item. In these cases, the booklets were reproduced and new subjects were tested with them.

4.3 Because of the a priori unequal cell frequencies in a few cells of the design due to the reclassification of a few of the items (see note 4.1), the obtained frequencies of the affected cells in Experiments 1, 2, and 3 were adjusted for each subject to reflect a score out of 8 for categorical transfer and a score out of 4 for the recognition test to equate the affected cells with the remaining cells of the design.

4.4 Although the ratio of estimated variance components provides a measure of the relative difference in magnitude of effect between grammatical status and specific similarity, it does not lend itself easily to assessment of the statistical significance of the difference. Since demonstrating that specific similarity produces a larger effect on responding than does grammatical status is key to a number of my arguments, an alternative measure seemed warranted. This measure was computed as follows. First, the F-ratio for each effect was converted to a point-biserial correlation coefficient (i.e., r = [F / (F + df)] ^ 0.5) to provide a direct measure of effect size. Since these effects have independent error terms, this is equivalent to calculating the partial correlation between each independent variable and the response measure. Second, these correlations were converted to z-scores by the Fisher r to z transformation (i.e., $z = 0.5 * \ln(1 + r) - 0.5 * \ln(1 - r)$). The difference between the two z-scores was then divided by the standard error of the difference between the two z-scores (i.e., $[2 / (N - 3)]$ ^ 0.5) to yield a z-score that was treated in the usual manner (see McNemar, 1969).

4.5 There are several ways of conceptualizing these interactions involving specific similarity. All of them must account for the correlation between the magnitude of the specific similarity effect and the overall rate of responding "grammatical" for a given set of items. This eliminates signal detection theory as a model since it usually assumes independence between d′ (i.e., effect size) and the criterion (i.e., overall response rate). The model presented here is a simple one that is predicated upon the properties of the normal distribution. Assume that each experience that the subjects have with each transfer stimulus produces a level of "familiarity" that may be thought of as a value along a dimension of similarity to specific, prior experiences. Assume also that these values are normally distributed both across different experiences with the same nominal stimulus and across stimuli. Each nominal stimulus, then, would have a mean level of "familiarity", and sets of stimuli (e.g., "close" items) could be described in terms of the average level of "familiarity" that they produce. If it also is assumed that the probability that a given stimulus (on a given trial for a given subject) or a given stimulus type is labelled "grammatical" is a function of its level of "familiarity", then the function relating responding to "familiarity", and, hence, responding to the independent variables, may be given as the familiar ogive produced by the intergral of the normal distribution over all levels of "familiarity". This is shown in Figure 4.5.1. Note that, for overall response rates less than 50%, response rate is a positively accelerated function of the "familiarity" level. Consequently, for equal differences in "familiarity", overall rates of responding close to 50% are associated with larger differences in response rates to stimuli that differ in "familiarity" than are overall rates of responding that are much less than 50%. For example, the difference in responding to "close" and "far" stimuli should be greater for sets of "close and "far" stimuli where the overall response rate for the set is close to 50% (e.g., substitution items on the first pass) than where the overall response rate is quite low (e.g., addition items on the second pass), and this reduction in response rates across sets of items should occur more as a function of reduced responding to "close" items than reduced responding to "far" items, exactly as was found in the experiment.

There are a number of predictions that arise from the model. For example, as overall response rates exceed 50%, the relationship between the magnitude of the specific similarity effect and overall response rate should reverse. Extremely high overall response rates should produce smaller response rate differences between "close" and "far" items than overall response rates closer to 50%. Unfortunately, overall response rates for sets of items in this and subsequent experiments rarely exceeded 50%, so this prediction may not be evaluated. Moreover, attempts to manipulate response rates

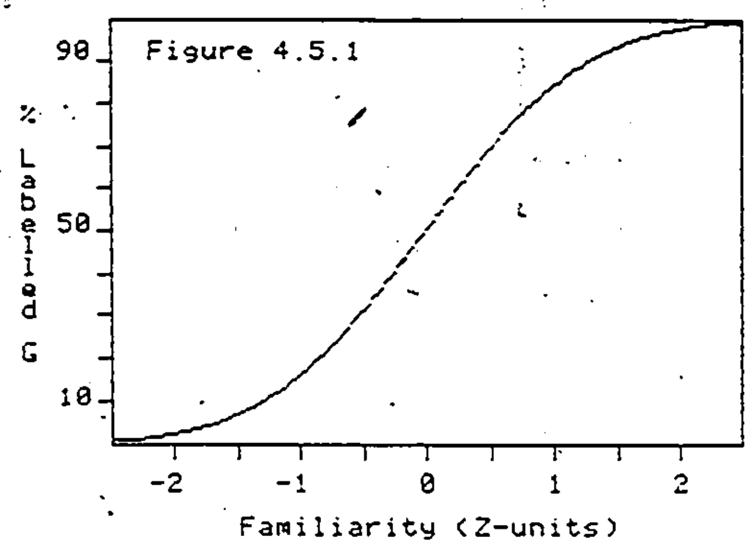

Figure 4.5.1: Responding as a function of "familiarity". The ogive produced by the integration of the standard normal distribution relating responding to the hypothesized underlying "familiarity" of the transfer items. Note that for rates of responding less than 50%, rate of responding "grammatical" is a positively accelerated function of "familiarity", and that, as a consequence, equal differences in "familiarity" are associated with larger differences in response rates as "familiarity" approaches the mean (Z = 0).

independently of the "familiarity" of the items (i.e., by disproportionate pay-offs and the like) would invalidate the model. However, the argument presented in this thesis is that the different global encoding operations (e.g., Observation training, No Label training) do affect the overall "familiarity" of the items and, hence, the overall response rate. Thus, the increased (but still less than 50%) overall response rate associated with one encoding condition relative to another should be associated with an increased effect of the specific similarity variable. Manipulations designed to increase the "familiarity" of specific items relative to other items are presented in Experiments 4 and 5. As discussed there, these manipulations to affect the "familiarity" of specific items should, in line with the simple model presented here, affect the magnitude of the effect of the specific similarity variable (in itself, a manipulation of the "familiarity" of specific items).

6.1 The discussion of different levels of representation of individual events raises an issue about the multiple meanings of the term "abstraction". "Abstraction" in psychology has at least two distinct meanings. One refers to the representation of specific events. The mental event corresponding to some specific occurrence is not a veridical copy of the experience, but rather is "abstract" in the sense that only some aspects of the event have been included (certain components have been "abstracted" or "lifted away" from the event), and many other components may have been added by a process of elaboration (e.g., the relation to other events, the "name" of the event, the significance to the organism, etc.). The second meaning refers to the summary of a number of events; the "abstract" representation that is gleaned from the multiple experiences of the same nominal event of events of the same type or from the same category. By the former meaning of "abstract", representations of items in terms of higher-order features are clearly "abstract". But the same may be said of representations in terms of letters in position since representing an item as a sequence of letters is certainly "abstract". However, in neither of these cases is the representation necessarily "abstract" in the latter sense since each item or even each experience with each item may yield a unique memorial representation rather than be combined with some summary representation of a number of different items or experiences. Brooks (1978) has suggested the terms "analytic" and "non-analytic" to refer to the presence or absence of "abstraction" in this latter sense. In his terms, then, the representation of the items that I am attributing to the Grammatical Mnemonic subjects is non-analytic, but at a higher level of "abstraction" than that of their counterparts in other training conditions.

7.1 To program the computer to generate all possible (legal) items less than or equal to some maximum length from any given finite-state grammar, it is necessary that the grammar be represented in a form that may be easily manipulated by the computer. The artificial grammar depicted in Figure 7.1 may be represented as 24 node to node transitions, where each transition produces either one of the 5 letters (M, V, R, X, T) or a space (for transitions that exit the grammar). These 24 transitions completely define the grammar and, consequently, may be interpreted as the minimal sub-rules or elements of the artificial grammar. All of the higher-order rules or patterns of the grammar may be produced by the recursive application of these sub-rules. Sets of items that are matched for use of these sub-rules, then, are "matched" with respect to the grammar. While matching at this level does not ensure matching for higher-order or "emergent" patterns or rules, visual inspection of random samples of matched sets (see text) indicated that pronounced differences at these higher-order levels, given matching of the sub-rules, were unlikely. This is not to say, however, that differences at these higher-order levels would not emerge if some other method of inspection or encoding of the items (e.g., the grammatical mnemonics in Experiment 3) was used.

7.2 Since there are 75 pairs of items, the total number of ways they may be combined into one list of 16 pairs is 75! / (16! (75-16)!). Similarly, given one list of 16 pairs, the total number of ways that a second list of 16 pairs may be produced from the remaining 75-16 = 59 pairs is (75-16)! / (16! (75-32)!). And the total number of ways of producing two lists of 16 pairs each is given by the product of these two values. In general, the total number of ways of that n items may be combined into p sets of r items each (pr <= n) is given by $n! / ((r!)^p (n-pr)!)$, which, in the present case, is approximately equal to 10^{30}. Since this is functionally infinite, some method other than systematic permutation and investigation of the possible lists of item pairs was required to produce the matched lists.

