This is to certify that the

dissertation entitled

CONSISTENCY IN THE USE OF PROTOTYPICAL
INFORMATION IN CATEGORY JUDGMENTS:
EFFECTS OF AGE AND CATEGORY LEVEL

presented by

Roxanne Lee Sullivan

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Major professor
Thomas H. Carr

4/26/84
Date August 27, 1984

# CONSISTENCY IN THE USE OF PROTOTYPICAL INFORMATION IN CATEGORY JUDGMENTS: EFFECTS OF AGE AND CATEGORY LEVEL

By

Roxanne Lee Sullivan

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1984

# ABSTRACT

## CONSISTENCY IN THE USE OF PROTOTYPICAL INFORMATION IN CATEGORY JUDGMENTS: EFFECTS OF AGE AND CATEGORY LEVEL

By

Roxanne Lee Sullivan

The ability to use category knowledge in a consistent manner was assessed by presenting subjects with a number of cognitive decision making tasks. Preschool children, second grade children, and adults performed an item generation task, a simple recognition memory task, a paried comparisons task, and a match-to-sample task with stimuli representing either a basic level category (dogs) or a superordinate level category (animals). According to a prototype approach to categorization, subjects operating with basic level category knowledge should demonstrate more efficient and consistent cognitive processing than subjects using superordinate level category information. The results of the present study were basically in contradiction to previous results in the area. Positive processing effects were not observed for the basic level category at any age level studied, with the exception of a decrease in the number of matching errors for basic level information. Both intra- and inter-subject consistency in the use of category information at either category level studied was not evident until the adult years.

The results were interpreted as indicating that caution should be taken in generalizing the idea of positive processing effects for basic level category information to all types of cognitive tasks and that expertise in an area of knowledge may be of importance in obtaining such effects.

To Dr. Sue R. Rosner and Dr. W. Fred Stultz, for their inspiration, guidance, and support throughout the past few years

## ACKNOWLEDGEMENTS

I would like to thank Dr. Thomas Carr for serving as my dissertation supervisor, especially for having the patience to see me through this project over both time and distance. I would also like to extend my appreciation to Dr. Gordon Wood, Dr. Ellen Strommen, Dr. J. Kathyrn Bock, and Dr. Santo Camilleri for their time and help in serving on my dissertation committee.

The completion of this part of my education would not have possible without the support of my family and friends. They helped me to "keep the faith" when the task at hand seemed impossible to handle.

Finally, I would like to thank the administration, faculty, staff, and, most importantly, the students of Bellevue College. During the past year, they have fulfilled the roles of both cheerleader and nagging mother that have supported me through the completion of this document. They have made earning my doctorate very special.

TABLE OF CONTENTS

Page

## LIST OF TABLES

LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

The purpose of this chapter is to provide an introduction to the area of categorization research, from both a theoretical and an empirical perspective. It is also the purpose of this chapter to provide a brief overview of the rationale, methodology, and hypotheses put forth for the study conducted.

## Background for the Problem

One area of research which has been a source of great interest and debate among investigators of human behavior is the manner in which people are able to reduce the complexity of the environment into a manageable number of meaningful units or categories. Especially of interest has been the ways in which categories are acquired, represented, and stored as part of the semantic knowledge base. Two of the major types of theory used to explain categorization behavior are traditional category theory, which is based on an Aristotelian model, and best example (or prototype) theory.

Proponents of the more traditional types of categorization theory have posited that categories may be defined by a set of arbitrarily determined criterial attributes. These criterial attributes are considered to be both necessary and sufficient for the determination of category membership. It is also posited that once a category's boundaries are established, those boundaries are absolute. The acquisition of a category involves the abstraction of this set of criterial attributes by observing the differences that occur between category members and non-members (Bourne, Dominowski, & Loftus, 1979; Bruner, Goodnow, & Austin, 1956; Farah, & Kossyln, 1982; Garner, 1978; Mervis, 1980; Mervis & Rosch, 1981; Smith & Medin, 1981).