7.3 The chi-square values (and degrees of freedom) for tests of independence for each of the variable criteria are as follows. Training List by Item Length: $\chi^2(4) = 2.36$. Training List by Sub-rules: $\chi^2(23) = 4.90$. Training - Transfer by Sub-rules: $\chi^2(23) = 5.02$. Training List by Position of Change: $\chi^2(2) = 0.58$.

8.1 As in all of the previous experiments, there was no significant interaction of the categorical status and specific similarity of the items. However, the effects of these two variables were related through a three-way interaction with pass. As in Experiment 2, the specific similarity effect was larger for grammatical items on the first pass (C-F = 9.0%) than for non-grammatical items on the second pass (C-F = 3.7%), with the size of the specific similarity effect decreasing systematically as a function of the decreasing rate of responding associated with the conjunction of grammaticality and pass.

8.2 As in Experiment 4, the effect of previous recognition test exposure interacted with a number of variables in the experiment. One of these was a significant interaction with the appropriateness of the transfer item labels. The significant main effect of prior exposure occurred almost exclusively on appropriately labelled items, possibly because these items received more attention during recognition due to their confusability with the training items. Thus, they may have appeared to be more "familiar" on the test of categorical transfer than previously exposed, inappropriately labelled items and "new" items associated with labels of either type. Prior exposure also entered into a three-way interaction with specific similarity and grammaticality. In this case, the interaction was similar to that involving pass and these two variables (see Note 8.1).

8.3 The non-significant interaction of the specific similarity of the items and the appropriateness of the associated labels depicted in Figure 8.3 interacted significantly with the grammatical status of the items. In fact, the interaction between appropriateness and specific similarity shown in Figure 8.3 represents a diluted (and, hence, non-significant) version of the expected interaction produced by collapsing across grammatical and non-grammatical items. For non-grammatical items, there was no evidence of the expected interaction of specific similarity and label appropriateness (C-F for appropriate labels = 4.5%; C-F for inappropriate labels = 7.4%). For grammatical items, however, the interaction was as expected (C-F for appropriate labels = 16.4%; C-F for inappropriate labels = 3.7%). Thus, for items that had the higher rate of acceptance as "grammatical" (i.e., the grammatical items), the basis of this acceptance was primarily the similarity between training and transfer items as encoded. For the acceptance of non-grammatical items, the appropriateness of the associated labels was of no consequence. Since specific similarity did not interact directly with grammaticality but was significant as a main effect, variation in the acceptance of non-grammatical items was primarily a function of the formal similarity between the training and transfer items. As discussed in the text, the present results suggest that the expected encoding context effects were merely weak in the present experiment, possibly due to the subjects encoding the items and associated labels relatively independently of each other, rather than absent altogether.

9.1 As in Experiments 4 and 5, prior exposure of the items (and associated mnemonics) interacted significantly with a number of the other variables. In fact, prior exposure in the present experiment interacted significantly with each of the remaining variables of the categorical transfer phase. For example, there was no effect of the grammatical status of the items on categorical transfer responses unless the items had been previously encountered on the test of recognition (G–NG for "old" items = 9.3%; G–NG for "new" items = -0.1%). Similarly, the effect of the specific similarity of the items was significantly larger for previously experienced items (C–F for "old" items = 24.3%) than items encountered for the first time during categorical transfer (C–F for "new" items = 16.3%). On the other hand, the interaction of the appropriateness of the mnemonics with prior exposure was significant in the opposite direction. The effect of the appropriateness of the mnemonic associates of the items was larger on "new" items than "old" items. Thus, prior exposure to the items (and their associated mnemonics) apparently reduced the subjects' ability to discriminate among the items on the basis of mnemonic appropriateness, possibly because the increased familiarity of the pre-exposed items renders them more "appropriate" than the remaining items, while enhancing their ability to discriminate among the transfer stimuli on the basis of grammaticality and specific similarity. Prior recognition exposure also interacted significantly with the variable of pass such that pre-exposed items received more "grammatical" responses than did "new" items on the first pass, with the reverse, to a limited degree, being true on the second pass. The remaining significant effect in the transfer phase of the experiment was a three-way interaction between grammaticality, mnemonic appropriateness, and, again, prior recognition exposure. The effect in this case was a moderation of the two interaction effects involving these three variables discussed above. Described in terms of the interaction between grammaticality and pre-exposure, the reduction in the effect of grammaticality due to the interaction with pre-exposure of the items was larger for items associated with appropriate mnemonics than for items associated with inappropriate mnemonics.

Appendix A

The mnemonic phrases associated with each of the items in the mnemonics conditions, and the animal names used in the Label condition of Experiment 5:

A. Overlapping Mnemonics, Experiment 1

List 1

VXM	"Vermont 10" and Montreal.
MVRXR	Montreal and Vermont are Rated "X" by Reviewers.
VXTVX	"Vermont 10" TV is "X".
MTTTTVT	Montreal's Thousands Take The TV Times.
MTTVRXM	Montreal's Top TV is Rated "X" by Many.
VXRRRRR	"Vermont 10" is Rated Racy, Risky, Racist and Repetitive.
VXTTTVT	"Vermont 10" Takes The TV Times.
MVRXTVX	Montreal Views Rated-"X" TV as eXcellent.

List 2

MVX	Mandy Viewed X-rays.
MTTVT	Montreal's Top TV is Trite.
VXRRM	View X-rays of Rowdy Randy's Mom.
MVRXRRM	Mandy viewed Rosy X-rays of Rowdy Randy's Mom.
MTVRXRR	Montreal's TV is Rated "X" in Remote Regions.
MTVRXVX	Montreal's TV is Rated "X" by Viewing eXperts.
VXTVRXM	"Vermont 10" TV is Rated "X" by Many.
VXVRXRR	Vivian Xeroxed Very Rosy X-rays of Rowdy Randy.

B. Unique Mnemonics, Experiment 2

List 1

VXM	Viruses eXcite Macrophages.
MVRXR	Muscular Venetians Rowed across the River.
VXTVX	Vertical crosses To Vertical crosses.
MTTTTVT	Missles, Tanks and Trucks Threaten Vietnamese Tribes.
MTTVRXM	Many Transient TorontoniansVoiced Rich seXual Music.
VXRRRRR	Vast eXcessive Rats Ravenously Ravaged the Retarded Rabbit.
VXTTTVT	VeXed Tiny Toads Talked Very Thoughtfully.
MVRXTVX	Mystified Venusians Ran across The Vanishing "X".

List 2

MVX	Many Varied eXtras.
MTTVT	Marsha's Too Tight Vest Tore.
VXRRM	View 10 Ravenous Rats Munching.
MVRXRRM	Medics Very Rightly eXposed Rick's and Robert's Muscles.
MTVRXRR	Mark That Village Railway crossing "RR".
MTVRXVX	Montreal's TV is Rated "X" by Viewing eXperts.
VXTVRXM	5 or 10 Tired Villa Residents crossed Mountains.
VXVRXRR	Vivian Xeroxed Very Rosy X-rays of Rowdy Randy.

C. Grammatical Mnemonics, Experiment 3

List 1

VXM	Virgins eXpect Miracles.
MVRXR	Many Varmints and Rats eXpect Rewards.
VXTVX	Virgins eXamining Terrified Varmints eXplode.
MTTTTVT	Many Terrified, Timid, Tense, Torontonian Varmints Tremble.
MTTVRXM	Many Terrified, Timid, Varmints and Rats eXpect Miracles.
VXRRRRR	Virgins eXpect Rewards, Reinforcements, Riches, Ridicule and Rejection.
VXTTTVT	Virgins eXamining Terrified, Timid, Tense Varmints Tremble.
MVRXTVX	Many Varmints and Rats eXamining Terrified Varmints eXplode.

List 2

MVX	Many Varmints eXplode.
MTTVT	Many Terrified, Timid Varmints Tremble.
VXRRM	Virgins eXpect Rewards and Reinforcements for Miracles.
MVRXRRM	Many Varmints and Rats eXpect Rewards and Reinforcements for Miracles.
MTVRXRR	Many Terrified Varmints and Rats eXpect Rewards and Reinforcements.
MTVRXVX	Many Terrified Varmints and Rats eXamining Varmints eXplode.
VXTVRXM	Virgins eXamining Terrified Varmints and Rats eXpect Miracles.
VXVRXRR	Virgins eXamining Varmints and Rats eXpect Rewards and Reinforcements.

D. Animal Names, Experiment 5

List	Training Item	Grammatical Item	Non-Grammatical Item
1	MXRVXT PYTHON	MXRMXT COBRA	MXRRXT JACKAL
2	MVXRM PYTHON	MVXRV COBRA	MVXRT JACKAL
1	VMTRRRR COYOTE	VMTRRRX ANACONDA	VMTRRRT WOLF
2	MVXRMVR COYOTE	MVXRMXR ANACONDA	MVXRMTR WOLF
1	MXTRRR MOUSE	VXTRRR GERBIL	TXTRRR GRIZZLY
2	VXVT MOUSE	VXVR GERBIL	VXVM GRIZZLY
1	VXVRMXT PANDA	VXVRVXT HAMSTER	VXVRTXT KOALA

2	MXRMXRV PANDA	MXRMVRV HAMSTER	MXRMTRV KOALA
1	VXVRVM DONKEY	VXVRVV BURRO	VXVRVT NEWT
2	MXRTVMT DONKEY	MXRTVMR BURRO	MXRTVMM NEWT
1	VMRVVVV LIZARD	VMRVVVM MULE	VMRVVVR SALAMANDER
2	VMT LIZARD	VMR MULE	VMM SALAMANDER
1	MXRTMVR BABOON	MXRTMXR CHIMPANZEE	MXRTMTR LOBSTER
2	MVXRVMR BABOON	MVXRVMT CHIMPANZEE	MVXRVMX LOBSTER
1	VMRMXTR CRAB	VMRVXTR GORILLA	VMRTXTR CRAYFISH
2	MVXRVVV CRAB	MVXRVVM GORILLA	MVXRVVR CRAYFISH
1	MXR PIG	MVR BOAR	MTR SALMON
2	MXRMXRM PIG	MXRMVRM BOAR	MXRMTRM SALMON
1	VMRVXVR PERCH	VMRVXVT HOG	VMRVXVX TROUT
2	VMRVMT PERCH	VMRVMR HOG	VMRVMM TROUT
1	MVRVM LEOPARD	MXRVM LION	MTRVM HORNET
2	MXRVVVM LEOPARD	MVRVVVM LION	MTRVVVM HORNET
1	VMRMVRV BEE	VMRMXRV TIGER	VMRMTRV WASP
2	VXVRMXR BEE	VXVRMVR TIGER	VXVRMTR WASP
1	VMRMVXR BEAVER	VMRMVXT OTTER	VMRMVXX PARROT
2	MXTRRRX BEAVER	VXTRRRX OTTER	TXTRRRX PARROT
1	MXRTVXT MACAW	MXRTMXT MUSKRAT	MXRTRXT TOUCAN
2	VXVTRRX MACAW	VXVTRRR MUSKRAT	VXVTRRM TOUCAN
1	MXRMVXR DEER	MXRMVXT ELK	MXRMVXX HIPPOPOTAMUS
2	MXTRRX DEER	VXTRRX ELK	TXTRRX HIPPOPOTAMUS
1	MVXTR RHINOCEROS	MVXTX MOOSE	MVXTT ELEPHANT
2	VMRTMXT RHINOCEROS	VMRTVXT MOOSE	VMRTTXT ELEPHANT

E. Unique Mnemonics, Experiment 6

List	Training Item	Grammatical Item	Non-Grammatical Item
1	MXRVXT Maniacs eXperience Rough, Vile, eXperimental Treatment.	MXRMXT Maniacs eXperience Rough, Mean, eXperimental Treatment.	MXRRXT Mary eXalts Red Raisins' eXtra Taste.
2	MVXRM Maniacs' Venturesome, eXperimental Rehabilitatiion (is) Mean.	MVXRV Maniacs' Venturesome, eXperimental Rehabilitation (is) Vile.	MVXRT Mary Vibrantly eXalts Raisins' Taste.
1	VMTRRRR Vibrant Mary Tastes Rum Rolled Red Raisins.	VMTRRRX Various Maniacs' Treatment Remains Rough, Ruthless, (and) eXperimental.	VMTRRRT Vibrant Mary Tastes Rum Rolled Raisins Tenderly.
2	MVXRMVR Mary Vibrantly eXalts Rum Moistened Vineyard Raisins.	MVXRMXR Maniacs Variously eXperience Rough, Mean, eXperimental Rehabilitation.	MVXRMTR Mary Vibrantly eXalts Rum Moistened, Tender Raisins.
1	MXTRRR Massive eXplosions Terrify Robins, Rabbits, (and) Rattlesnakes.	VXTRRR Violent eXplosions Terrify Robins Rabbits, (and) Rattlesnakes.	TXTRRR Teasingly eXecuted Trick Rollerskating Requires Rest.
2	VXVT Violent eXplosions - Varmints Terrified.	VXVR Violent eXplosions - Varmints Revolted.	VXVM Vivian eXhausts Vance Mischievously.
1	VXVRMXT Vivian eXhausts Vance Rollerskating, Mischievously eXecuting Tricks.	VXVRVXT Violent eXplosions Viciously Render Varmints eXceptionally Terrified.	VXVRTXT Vivian eXhausts Vance Rollerskating, Teasingly eXecuting Tricks.
2	MXRMXRV Mischievously eXecuted Rollerskating Moves eXhaust Resting Vance.	MXRMVRV Massive eXplosions Render Miserable Varmints Revolted Violently.	MXRMTRV Mischievously eXecuted Rollerskating Moves Tire Resting Vance.
1	VXVRVM Vandals eXamine Vaults, Removing Volumes (of) Mushrooms.	VXVRVV Vandals eXamine Vaults, Removing Volumes (of) Vegetables.	VXVRVT Virginia eXhibits Vases, Remarkable Violins, (and) Tulips.
2	MXRTVMT Masterfully eXamining Ripe Toadstools, Vandals Munch (and) Transplant.	MXRTVMR Masterfully eXamining Ripe Toadstools, Vandals Munch (and) Remove.	MXRTVMM Modelling (and) eXhibiting Remarkable Talent, Viginia Makes Mums.

1	VMRVVVV Viginia Models Remarkable Vases, Violins, Veils, (and) Violets.	VMRVVVM Vandals Masterfully Removing Volumes (of) Vegetables Venerate Mushrooms.	VMRVVVR Viginia Models Remarkable Vases, Violins, Veils, (and) Roses.
2	VMT Viginia Models Tulips.	VMR Vandals Munch Rhutabagas.	VMM Viginia Models Moms.
1	MXRTMVR Money eXperts Report Trouble Memorizing Volatile Rates.	MXRTMXR Money eXperts Report Trouble Memorizing eXcessive Rates.	MXRTMTR Momentarily eXcluding Reason, Thoughtless Matadors Temporarily Relax.
2	MVXRVM Money Viewing eXperts Report Volatile, Mixed Rates,	MVXRVMT Money Viewing eXperts Report Volatile, Mixed Tariffs	MVXRVMX Matadors Vulnerably eXcluding Reason Vaguely Malinger eXposed.
1	VMRMXTR Vague Matadors Relax Momentarily, eXposing Themselves Rashly.	VMRVXTR Viewers (of) Money Report Volatile, eXcessive Tariff Rates.	VMRTXTR Vague Matadors Relax Temporarily, eXposing Themselves Rashly.
1	MXR Monks eXcavate Rubies.	MVR Monks Value Rubies.	MTR Malicious Totalitarian Robots.
2	MXRMXRM Millions (of) eXpensive Rubies Mystify eXcavating Religious Monks.	MXRMVRM Millions (of) eXpensive Rubies Mystify Vowing Religious Monks.	MXRMTRM Malicious eXasperated Robots Mercilessly Terminate Radical Mercenaries.
1	VMRVXVR Very Malicious Robots Vapourize eXasperated Villainous Radicals.	VMRVXVT Vowing Monks' Rubies Vex; eXpensively, Vulgarly Tempting.	VMRVXVX Very Malicious Robots Vapourize eXasperated Villainous eXtremists.
2	VMRVMT Very Malicious Robots Vapourize Mercenary Totalitarians.	VMRVMR Vowing Monks' Rubies Vex (and) Mystify Religious.	VMRVMM Very Malicious Robots Vapourize Mercenary Militants.
1	MVRVM Muppets Vacate Restaurants Visibly Miffed.	MXRVM Muppets eXit Restaurants Visibly Miffed.	MTRVM Monkeys Trail Relatives (and) Vagrants Merrily.
2	MXRVVVM Muppets eXit Restaurants Vehement, Verbose, (and) Visibly Miffed.	MVRVVVM Muppets Vacate Restaurants Vehement, Verbose, (and) Visibly Miffed.	MTRVVVM Monkeys Trail Relatives, Vagrants, (and) Voyaging Vagabonds Merrily.
1	VMRMVRV Vagabond Monkeys Roam Merrily, Visting	VMRMXRV Visibly Miffed, Riled Muppets eXit	VMRMTRV Vagabond Monkeys Roam Merrily Trailing