Traditional category theory has proven to be quite useful in explaining the manner in which certain types of artificially designed categories are learned in laboratory settings (cf. Bruner et al., 1956; Vygotsky, 1962). The theory does have difficulty in explaining other types of categorization phenomena. For example, traditional category theory has difficulty explaining the results of studies using natural categories (e.g., categories that may be named with the nouns of a language) as the focus of investigation. Such studies have demonstrated that natural categories have "fuzzy", rather than absolute boundaries (Berlin & Kay, 1969; Mervis & Rosch, 1981; Smith, Shoben, & Rips, 1974). Traditional theories of categorization also have difficulties in explaining how very young children are able to acquire and use categorical information. Children under the age of seven years do not possess or are unable to use the kinds of abstraction processes (e.g., simultaneous hypothesis testing) that are required for category acquisition as posited by traditional theories of categorization (Farah & Kosslyn, 1982; Mervis, 1980; Mervis & Rosch, 1981; Rosch, 1973a, 1973b, 1976).

These and other issues of concern regarding the more traditional theories of categorization are discussed in more detail later in this paper.

According to prototype theory, categories are structured with respect to the idea of "family resemblance." That is, each member of a category may have one or more attributes in common with one or more other category members, but <u>no one attribute or group of attributes must be common to all category members</u>. In addition, those category members sharing the most attributes in common are judged to be the most representative or "best examples" of that category than are category members with few attributes in common. Those category items judged to be most representative or typical of the category form the category's core or <u>prototype</u>. Category boundaries are believed to be "fuzzy" or ill-defined, rather than absolute and, rather than being arbitrarily determined as posited by the more traditional theories, reflect the correlational attribute structure of the environment (i.e., the features of the environment are believed to have a natural structure and are not randomly combined). Finally, categories are posited to have several <u>levels of inclusiveness</u> which indicate the degree to which a category reflects the correlational attribute structure of the world, e.g., superordinate versus basic versus subordinate levels (Mervis, 1980; Mervis & Rosch, 1981; Rosch, 1973a, 1973b, 1975, 1977, 1978; Smith & Medin, 1981; Wittenstein, 1953).

Prototype theory has become popular with investigators of categorization behavior since the theory seems better able to explain many of the phenomena that are difficult to handle using a more traditional approach of categorization. For example, it has been demonstrated that people are able to judge how typical (i.e., representative) an item is of a specific category, that people are consistent with respect to each other in making this type of judgment, and that items judged as being typical of a category are processed in a more efficient manner than are items judged to be atypical (e.g., Berlin & Kay, 1969; Rosch, 1973a, 1973b; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Smith et al., 1974).

Prototype theory is also better able to cope with the developmental findings in the area of categorization, such as the ability of 10 to 24 month olds to form basic concepts like "female face" or "man." The theory works well because, according to prototype theory, all that is required for acquisition of a category is the ability to detect similarity among category instances, rather than the abstraction of criterial attributes or features (Cohen & Strauss, 1978; Farah & Kosslyn, 1982; Mervis, 1980; Rescorla, 1980; Ross, 1981; Sherman, 1981; Younger & Cohen, 1983). It has also been demonstrated that young children are able to make judgments on the basis of category representativeness and that the first categories that a child acquires correspond to the level of inclusiveness which most closely reflects the correlational attribute structure of the environment (i.e., the basic level). It is not until later in development that children acquire categories whose level of inclusion does not match the correlational attribute structure of the environment as closely as at the basic level, such as superordinate and subordinate levels (Anglin, 1977; Mervis & Crisafi, 1982; Mervis & Pani, 1981; Thompson & Bjorkland, 1981).

In summary, the data for both children and adults, with respect to natural categories, seem to be better accounted for by a prototype theory than by a more traditional-Aristotelian theory of categorization. Both children and adults have provided evidence to justify the presence of a "family resemblance" structure and of different levels of category inclusiveness as posited by the prototype approach. Also the finding that even young infants are able to form basic level categories add special emphasis to the role played by the levels of inclusiveness factor, in particular the basic level of inclusion, in human categorization. Thus, based on the empirical findings and theoretical principles of prototype theory described above the following study was conducted.

## Rationale and Statement of the Problem

One area for which we have little information concerns the consistency with which people use their personal knowledge of category structure (either consciously or unconsciously) when making various kinds of cognitive decisions. If one is to have a complete theory of human categorization, it is important to be able to provide not only a description of the way in which category knowledge is represented and organized in the cognitive system, but also a description of the manner in which the knowledge is used and the situations in which the knowledge is used. In other words, "Are people consistent in applying their knowledge of a category's prototypical structure when making cognitive decisions relating to that category or does the type of decision to be made (e.g., recognition versus membership judgments versus production of category instances) affect the way in which prototypical information is used?" A related question is: "Does consistency of use of prototypical information vary with respect to age or category inclusion level?" Previous studies which have been indirectly related to the issue of consistency have shown that young children (children under the age of four years) are not very consistent with respect to using attributes to assign items to a given category. Anglin (1977, 1978) had children produce definitions for various natural language categories (e.g., "animals," "plants") and then observed the children's classification behavior with respect to their definitions. Anglin found that children were rarely consistent in assigning instances to a category in accordance with their (the children's) definitions, even though the children's definition seemed quite appropriate according to adult standards. Anglin interpreted this finding as a lack of coordination between the child's knowledge of category members and knowledge of category attributes, suggesting an initial independence in the development of these types of behaviors.