	Relatives (and) Vagrants.	Restaurants Vehemently.	Relatives (and) Vagrants.
2	VXVRMXR Visiting eXiled Vagrants (and) Relatives, Monkeys eXplicitly Roam.	VXVRMVR Visibly eXiting Vehemently, Riled Muppets Vacate Restaurants.	VXVRMTR Visting eXiled Vagrants (and) Relatives, Monkeys' Trails Roam.
1	VMRMVXR Virgins (and) Maidens Rally Monthly Virtuously eXpecting Rewards.	VMRMVXT Virgins (and) Maidens Rally Monthly Virtuously eXpecting Tributes.	VMRMVXX Vigorously, Many Ranchers Mail Valentines (and) eXotic eXports.
2	MXTRRRX Maidens eXpect Tributes (and) Rewards, Regularly Rallying eXuberantly.	VXTRRRX Virgins eXpect Tributes (and) Rewards, Regularly Rallying eXuberantly.	TXTRRRX Trading eXotic Toffee, Really Romantic Ranchers eXport.
1	MXRTVXT Many eXotic Ranchers Trade Valentines (and) eXport Toffee.	MXRTMXT Monthly eXpecting Rewards (and) Tributes, Maidens eXuberantly Throng.	MXRTRXT Many eXotic Ranchers Trade Romance (and) eXport Toffee.
2	VXVTRRX Vigorously (and) eXotically, Valentine Trading Romantic Ranchers eXport.	VXVTRRR Virgins eXpect Virtuous Tributes (and) Rewards, Rallying Regularly.	VXVTRRM Vigorously (and) eXotically, Valentine Trading Romantic Ranchers Mail.
1	MXRMVXR Messy, eXtroverted, Rebellious Monsters Vocally eXpress Rigidity.	MXRMVXT Messy, eXtroverted, Rebellious Monsters Vocally eXpress Tension.	MXRMVXX Most X-rayed Rumps, Mr. Vokey eXpediently Xeroxes.
2	MXTRRX Monsters eXpress Tension Rigidly (and) Rebel eXtrovertly.	VXTRRX Vampires eXpress Tension Rigidly (and) Rebel eXtrovertly.	TXTRRX Teenage X-rayed Thighs (and) Rumps Regarded "X".
1	MVXTR Mr. Vokey X-rays Teenage Rumps.	MVXTX Monsters Vocally eXpress Tension eXtrovertly.	MVXTT Mr. Vokey X-Rays Teenage Thighs.
2	VMRTMXT Vokey Magnifies Rumps (of) Teenagers, Mostly X-raying Thighs.	VMRTVXT Vampire Monsters, Rebellious (and) Tacky, Vocally eXpress Tension.	VMRTTXT Vokey Magnifies Rumps (of) Teenagers, Torridly X-raying Thighs.

Appendix B

Following is the text of the instructions for the Interview phase used in Experiments 4, 5, and 6.

> We are interested in assessing how you went about determining whether an item did or did not obey the complex rules used to construct the items you were asked to learn. On the next pages are possible strategies that you may have used. For each strategy, we would like you to indicate your estimate of the proportion of items for which you used the strategy.
>
> For each strategy, please indicate your estimated use of the strategy by choosing a number from zero (meaning you never used that strategy) to one-hundred (meaning you always used that strategy). For example, if you think you used a given strategy for 60 percent of the items, then write the number 60 next to the strategy.
>
> Since you may have used more than one strategy for any given item, or used strategies not mentioned, there is no necessity to have your estimated proportions add up to 100.

After presentation of the instructions, the subjects received, in a random order, the following pairs of strategy statements. In each pair, the first statement refers to the acceptance of the items as obeying the rules, and the second refers to the rejection of items. The presentation of each pair of statements on a given page of the booklets was preceded by the instruction "Please estimate the proportion of items for which you used each strategy."

Specific similarity statements:

If a new item reminded me of one of the items I had been asked to learn, I said that the new item obeyed the complex rules.

If a new item did not remind me of one of the items I had been asked to learn, I said the new item did not obey the complex rules.

General similarity statements:

If a new item seemed to be familiar, even though I could not be sure which learning item it was similar to, I said that the new item obeyed the complex rules.

If a new item seemed unfamiliar (it seemed quite different from any of the items I had been asked to learn), I said that the new item did not obey the complex rules.

Specific rules statements:

From the learning items, I knew that only certain letters could occur in certain positions of the items. If a new item did not have wrong letters in these positions, I said that it obeyed the complex rules.

If a new item did have wrong letters in the letter positions I knew from the learning items, I said that it did not obey the complex rules.

General rules statements:

From the learning items, I knew that only certain combinations of letters obeyed the complex rules. If a new item had some of these letter combinations and did not have any new letter combinations, I said that it obeyed the complex rules.

If a new item did not have any of the letter combinations that I knew from the learning items, or had new letter combinations, I said that it did not obey the complex rules.

Appendix C: Analysis of Variance Tables

1.0 Experiment 1: Training Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	1.31	1	1.31	0.24	.6228
List	ST(E)	38.51	1	38.51	15.88	.0002
Trial	ST(E)	163.10	4	40.78	70.09	.0000
S(E)		333.46	62	5.38		
EL	SL(E)	0.45	1	0.45	0.19	.6676
ET	ST(E)	8.02	4	2.01	3.45	.0092
LT	SLT(E)	12.35	4	3.09	6.89	.0000
SL(E)		150.33	62	2.42		
ST(E)		144.28	248	0.58		
ELT	SLT(E)	6.70	4	1.67	3.74	.0057

1.1 Experiment 1: Recognition Phase

1.1.1 Analysis on Hits

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	0.27	2	0.14	0.05	.9555
S(E)		276.71	93	2.98		

1.1.2 Analysis on False-Positives

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	8.44	2	4.22	7.10	.0013
Gram	SG(E)	7.88	1	7.88	26.60	.0000
Dist	SD(E)	4.82	1	4.82	13.00	.0005
S(E)		55.26	93	0.59		
EG	SG(E)	0.05	2	0.03	0.09	.9132
ED	SD(E)	2.21	2	1.11	2.98	.0555
GD	SGD(E)	1.15	1	1.15	4.39	.0389
SG(E)		27.54	93	0.29		
SD(E)		34.45	93	0.37		
EGD	SGD(E)	1.99	2	1.00	3.81	.0258
SGD(E)		24.33	93	0.26		

1.2 Experiment 1: Transfer Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	71.24	2	35.62	0.96	.3858
Altype	SA(E)	105.29	1	105.29	25.64	.0000
Gram	SG(E)	46.02	1	46.02	8.83	.0038
Pass	SP(E)	4.06	1	4.06	5.82	.0178
Dist	SD(E)	475.82	1	475.82	172.34	.0000
S(E)		3442.90	93	37.02		

EA	SA(E)	5.50	2	2.75	0.67	.5141
EG	SG(E)	1.93	2	0.97	0.19	.8311
AG	SAG(E)	0.00	1	0.00	0.00	.9982
EP	SP(E)	1.41	2	0.71	1.01	.3670
AP	SAP(E)	1.21	1	1.21	1.30	.2564
GP	SGP(E)	0.05	1	0.05	0.07	.7873
ED	SD(E)	21.77	2	10.89	3.94	.0227
AD	SAD(E)	6.92	1	6.92	7.32	.0081
GD	SGD(e)	0.40	1	0.40	0.21	.6475
PD	SPD(E)	5.10	1	5.10	7.20	.0086
SA(E)		381.96	93	4.11		
SG(E)		484.70	93	5.21		
SP(E)		64.91	93	0.70		
SD(E)		256.76	93	2.76		
EAG	SAG(E)	1.11	2	0.56	0.20	.8202
EAP	SAP(E)	5.59	2	2.79	3.02	.0535
EGP	SGP(E)	3.91	2	1.96	2.73	.0701
AGP	SAGP(E)	0.15	1	0.15	0.21	.6493
EAD	SAD(E)	0.49	2	0.25	0.26	.7719
EGD	SGD(E)	2.05	2	1.02	0.53	.5884
AGD	SAGD(E)	2.26	1	2.26	2.37	.1268
EPD	SPD(E)	0.47	2	0.23	0.33	.7192
APD	SAPD(E)	6.70	1	6.70	6.21	.0145
GPD	SGPD(E)	0.12	1	0.12	0.13	.7192
SAG(E)		260.38	93	2.80		
SAP(E)		85.98	93	0.92		
SGP(E)		66.55	93	0.72		
SAD(E)		87.97	93	0.95		
SGD(E)		178.37	93	1.92		
SPD(E)		65.87	93	0.71		
EAGP	SAGP(E)	0.03	2	0.01	0.02	.9822
EAGD	SAGD(E)	0.19	2	0.10	0.10	.9041
EAPD	SAPD(E)	0.13	2	0.06	0.06	.9432
EGPD	SGPD(E)	0.11	2	0.05	0.06	.9436
AGPD	SAGPD(E)	0.01	1	0.01	0.01	.9184
SAGP(E)		65.20	93	0.70		
SAGD(E)		88.50	93	0.95		
SAPD(E)		100.45	93	1.08		
SGPD(E)		84.91	93	0.91		
EAGPD	SAGPD(E)	0.86	2	0.43	0.57	.5690
SAGPD(E)		70.48	93	0.76		

2.0 Experiment 2: Training Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	32.40	1	32.40	4.57	.0364
List	SL(E)	14.40	1	14.40	8.40	.0052
Trial	ST(E)	129.15	4	32.29	65.52	.0000
S(E)		439.19	62	7.08		
EL	SL(E)	0.16	1	0.16	0.09	.7637

ET	ST(E)	5.24	4	1.13	2.66	.0334
LT	SLT(E)	9.24	4	2.31	4.64	.0013
SL(E)		106.24	62	1.71		
ST(E)		122.21	248	0.49		
ELT	SLT(E)	3.36	4	0.84	1.69	.1540
SLT(E)		123.60	248	0.50		

2.1 Experiment 2: Recognition Phase
2.1.1 Analysis on Hits

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	11.39	1	11.39	3.60	.0625
S(E)		196.21	62	3.16		

2.1.2 Analysis on False-Positives

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	4.52	1	4.52	11.80	.0011
Gram	SG(E)	1.67	1	1.67	6.25	.0151
Dist	SD(E)	3.21	1	3.21	10.62	.0018
S(E)		23.74	62	0.38		
EG	SG(E)	0.06	1	0.06	0.23	.6301
ED	SD(E)	1.00	1	1.00	3.31	.0737
GD	SGD(E)	0.17	1	0.17	0.62	.4348
SG(E)		16.55	62	0.27		
SD(E)		18.73	62	0.30		
EGD	SGD(E)	0.02	1	0.02	0.06	.8144
SGD(E)		17.42	62	0.28		

2.2 Experiment 2: Transfer Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	36.93	1	36.93	0.77	.3844
Altype	SA(E)	38.79	1	38.79	4.69	.0343
Gram	SG(E)	0.12	1	0.12	0.03	.8646
Pass	SP(E)	0.27	1	0.27	0.24	.6264
Dist	SD(E)	344.94	1	344.94	106.15	.0000
S(E)		2983.87	62	48.13		
EA	SA(E)	8.35	1	8.35	1.01	.3192
EG	SG(E)	5.71	1	5.71	1.34	.2511
AG	SAG(E)	5.84	1	5.84	2.28	.1365
EP	SP(E)	0.59	1	0.59	0.52	.4727
AP	SAP(E)	0.63	1	0.63	0.49	.4874
GP	SGP(E)	0.21	1	0.21	0.22	.6386
ED	SD(E)	23.76	1	23.76	7.31	.0088
AD	SAD(E)	0.75	1	0.75	0.48	.4890
GD	SGD(E)	3.03	1	3.03	1.87	.1770
PD	SPD(E)	3.59	1	3.59	5.95	.0176
SA(E)		513.23	62	8.28		

SG(E)		263.71	62	4.25		
SP(E)		70.10	62	1.13		
SD(E)		201.40	62	3.25		
EAG	SAG(E)	1.38	1	1.38	0.54	.4666
EAP	SAP(E)	2.39	1	2.39	1.85	.1783
EGP	SGP(E)	0.03	1	0.03	0.03	.8619
AGP	SAGP(E)	0.53	1	0.53	0.83	.3672
EAD	SAD(E)	0.03	1	0.03	0.02	.8809
EGD	SGD(E)	0.47	1	0.47	0.29	.5917
AGD	SAGD(E)	5.30	1	5.30	5.06	.0280
EPD	SPD(E)	1.16	1	1.16	1.92	.1705
APD	SAPD(E)	1.17	1	1.17	1.16	.2857
GPD	SGPD(E)	2.50	1	2.50	3.67	.0601
SAG(E)		159.00	62	2.56		
SAP(E)		79.76	62	1.29		
SGP(E)		58.28	62	0.94		
SAD(E)		95.75	62	1.54		
SGD(E)		100.63	62	1.62		
SPD(E)		37.45	62	0.60		
EAGP	SAGP(E)	0.01	1	0.01	0.02	.8947
EAGD	SAGD(E)	0.14	1	0.14	0.13	.7150
EAPD	SAPD(E)	0.07	1	0.07	0.07	.7928
EGPD	SGPD(E)	0.13	1	0.13	0.19	.6648
AGPD	SAGPD(E)	0.01	1	0.01	0.01	.9120
SAGP(E)		40.03	62	0.65		
SAGD(E)		64.94	62	1.05		
SAPD(E)		62.66	62	1.01		
SGPD(E)		42.35	62	0.68		
EAGPD	SAGPD(E)	0.04	1	0.04	0.05	.8309
SAGPD(E)		55.38	62	0.89		

3.0 Experiment 3: Training Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
List	SL	0.38	1	0.38	0.14	.7085
Trial	ST	58.11	4	14.53	16.13	.0000
SL		82.32	31	2.66		
ST		111.69	124	0.90		
LT	SLT	3.36	4	0.84	1.23	.3007
SLT		84.44	124	0.68		

3.1 Experiment 3: Recognition Phase
3.1.1 Analysis on Hits

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	1.00	1	1.00	0.52	.4727
S(E)		118.75	62	1.92		

3.1.2 Analysis on False-Positives

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	4.69	1	4.69	4.60	.0358
Gram	SG(E)	3.06	1	3.06	4.60	.0359
Dist	SD(E)	3.06	1	3.06	4.06	.0482
S(E)		63.23	62	1.02		
EG	SG(E)	0.92	1	0.92	1.38	.2448
ED	SD(E)	1.46	1	1.46	1.94	.1691
GD	SGD(E)	1.27	1	1.27	2.78	.1005
SG(E)		41.38	62	0.67		
SD(E)		46.76	62	0.75		
EGD	SGD(E)	0.11	1	0.11	0.24	.6231
SGD(E)		28.23	62	0.46		

3.2 Experiment 3: Transfer Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Encode	S(E)	45.91	1	45.91	2.54	.1164
Antype	SA(E)	0.77	1	0.77	0.34	.5632
Gram	SG(E)	22.54	1	22.54	4.75	.0332
Pass	SP(E)	1.90	1	1.90	2.08	.1544
Dist	SD(E)	188.85	1	188.85	66.20	.0000
S(E)		1122.40	62	18.10		
EA	SA(E)	3.19	1	3.19	1.40	.2408
EG	SG(E)	77.20	1	77.20	16.25	.0002
AG	SAG(E)	32.37	1	32.37	11.64	.0011
EP	SP(E)	0.70	1	0.70	0.77	.3828
AP	SAP(E)	0.03	1	0.03	0.05	.8170
GP	SGP(E)	0.48	1	0.48	0.56	.4568
ED	SD(E)	3.51	1	3.51	1.23	.2715
AD	SAD(E)	0.46	1	0.46	0.29	.5928
GD	SGD(E)	0.86	1	0.86	0.50	.4808
PD	SPD(E)	0.07	1	0.07	0.08	.7731
SA(E)		141.11	62	2.28		
SG(E)		294.55	62	4.75		
SP(E)		56.55	62	0.91		
SD(E)		176.87	62	2.85		
EAG	SAG(E)	0.95	1	0.95	0.34	.5612
EAP	SAP(E)	1.33	1	1.33	2.08	.1539
EGP	SGP(E)	1.82	1	1.82	2.14	.1483
AGP	SAGP(E)	1.57	1	1.57	2.19	.1440
EAD	SAD(E)	11.82		11.82	7.39	.0085
EGD	SGD(E)	6.57	1	6.57	3.82	.0550
AGD	SAGD(E)	0.06	1	0.06	0.05	.8155
EPD	SPD(E)	1.69	1	1.69	1.90	.1727
APD	SAPD(E)	0.50	1	0.50	0.37	.5436
GPD	SGPD(E)	2.13	1	2.13	1.80	.1850
SAG(E)		172.37	62	2.78		
SAP(E)		39.48	62	0.64		
SGP(E)		52.58	62	0.85		
SAD(E)		99.12	62	1.60		

SGD(E)		106.51	62	1.72		
SPD(E)		55.09	62	0.89		
EAGP	SAGP(E)	0.03	1	0.03	0.04	.8501
EAGD	SAGD(E)	2.25	1	2.25	2.11	.1515
EAPD	SAPD(E)	0.24	1	0.24	0.18	.6763
EGPD	SGPD(E)	0.01	1	0.01	0.01	.9177
AGPD	SAGPD(E)	0.00	1	0.00	0.00	.9787
SAGP(E)		44.38	62	0.72		
SAGD(E)		66.23	62	1.07		
SAPD(E)		83.90	62	1.35		
SGPD(E)		73.60	62	1.19		
EAGPD	SAGPD(E)	0.33	1	0.33	0.45	.5034
SAGPD(E)		45.16	62	0.73		

4.0 Experiment 4: Training Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Task	GS(T)	5.49	1	5.49	1.23	.2698
List	SL(T)	17.43	3	5.81	4.36	.0049
Trial	S(T)	81.56	3	27.19	7.13	.0002
S(T)		472.77	124	3.81		
GL	GSL(T)	1.58	3	0.53	0.62	.5996
GT	GS(T)	1.78	3	0.59	0.13	.9402
LT	SL(T)	3.92	9	0.44	0.33	.9661
GS(T)		554.35	124	4.47		
SL(T)		495.53	372	1.33		
GLT	GSL(T)	8.34	9	0.93	1.10	.3628
GSL(T)		313.95	372	0.84		

4.1 Experiment 4: Recognition Phase

4.1.1 Analysis on Hits

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Task	S(T)	3.06	1	3.06	0.65	.4222
S(T)		290.88	62	4.69		

4.1.2 Analysis on False-Positives: Grammatical Responses

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Task	S(T)	2.25	1	2.25	0.36	.5534
Dist	SD(T)	23.77	1	23.77	15.31	.0002
Gram	SG(T)	132.25	1	132.25	75.44	.0000
S(T)		392.75	62	6.33		
TD	SD(T)	0.02	1	0.02	0.01	.9204
TG	SG(T)	0.06	1	0.06	0.04	.8509
DG	SDG(T)	0.77	1	0.77	0.77	.3844
SD(T)		96.22	62	1.55		
SG(T)		108.69	62	1.75		