Rosner and Poole (1981) have also looked at the consistency that young children show between the attributes ascribed to superordinate categories and the types of items which they include in those categories. First graders were asked attribute questions (e.g., "Is some (all) clothing worn on the feet?") and their responses were compared with their underinclusion and overinclusion of category instances. There was no consistency found between the occurrence of overgeneralized attribute errors and the overinclusion of marginal category items. In the case of underinclusion, the children did show some consistency between the attributes ascribed to a category and the items assigned to that category. Thus, with respect to the coordination between attribute and item assignment, young children seem to be quite inconsistent in their behavior.

The research on adults, on the other hand, suggest that their use of prototypical information may be much more consistent than that of the children. For example, Rosch et al.'s (1976) series of convergent validation experiments showed that for a number of different tasks (e.g., speed of verification, attribute production), people's behavior was consistent with the way in which the stimuli were ranked according to category typicality. That is, high typical items were responded to faster, were assigned more attributes in common, etc. Thus, the fact that these types of results have been found in several other studies (e.g., Mervis, Catlin, & Rosch, 1976; Rosch, Simpson, & Miller, 1976) helps to strengthen the position that adults can use prototypical and structural information about categories in essentially the same way in a number of task situations.

The data with respect to consistency of use of category knowledge reported so far has been concerned with the use of typicality norms (especially for the adults) based on group responses. What about consistency of use of one's <u>own personal</u> prototypical structure for a category? That individual differences in the use of prototypical structure may be important in making cognitive decisions has been suggested by two studies, one conducted with adults and one which was developmental in nature. In the first study, Coltheart and

Evans (1981) had adults generate bird names, a set of semantic scales for those names, and then had them make judgments about the generated names based on the semantic scales. These investigators found that, although a great deal of knowledge is shared concerning categories, the semantic spaces generated by their subjects were to an important degree unique for different individuals. In the second study (Thompson and Bjorkland, 1981), kindergarten, third, and sixth graders received a cue-at-input/cue-at-output recall task using atypical and typical category items selected from either age-appropriate norms or adult-defined norms. Children who received the age-appropriate lists performed significantly better on the recall task than children receiving the adult-defined lists. This finding stresses the importance of using child-generated norms when assessing children's categorization behavior.

Taking these types of findings into account, the following study was designed to assess both the within and between person consistency with which the prototypical information present in the structure of natural categories is used when making cognitive decisions as a function of age and category level. Children (preschoolers and second graders) and adults participated in each of four tasks designed to gain an assessment of a person's knowledge of either a superordinate level natural category ("animals") or a basic level natural category ("dogs"). The four tasks used were an item generation task, a simple recognition memory task, a paired comparisons task and a match-to-sample task. Each of these choices will be defended below.

It was decided to test both a basic level and a superordinate level category since the special status given to the basic level of category inclusion in the literature suggested that the prototypical structure of this level might be used in a more consistent fashion by both children and adults than might be the case for superordinate level. Part of the consistency of use may be based on the greater perceptual similarity of basic level category members and the fact that children tend to rely more on perceptual properties in their categorizations (Anglin, 1977; Horton & Markman, 1980; Mervis, 1980). The specific categories were selected based on the fact that norms (both child and adult) were available for them (Blewitt, 1981 and Thompson & Bjorkland, 1981, for the category of "dogs" and Rosner & Hayes, 1977, and Rosner & Poole, 1978, for the category "animals') and that there was a representative literature established for each of the category levels which shows that even very young children have some rudimentary knowledge for each of the levels investigated (Anglin, 1977; Blewitt, 1981; Carey, 1978; Nelson, 1974; Rosner & Hayes, 1977; Thompson & Bjorkland, 1981). This last point (the presence of category knowledge) is important since if one is to assess the consistency with which a person uses her category knowledge, one should be sure that the person being tested has some knowledge about the category of interest.