TDG	SDG(T)	0.39	1	0.39	0.39	.5338
SDG(T)		61.84	62	1.00		

4.1.3 Analysis on False-Positives: Mean Response

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Task	S(T)	5.94	1	5.94	0.04	.8375
Dist	SD(T)	299.72	1	299.72	16.55	.0001
Gram	SG(T)	2156.40	1	2156.40	102.35	.0000
S(T)		8679.80	62	140.00		
TD	SD(T)	6.57	1	6.57	0.36	.5493
TG	SG(T)	0.47	1	0.47	0.02	.8814
DG	SDG(T)	26.91	1	26.91	1.94	.1683
SD(T)		1123.00	62	18.11		
SG(T)		1306.30	62	21.07		
TDG	SDG(T)	12.69	1	12.69	0.92	.3421
SDG(T)		85.87	62	13.85		

4.2 Experiment 4: Transfer Phase
4.2.1 Analysis on Grammatical Responses

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Task	S(T)	7.39	1	7.39	0.34	.5630
Pass	SP(T)	19.97	1	19.97	7.25	.0091
Dist	SD(T)	205.74	1	205.74	78.92	.0000
Rec	SR(T)	0.02	1	0.02	0.03	.8615
Gram	SG(T)	277.43	1	277.43	110.46	.0000
S(T)		1355.00	62	21.86		
TP	SP(T)	0.01	1	0.01	0.00	.9551
TD	SD(T)	21.68	1	21.68	8.32	.0054
PD	SPD(T)	0.52	1	0.52	0.43	.5122
TR	SR(T)	0.00	1	0.00	0.00	.9722
PR	SPR(T)	1.06	1	1.06	0.98	.3250
DR	SDR(T)	0.52	1	0.52	0.34	.5623
TG	SG(T)	2.16	1	2.16	0.86	.3576
PG	SPG(T)	1.06	1	1.06	0.68	.4142
DG	SDG(T)	0.43	1	0.43	0.29	.5918
RG	SRG(T)	0.01	1	0.01	0.00	.9540
SP(T)		170.83	62	2.76		
SD(T)		161.64	62	2.61		
SR(T)		49.29	62	0.79		
SG(T)		155.72	62	2.51		
TPD	SPD(T)	0.08	1	0.08	0.07	.7973
TPR	SPR(T)	0.02	1	0.02	0.02	.8810
TDR	SDR(T)	0.94	1	0.94	0.62	.4353
PDR	SPDR(T)	0.82	1	0.82	0.98	.3264
TPG	SPG(T)	0.43	1	0.43	0.27	.6027
TDG	SDG(T)	1.98	1	1.98	1.33	.2525
PDG	SPDG(T)	1.06	1	1.06	0.81	.3709
TRG	SRG(T)	15.75	1	15.75	6.02	.0170

PRG	SPRG(T)	0.00	1	0.00	0.00	.9754
DRG	SDRG(T)	42.66	1	42.66	17.12	.0001
SPD(T)		73.72	62	1.19		
SPR(T)		66.98	62	1.08		
SDR(T)		94.36	62	1.52		
SPG(T)		97.57	62	1.57		
SDG(T)		91.90	62	1.48		
SRG(T)		162.30	62	2.62		
TPDR	SPDR(T)	1.20	1	1.20	1.43	.2371
TPDG	SPDG(T)	2.34	1	2.34	1.79	.1857
TPRG	SPRG(T)	1.81	1	1.81	1.78	.1874
TDRG	SDRG(T)	7.39	1	7.39	2.97	.0908
PDRG	SPDRG(T)	1.98	1	1.98	1.27	.2639
SPDR(T)		52.05	62	0.84		
SPDG(T)		81.15	62	1.31		
SPRG(T)		63.01	62	1.02		
SDRG(T)		154.51	62	2.49		
TPDRG	SPDRG(T)	3.40	1	3.40	2.19	.1444
SPDRG(T)		96.44	62	1.56		

4.2.2 Analysis on Mean Response

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Task	S(T)	78.21	1	78.21	0.22	.6419
Pass	SP(T)	379.03	1	379.03	10.55	.0019
Dist	SD(T)	2926.10	1	2926.10	71.52	.0000
Rec	SR(T)	0.94	1	0.94	0.09	.7648
Gram	SG(T)	4180.40	1	4180.40	155.11	.0000
S(T)		22208.00	62	358.19		
TP	SP(T)	2.54	1	2.54	0.07	.7912
TD	SD(T)	216.64	1	216.64	5.29	.0248
PD	SPD(T)	39.45	1	39.45	3.61	.0620
TR	SR(T)	1.20	1	1.20	0.12	.7355
PR	SPR(T)	4.65	1	4.65	0.53	.4688
DR	SDR(T)	43.48	1	43.48	1.84	.1793
TG	SG(T)	3.17	1	3.17	0.12	.7327
PG	SPG(T)	4.38	1	4.38	0.26	.6144
DG	SDG(T)	21.10	1	21.10	1.37	.2461
RG	SRG(T)	0.71	1	0.71	0.02	.8787
SP(T)		2226.70	62	35.92		
SD(T)		2536.80	62	40.92		
SR(T)		644.18	62	10.39		
SG(T)		1671.00	62	26.95		
TPD	SPD(T)	-0.22	1	0.22	0.02	.8876
TPR	SPR(T)	0.71	1	0.71	0.08	.7764
TDR	SDR(T)	0.02	1	0.02	0.00	.9744
PDR	SPDR(T)	32.70	1	32.70	2.93	.0920
TPG	SPG(T)	3.63	1	3.63	0.21	.6464
TDG	SDG(T)	58.62	1	58.62	3.81	.0555
PDG	SPDG(T)	37.90	1	37.90	2.92	.0927
TRG	SRG(T)	160.18	1	160.18	5.28	.0249

PRG	SPRG(T)	0.01	1	0.01	0.00	.9780
DRG	SDRG(T)	974.61	1	974.61	36.00	.0000
SPD(T)		676.89	62	10.92		
SPR(T)		542.45	62	8.75		
SDR(T)		1461.10	62	23.57		
SPG(T)		1060.00	62	17.10		
SDG(T)		954.09	62	15.39		
SRG(T)		1880.70	62	30.33		
TPDR	SPDR(T)	21.10	1	21.10	1.89	.1742
TPDG	SPDG(T)	8.81	1	8.81	0.68	.4133
TPRG	SPRG(T)	45.14	1	45.14	3.95	.0514
TDRG	SDRG(T)	18.87	1	18.87	0.70	.4070
PDRG	SPDRG(T)	6.41	1	6.41	0.35	.5535
SPDR(T)		692.26	62	11.17		
SPDG(T)		805.60	62	12.99		
SPRG(T)		709.41	62	11.44		
SDRG(T)		1678.30	62	27.07		
TPDRG	SPDRG(T)	23.31	1	25.31	1.40	.2408
SPDRG(T)		1119.10	62	18.05		

4.3 Experiment 4: Position of Change

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Task	S(T)	0.48	1	0.48	0.00	.9478
Gram	S(T)	1757.45	1	1757.45	82.03	.0000
Position	SP(T)	1404.71	2	702.36	15.17	.0000
S(T)		6943.15	62	111.99		
TG	SG(T)	6.68	1	6.68	0.31	.5786
TP	SP(T)	156.74	2	78.37	1.69	.1882
GP	SGP(T)	52.02	2	26.01	1.82	.1662
SG(T)		1328.25	62	21.42		
SP(T)		5739.44	124	46.29		
TGP	SGP(T)	16.77	2	8.39	0.59	.5575
SGP(T)		1771.34	124	14.28		

4.4 Experiment 4: Interview

Source	Error Term	Sum of Squares	BF	Mean Square	F	Prob.
Subject		64309.00	31	2074.48		
Type	ST	18428.06	1	18428.06	13.14	.0010
Gen-Spec.	SG	833.77	1	833.77	1.27	.2684
Pos-Neg.	SP	1691.27	1	1691.27	2.11	.1567
ST		43464.94	31	1402.09		
SG		20349.23	31	656.43		
TG	STG	1795.64	1	1795.64	3.40	.0747
SP		24884.23	31	802.72		
TP	STP	356.27	1	356.27	1.04	.3158
GP	SGP	529.00	1	529.00	0.94	.3390
STG		16367.36	31	527.98		
STP		10624.23	31	342.72		

SGP		17391.50	31	561.02		
TGP	STGP	3660.25	1	3660.25	13.33	.0010
STGP		8515.25	31	274.69		

5.0 Experiment 5: Training Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		314.47	31	10.14		
List	SL	52.31	3	17.44	6.17	.0007
Trial	ST	71.61	3	23.87	32.70	.0000
SL		262.69	93	2.82		
ST		67.89	93	0.73		
LT	SLT	2.86	9	0.32	0.43	.9184
SLT		206.14	279	0.74		

5.1 Experiment 5: Recognition Phase

5.1.1 Analysis on False-Positives

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		132.21	31	4.26		
Dist	SD	0.19	1	0.19	0.37	.5461
Appin	SA	0.19	1	0.19	0.38	.5398
Gram	SG	14.54	1	14.54	17.96	.0002
SD		15.93	31	0.51		
SA		15.43	31	0.50		
DA	SDA	0.32	1	0.32	1.05	.3126
SG		25.09	31	0.81		
DG	SDG	1.72	1	1.72	3.84	.0590
AG	SAG	0.10	1	0.10	0.34	.5667
SDA		9.31	31	0.30		
SDG		13.90	31	0.45		
SAG		9.03	31	0.29		
DAG	SDAG	0.32	1	0.32	0.83	.3691
SDAG		11.81	31	0.38		

5.1.2 Analysis on Mean Response

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		2498.36	31	80.59		
Dist	SD	5.06	1	5.06	0.59	.4494
Appin	SA	31.64	1	31.64	5.36	.0273
Gram	SG	182.25	1	182.25	18.24	.0002
SD		267.44	31	8.63		
SA		182.86	31	5.90		
DA	SDA	7.56	1	7.56	2.07	.1606
SG		309.75	31	9.99		
DG	SDG	15.02	1	15.02	3.93	.0564
AG	SAG	4.00	1	4.00	1.05	.3143
SDA		113.44	31	3.66		

SDG		118.48	31	3.82		
SAG		118.50	31	3.82		
DAG	SDAG	19.14	1	19.14	3.67	.0648
SDAG		161.86	31	5.22		

5.2 Experiment 5: Transfer Phase
5.2.1 Analysis on Grammatical Responses

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		386.15	31	12.46		
Pass	SP	4.79	1	4.79	1.96	.1719
Dist	SD	26.26	1	26.26	20.24	.0001
Appin	SA	0.39	1	0.39	0.29	.5940
Rec	SR	19.69	1	19.69	25.37	.0000
Gram	SG	39.06	1	39.06	17.82	.0002
SP		75.84	31	2.45		
SD		40.23	31	1.30		
PD	SPD	0.14	1	0.14	0.15	.7038
SA		41.73	31	1.35		
PA	SPA	0.06	1	0.06	0.13	.7201
DA	SDA	2.44	1	2.44	2.93	.0968
SR		24.06	31	0.78		
PR	SPR	0.66	1	0.66	1.22	.2788
DR	SDR	0.02	1	0.02	0.02	.8969
AR	SAR	7.56	1	7.56	6.10	.0192
SG		67.94	31	2.19		
PG	SPG	0.00	1	0.00	0.00	1.0000
DG	SDG	1.72	1	1.72	3.03	.0919
AG	SAG	0.19	1	0.19	0.24	.6282
RG	SRF	4.00	1	4.00	3.23	.0820
SPD		29.61	31	0.96		
SPA		14.81	31	0.48		
SDA		25.81	31	0.83		
PDA	SPDA	0.66	1	0.66	1.45	.2373
SPR		16.84	31	0.54		
SDR		28.36	31	0.91		
PDR	SPDR	0.14	1	0.14	0.23	.6374
SAR		38.44	31	1.24		
PAR	SPAR	0.39	1	0.39	0.80	.3776
DAR	SDAR	1.13	1	1.13	1.61	.2140
SPG		25.50	31	0.82		
SDG		17.65	31	0.57		
PDG	SPDG	1.13	1	1.13	4.67	.0386
SAG		24.81	31	0.80		
PAG	SPAG	0.04	1	0.04	0.04	.8350
DAG	SDAG	6.25	1	6.25	6.83	.0137
SRG		38.38	31	1.24		
PRG	SPRG	1.56	1	1.56	2.54	.1211
DRG	SDRG	9.38	1	9.38	13.14	.0010
ARG	SARG	0.19	1	0.19	0.25	.6221
SPDA		14.09	31	0.45		

SPDR		19.23	31	0.62		
SPAR		15.11	31	0.49		
SDAR		21.75	31	0.70		
PDAR	SPDAR	0.10	1	0.10	0.13	.7164
SPDG		7.50	31	0.24		
SPAG		24.71	31	0.80		
SDAG		28.38	31	0.92		
PDAG	SPDAG	0.39	1	0.39	0.77	.3871
SPRG		19.06	31	0.61		
SDRG		22.12	31	0.71		
PDRG	SPDRG	0.47	1	0.47	0.79	.3807
SARG		23.93	31	0.77		
PARG	SPARG	0.88	1	0.88	1.12	.2973
DARG	SDARG	0.14	1	0.14	0.34	.5646
SPDAR		22.53	31	0.73		
SPDAG		15.73	31	0.51		
SPDRG		18.53	31	0.60		
SPARG		24.25	31	0.78		
SDARG		12.86	31	0.41		
PDARG	SPDARG	0.06	1	0.06	0.11	.7393
SPDARG		17.19	31	0.55		

5.2.2 Analysis on Mean Response

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		6519.60	31	210.31		
Pass	SP	98.75	1	98.75	2.96	.0951
Dist	SD	268.14	1	268.14	22.35	.0000
Appin	SA	6.25	1	6.25	0.60	.4455
Rec	SR	246.10	1	246.10	27.05	.0000
Gram	SG	484.00	1	484.00	20.19	.0001
SP		1033.10	31	33.33		
SD		371.98	31	12.00		
PD	SPD	3.06	1	3.06	0.70	.4093
SA		324.38	31	10.46		
PA	SPA	9.77	1	9.77	1.79	.1908
DA	SDA	11.82	1	11.81	1.62	.2124
SR		282.03	31	9.10		
PR	SPR	11.82	1	11.82	2.21	.1468
DR	SDR	1.89	1	1.89	0.26	.6125
AR	SAR	40.64	1	40.64	2.67	.1121
SG		743.25	31	23.98		
PG	SPG	3.06	1	3.06	0.63	.4321
DG	SDG	0.32	1	0.32	0.05	.8235
AG	SAG	7.91	1	7.91	0.86	.3603
RG	SRG	5.06	1	5.06	0.52	.4744
SPD		135.69	31	4.38		
SPA		169.23	31	5.46		
SDA		225.93	31	7.29		
PDA	SPDA	0.66	1	0.66	0.21	.6491
SPR		165.43	31	5.34		

SDR		223.86	31	7.22		
PDR	SPDR	6.89	1	6.89	1.72	.1998
SAR		471.11	31	15.20		
PAR	SPAR	1.27	1	1.27	0.42	.5194
DAR	SDAR	4.79	1	4.79	0.55	.4652
SPG		149.81	31	4.83		
SDG		193.81	31	6.25		
PDG	SPDG	0.04	1	0.04	0.01	.9840
SAG		284.46	31	9.18		
PAG	SPAG	0.66	1	0.66	0.16	.6892
DAG	SDAG	78.77	1	78.77	6.88	.0134
SRG		299.31	31	9.66		
PRG	SPRG	6.89	1	6.89	1.31	.2619
DRG	SDRG	36.75	1	36.75	6.20	.0183
ARG	SARG	1.72	1	1.72	0.20	.6575
SPDA		96.97	31	3.13		
SPDR		124.48	31	4.02		
SPAR		92.36	31	2.98		
SDAR		271.34	31	8.75		
PDAR	SPDAR	8.63	1	8.63	0.95	.3375
SPDG		73.72	31	2.38		
SPAG		125.59	31	4.05		
SDAG		354.73	31	11.44		
PDAG	SPDAG	8.27	1	8.27	2.08	.1591
SPRG		163.61	31	5.28		
SDRG		183.75	31	5.93		
PDRG	SPDRG	15.50	1	15.50	3.61	.0668
SARG		266.53	31	8.60		
PARG	SPARG	5.94	1	5.94	0.99	.3263
DARG	SDARG	1.56	1	1.56	0.25	.6182
SPDAR		281.87	31	9.09		
SPDAG		123.11	31	3.97		
SPDRG		133.12	31	4.29		
SPARG		185.18	31	5.97		
SDARG		191.06	31	6.16		
PDARG	SPDARG	3.52	1	3.52	0.90	.3499
SPDARG		120.98	31	3.90		

5.3 Experiment 5: Position of Change

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		4589.95	31	148.06		
Gram	SG	639.58	1	639.58	19.06	.0001
Position	SP	472.04	2	236.02	6.47	.0028
SG		1040.12	31	33.55		
SP		2261.24	62	36.472		
GP	SGP	97.52	2	48.76	3.13	.0505
SGP		964.41	62	15.56		

5.4 Experiment 5: Interview

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		93518.11	31	3016.71		
Type	ST	16608.77	1	16608.77	10.27	.0031
Gen-Spec.	SG	280.56	1	280.56	0.50	.4849
Pos-Neg.	SP	2013.77	1	2013.77	2.82	.1032
ST		50138.98	31	1617.29		
SG		17407.19	31	561.52		
TG	STG	306.25	1	306.25	0.76	.3897
SP		22143.98	31	714.32		
TP	STP	2956.64	1	2956.64	13.78	.0008
GP	SGP	6.25	1	6.25	0.03	.8571
STG		12476.00	31	402.45		
STP		6650.61	31	214.54		
SGP		5866.00	31	189.23		
TGP	STGP	2304.00	1	2304.00	7.09	.0122
STGP		10078.75	31	325.75		

6.0 Experiment 6: Training Phase

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		504.55	31	16.28		
List	SL	9.84	3	3.28	1.92	.1319
Trial	ST	80.46	3	26.82	31.86	.0000
SL		158.91	93	1.71		
ST		78.29	93	0.84		
LT	SLT	6.13	9	0.68	1.58	.1220
SLT		120.62	279	0.43		