One procedure which has been used in the past to assess the extent and qualities of a person's knowledge of natural language categories is a category item production task (Nelson, 1974; Rosner & Hayes, 1977; Sullivan, 1979). In this task, a person is presented with a category label and is then asked to provide as many examples of that concept as he/she can in a limited amount of time. The greater the number of items produced and the greater the agreement among people about the items making up a category indicates a more thorough and interpersonally-shared knowledge of a category. Studies such as those conducted by Nelson (1974) and Rosner and Hayes (1977), have demonstrated that the ability to give appropriate category instances increases with age (youngest age group = five year olds), that the categories of the young children are generally smaller and showed more diversity than the categories of older children and adults. Furthermore, the ability to produce a large number of category instances is positively related to recall memory performance (Sullivan, 1979). Thus, it is believed that these types of results and the kinds of data obtained from an item production task (or item generation task, as it will be referred to in the remainder of this paper) is believed to reflect the amount of experience and exposure, both of a sensory and linguistic nature, that a person has had with the stimuli comprising a given category. The order in which items are produced may also be taken as an indication of the strength of those items in the person's knowledge base (with initial items being assigned greater strength than later items) and as an indicator of the organization of those items in the knowledge base (with initial items being considered the most representative or typical of a category for a person). The item generation task was also included to provide an indication of idiosyncratic knowledge of a category (i.e., items not represented by the experimental stimuli).

Unfortunately, an item generation task, while useful in its own way, is not a perfect measure of a person's category knowledge.

For example, does the inability to produce labels for category instances indicate a lack of knowledge for a given category? This does not seem very likely, since even infants have been observed to behave in such a way as to indicate limited categorical knowledge (e.g., Ross, 1980; Sherman, 1981) and since some current positions on children's acquisition of conceptual knowledge posit that a category is formed first on the basis of experience with category instances and then the category label is acquired (e.g., Nelson, 1978). That is, it is assumed that young children base their initial concepts on idiosyncratic, episodic knowledge of one or a few category instances (which may or may not be the same as the instances most associated with the adult version of that concept). As the child gains greater experience with instances of a concept, it is assumed that knowledge of that concept will become more like that of the mature concept user (Bowerman, 1978; Mervis & Canada, 1981; Nelson, 1978; Nelson & Brown, 1978). It has also been observed that when children (in this case two to six year olds) are asked to name different members of a category (e.g., provide names for different flowers), they will provide the class name ("flower") for each picture rather than a more specific name ("rose"), and if they are presented with an unfamiliar category instance they are able to correctly assign it to its appropriate category (Anglin, 1977). Thus, being able to produce specific names for category instances is only one indication (and only a partial one at that) of a person's knowledge of a category. Other types of behavior should be observed in addition to an item generation task. In the present study, both the ability to generate category item labels and the way in which people named the experimental stimuli were observed as indications of a person's category knowledge.

Since an item generation task does not provide a complete reflection of a person's category knowledge with respect to which items a person includes as category members, it probably does not completely reflect a person's organization of category knowledge, since only verbalizable category knowledge is tapped by the task. Thus, especially for children, a task is needed which will provide an assessment of a person's internal organization of a category that is not dependent upon having verbalizable knowledge of a category's members. A procedure which holds promise with respect to this issue is a paired comparisons task as described by Howard and Howard (1977). In their study, Howard and Howard (1977) presented their subjects (first, third, sixth graders and college students) with all possible pairs of 10 animal pictures and asked their subjects to make a dissimilarity judgment for each stimulus pair by having their subjects place the members of each pair in spatial proximity to one another. Thus, items believed to be very similar would be placed close together and those judged to be very dissimilar would be placed farther apart. The results of these judgments were entered into a multidimensional scaling procedure. It was found that even children in the youngest age group yielded a systematic dimensional structure reflecting the dimensions of size, domesticity, and predativity of the portrayed animals. Howard and Howard conclude that this structure indicated that the children were not responding in a random fashion.