6.1 Experiment 6: Recognition Phase

6.1.1 Analysis on False-Positives

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		36.44	31	1.18		
Dist	SD	3.06	1	3.06	9.55	.0042
Appin	SA	2.25	1	2.25	5.94	.0208
Gram	SG	2.25	1	2.25	6.39	.0160
SD		9.94	31	0.32		
SA		11.75	31	0.38		
DA	SDA	1.56	1	1.56	5.74	.0228
SG		10.75	31	0.35		
DG	SDG	0.06	1	0.06	0.24	.6247
AG	SAG	0.00	1	0.00	0.00	1.0000
SDA		8.44	31	0.27		
SDG		7.94	31	0.26		
SAG		6.00	31	0.19		
DAG	SDAG	1.00	1	1.00	3.44	.0730
SDAG		9.00	31	0.29		

5.1.2 Analysis on Mean Response

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		1647.90	31	53.16		
Dist	SD	38.29	1	38.29	5.48	.0258
Appin	SA	43.07	1	43.07	6.96	.0129
Gram	SG	30.94	1	30.94	7.93	.0084
SD		216.59	31	6.99		
SA		191.81	31	6.18		
DA	SDA	9.38	1	9.38	1.65	.2082
SG		120.93	31	3.90		
DG	SDG	6.57	1	6.57	1.63	.2111
AG	SAG	0.00	1	0.00	0.00	.9739
SDA		176.00	31	5.68		
SDG		124.81	31	4.03		
SAG		111.37	31	3.59		
DAG	SDAG	8.63	1	8.63	1.37	.2507
SDAG		192.25	31	6.30		

6.2 Experiment 6: Transfer Phase
6.2.1 Analysis on Grammatical Responses

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		547.40	31	17.66		
Pass	SP	0.06	1	0.06	0.05	.8293
Dist	SD	169.00	1	169.00	67.11	.0000
Appin	SA	69.10	1	69.10	48.45	.0000
Rec	SR	0.39	1	0.39	0.45	.5054
Gram	SG	8.63	1	8.63	6.78	.0140
SP		41.00	31	1.32		
SD		78.06	31	2.52		
PD	SPD	1.41	1	1.41	2.27	.1422
SA		44.22	31	1.43		
PA	SPA	0.06	1	0.06	0.08	.7742
DA	SDA	81.00	1	81.00	32.43	.0000
SR		26.67	31	0.86		
PR	SPR	3.29	1	3.29	8.56	.0064
DR	SDR	6.57	1	6.57	6.54	.0157
AR	SAR	16.00	1	16.00	10.86	.0025
SG		39.43	31	1.27		
PG	SPG	0.39	1	0.39	0.38	.5401
DG	SDG	1.00	1	1.00	2.90	.0986
AG	SAG	0.88	1	0.88	1.45	.2379
RG	SRG	9.00	1	9.00	8.47	.0066
SPD		19.28	31	0.62		
SPA		23.13	31	0.75		
SDA		77.44	31	2.50		
PDA	SPDA	0.10	1	0.10	0.22	.6448
SPR		11.90	31	0.38		
SDR		31.12	31	1.00		
PDR	SPDR	0.02	1	0.02	0.05	.8185

SAR		45.69	31	1.47		
PAR	SPAR	0.00	1	0.00	0.01	.9260
DAR	SDAR	0.04	1	0.04	0.03	.8560
SPG		31.55	31	1.02		
SDG		10.69	31	0.34		
PDG	SPDG	0.47	1	0.47	1.14	.2937
SAG		18.81	31	0.61		
PAG	SPAG	1.00	1	1.00	1.44	.2396
DAG	SDAG	0.39	1	0.39	0.55	.4655
SRG		32.94	31	1.06		
PRG	SPRG	1.41	1	1.41	1.85	.1838
DRG	SDRG	0.00	1	0.00	0.00	.9473
ARG	SARG	2.25	1	2.25	5.66	.0236
SPDA		13.97	31	0.45		
SPDR		9.05	31	0.29		
SPAR		13.81	31	0.45		
SDAR		32.53	31	1.05		
PDAR	SPDAR	0.56	1	0.56	0.95	.3375
SPDG		12.84	31	0.41		
SPAG		21.56	31	0.70		
SDAG		22.17	31	0.72		
PDAG	SPDAG	0.66	1	0.66	1.05	.3139
SPRG		23.65	31	0.76		
SDRG		27.31	31	0.88		
PDRG	SPDRG	0.77	1	0.77	1.59	.2167
SARG		12.31	31	0.40		
PARG	SPARG	0.10	1	0.10	0.26	.6129
DARG	SDARG	0.04	1	0.04	0.05	.8176
SPDAR		18.38	31	0.59		
SPDAG		19.53	31	0.63		
SPDRG		14.92	31	0.48		
SPARG		11.59	31	0.37		
SDARG		21.15	31	0.65		
PDARG	SPDARG	0.77	1	0.77	1.00	.3257
SPDARG		23.80	31	0.77		

6.2.2 Analysis on Mean Response

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		6590.60	31	212.60		
Pass	SP	4.13	1	4.13	0.30	.5908
Dist	SD	1814.20	1	1814.20	43.86	.0000
Appin	SA	1095.20	1	1095.20	51.66	.0000
Rec	SR	27.24	1	27.24	2.58	.1184
Gram	SG	81.56	1	81.56	8.83	.0057
SP		433.28	31	13.98		
SD		1282.20	31	41.36		
PD	SPD	0.17	1	0.17	0.04	.8495
SA		657.21	31	21.20		
PA	SPA	1.20	1	1.20	0.33	.5678
DA	SDA	1116.00	1	1116.00	36.67	.0000

SR		327.30	31	10.56		
PR	SPR	22.27	1	22.27	6.27	.0177
DR	SDR	49.44	1	49.44	5.10	.0311
AR	SAR	178.06	1	178.06	11.53	.0019
SG		286.34	31	9.24		
PG	SPG	11.60	1	11.60	1.45	.2369
DG	SDG	21.10	1	21.10	6.38	.0169
AG	SAG	1.06	1	1.06	0.17	.6788
RG	SRG	161.77	1	161.77	15.25	.0005
SPD		139.74	31	4.51		
SPA		111.21	31	3.59		
SDA		943.43	31	30.43		
PDA	SPDA	7.06	1	7.06	2.53	.1222
SPR		110.01	31	3.55		
SDR		300.34	31	9.69		
PDR	SPDR	0.28	1	0.28	0.16	.6954
SAR		478.73	31	15.44		
PAR	SPAR	2.54	1	2.54	0.95	.3383
DAR	SDAR	0.94	1	0.94	0.08	.7726
SPG		247.30	31	7.98		
SDG		102.55	31	3.31		
PDG	SPDG	5.20	1	5.20	1.48	.2334
SAG		188.59	31	6.08		
PAG	SPAG	9.57	1	9.57	1.49	.2311
DAG	SDAG	16.25	1	16.25	2.45	.1275
SRG		328.76	31	10.61		
PRG	SPRG	9.57	1	9.57	1.40	.2452
DRG	SDRG	0.28	1	0.28	0.04	.8501
ARG	SARG	27.24	1	217.24	5.36	.0274
SPDA		86.60	31	2.79		
SPDR		56.00	31	1.81		
SPAR		83.24	31	2.69		
SDAR		342.34	31	11.04		
PDAR	SPDAR	5.20	1	5.20	0.91	.3476
SPDG		109.20	31	3.52		
SPAG		198.83	31	6.41		
SDAG		205.41	31	6.63		
PDAG	SPDAG	0.01	1	0.01	0.00	.9708
SPRG		211.46	31	6.82		
SDRG		241.00	31	7.77		
PDRG	SPDRG	6.73	1	6.73	1.56	.2212
SARG		157.55	31	5.08		
PARG	SPARG	0.17	1	0.17	0.05	.8306
DARG	SDARG	1.20	1	1.20	0.13	.7242
SPDAR		177.33	31	5.72		
SPDAG		200.65	31	6.47		
SPDRG		133.80	31	4.32		
SPARG		109.87	31	3.54		
SDARG		292.58	31	9.44		
PDARG	SPDARG	1.49	1	1.49	0.18	.6725
SPDARG		252.80	31	8.15		

6.3 Experiment 6: Position of Change

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		5776.81	31	186.35		
Gram	SG	528.35	1	528.35	23.94	.0000
Position	SP	266.82	2	133.41	4.39	.0164
SG		684.27	31	22.07		
SP		1882.19	62	20.36		
GP	SGP	712.55	2	356.28	28.33	.0000
SGP		779.83	62	12.58		

6.4 Experiment 6: Interview

Source	Error Term	Sum of Squares	DF	Mean Square	F	Prob.
Subject		86254.68	31	2782.41		
Type	ST	16304.10	1	16304.10	10.09	.0034
Gen-Spec.	SG	286.88	1	286.88	0.45	.5064
Pos-Neg.	SP	1345.97	1	1345.97	2.20	.1478
ST		50085.78	31	1615.67		
SG		19675.00	31	634.68		
TG	STG	96.29	1	96.29	0.32	.5772
SP		18936.40	31	610.85		
TP	STP	486.75	1	486.75	0.89	.3525
GP	SGP	10.97	1	10.97	0.03	.8627
STG		9403.09	31	303.33		
STP		16938.12	31	546.39		
SGP		11185.90	31	360.84		
TGP	STGP	1401.57	1	1401.57	4.09	.0518
STGP		10617.81	31	342.51		