Given the finding that young children could produce a systematic structuring for a set of stimulus materials, a variation of the Howard and Howard task was developed to assess a person's internal category structure based on item representativeness. In the present study, as in the Howard and Howard study, subjects were presented with all possible pairs of a set of picture stimuli, representing either a basic level category or a superordinate category. Unlike the Howard and Howard procedure, in which the subject was asked to make a spatial placement of the stimulus pairs to indicate how similar the items were to one another, the procedure used in this study involved having the subject select which stimulus in a given pair he/she believe was the "best" example of a stated category label. A structure for the stimulus set could then be constructed on the basis of the frequency with which each stimulus was chosen with respect to all other stimuli in the set. Those stimuli with a high frequency of choice would be assumed to reflect the more representative category instances (more "typical") for a person, while items with a low frequency of choice would be assumed to be less representative of a category for a given person. The idea that frequency of choice reflects category representativeness of category instances has been confirmed by Mervis and Rosch (1981). This procedure can then be used not only to provide data about the internal structure of an individual's knowledge of a category, but the data can also be used to assess the amount of agreement among people with respect to a given category's internal structure (i.e., How consistent are people in rank ordering the stimuli?). Finally, the data can be used to assess if the structure of a category changes with age.

Tasks such as the item generation task and paired comparisons task described above can give us some indication of the personal knowledge and structure of specific categories and even some information about consistency of knowledge within and across different groups of people (for the purpose of the present study, people of different age groups), yet they give us little or no information regarding the consistency across different situations with which a given individual uses her knowledge of category structure or is affected by that structure (since generation data may not match paired comparisons stimuli). In order to obtain this type of information, a match-to-sample task was designed in which the standard and composition of the comparison stimulus pairs were based on each subject's own structuring of the stimulus set in the paired comparisons task. For each subject, the standard was that stimulus which had the highest frequency of choice in the paired comparisons task. The next four most frequently chosen stimuli (plus the standard) were considered to be highly representative of the category for that subject and the remaining stimuli were considered as being less representative (or having low typicality) of that category for that subject. Five types of stimulus pairs were constructed: (a) high-high pairs, (b) low-low pairs, (c) high-low pairs, (d) high-different category pairs, and (e) low-different category pairs. Subjects were given several examples of each type of stimulus pair and were asked to select which pair member they thought was "most like" the standard. If a person is using category knowledge in a consistent fashion, then it could be assumed that the person's decisions will reflect that knowledge.

For example, high ranked stimuli should be selected over low ranked stimuli when paired together and rank ordering of the high and low stimuli (which can be obtained from choices in the high-high and low-low pairs, respectively) should be quite similar to those rankings observed in the original paired comparisons task if category knowledge is being used consistently. The procedure also allows the investigator to take into account differences in category structuring that may occur between age groups and which may influence the cognitive decisions to be made (Mervis & Canada, 1981; Thompson & Bjorkland, 1981).

A person's category structure may also produce more indirect effects on cognitive processing than those assessed by a match-to-sample task. For instance, how might an item's typicality affect memory processing? Since Thompson and Bjorkland (1981) had already demonstrated that recall is affected by category structure, it was decided to determine whether recognition memory would be affected by this type of knowledge in a similar manner. For example, is a person more likely to correctly recognize a target stimulus if she has rated that stimulus as being very typical of a category than if a target is of low typicality? Another question of interest is: Is a person more likely to say they recognize a distractor item as being part of the target list if the distractor is highly typical of the category than if a distractor is more atypical? In order to provide some answers to the above questions a simple (yes/no) recognition task was administered to the subjects and the subjects' responses were evaluated with respect to each individual subject's structuring of the stimulus materials. It has been demonstrated that this type of recognition task can be performed successfully by children as young as two years of age (Meyers & Perlmutter, 1978) and thus should be well within the capacity (as are the other tasks used) of the youngest age group observed in the study.

## Hypotheses

Based on the literature that has been reviewed thus far, the following hypotheses have been derived concerning the ability of children and adults to utilize various aspects of category structure when involved in making different types of cognitive decisions.

### Hypothesis I

Overall cognitive task performance will be positively related to chronological age.

### Hypothesis II

Cognitive task performance will vary as a function of the level of category inclusion about which a person is making decisions.

A. Cognitive task performance will be facilitated for subjects operating at the basic level of category inclusiveness.

B. Given the special status assigned to the basic level of category inclusion at very young ages, age differences will be smaller for subjects operating at the basic level of a category hierarchy than for subjects operating at the superordinate level of a category hierarchy.

Hypothesis III

The ability to provide an internal structuring for a natural category based on the typicality of category exemplars for different levels of category inclusion will be shown by both children and adults.

Hypothesis IV

The ability to use and/or to be affected by the internal structure of natural categories in making cognitive decisions will be positively related to chronological age.

Hypothesis V

The ability to consistently use the prototypical information available in the internal structure of natural categories will vary with the level of inclusion in a category hierarchy and with chronological age.

A. Consistency of use of prototypical information will be positively related to decisions made at the basic level of a natural category hierarchy both within and across age groups.

B. Consistency of use of prototypical information will be positively related with chronological age.

# CHAPTER II

## REVIEW OF THE LITERATURE

The world presents us with an incredibly large number of discriminably different objects, events, and impressions. Yet, if one observes human beings in interaction with the environment, much of the interaction seems to involve people treating clusters of these discriminably different items "as if" they were equivalent to one another, rather than as unique objects or events. That is, people group items into categories. The process of categorization has several functions. Among the functions of categorization are the reduction of environmental complexity, provision of labels for objects and events in the environment, reduction of the necessity of constant learning, and the ordering and relating of classes of events, i.e., the ability to go beyond the information given (Bruner, Goodnow, & Austin, 1956). Since the process of categorization is such a pervasive part of human behavior, the study of this process is crucial if we are to fully understand the human cognitive system.

The purpose of this chapter is to provide an overview of the major theoretical issues and empirical findings in the area of human category behavior that hold special importance for our ability to understand how categorization behavior develops. In order to provide sufficient background information for the study at hand, several topics in the realm of category behavior are reviewed. These topics include different theoretical approaches to categorization (with particular emphasis on prototype theory), infant categorization skills, and semantic memory development. Of special importance, for the purpose of the present study, is a review of research which has investigated the types of category knowledge that young children acquire during development and the manner in which children are able to use the category knowledge they have acquired.

## The Category/Concept Distinction

One potential source of confusion in the area of human categorization is the distinction, or lack of one, made between the terms, category and concept. Dictionary definitions seem to make the distinction on the basis of generality. For example, the Random House Dictionary of the English Language (1968) defines category as "a classificatory division in the system; a class or group." In contrast, a concept is defined as "a general notion or idea." This notion of a category representing some degree of specific information and a concept representing more general information is reflected in psychological descriptions of categorization. A category is often defined as a situation where a number of objects or events are considered to be and treated as if they are equivalent (Brunner et al., 1956; Mervis, 1980; Mervis & Rosch, 1981; Rosch, 1978; Rosch, Mervis, Johnson, Gray, & Boyes-Braem, 1956). A specific theory may add further to the definition of the term, category. For example, traditional approaches may define a category as a set of criterial attributes (Mervis, 1980). A prototype approach would define a category as a group of objects or events whose membership is based on the "family resemblance" of the items or that a category is a group of items named by words in natural languages (Mervis, 1980; Mervis & Rosch, 1981; Rosch, 1975). A feature list approach would define a category with respect to the relations among a set of relevant features characteristic of the category members (Bourne, Dominowski, & Loftus, 1979; Smith & Medin, 1981; Smith, Shoben, & Rips, 1974).

The definitions of the term, concept, while longer and more complex than the defintions for category, imply a more general notion, which often subsumes the ideas concerning categories. Bruner et al. (1956) in the book, A Study of Thinking, provide a formal definition of the term, concept, which states that "a concept is a network of sign-significate inferences by which one goes beyond a set of observed criterial properties exhibited by an object or event to the class identity of the object or event in question and therefore to additional inferences about other unobserved properties of the object or event." Anglin (1977) has simplified this definition somewhat by defining a concept as "all of the knowledge possessed by an individual about the category of objects denoted by a term." This knowledge may include such things as the knowledge of a set of identifying features and general spatio-temporal and functional relations among previously experienced category items (cf., Nelson, 1977b). Smith and Medin (1981) go further by proposing that concepts are essentially pattern recognition devices which a person can utilize to help in the classification of new objects or events and to draw inferences about those new objects or events. Thus, a category often refers to the specific items and their characteristics which form some type of consensually agreed upon group, while a concept refers to all of the knowledge about a category, plus the ability to use that knowledge to go beyond what is observed.

Three problems arise from the kinds of definition that are usually offered for the terms, category, and concept. First, is the notion that a concept or category is somehow equivalent to the word that labels it. Nelson (1977a) has stressed that one can have a concept prior to having a label for that concept and thus, it is important to separate the linguistic and conceptual systems, especially when stuyding the acquisition of concepts by young children. Second, the two terms are often used interchangeably in the literature, even when a distinction has been made between the two terms. This state of affairs may result from an incomplete understanding of the distinction that is made between concepts and categories or more likely, from the possibility that one may not be able to study a category empirically without invoking more general knowledge of the underlying concept or vice versa.

Finally, the various definitions of concept and category often fail to take note of the purpose behind the formation of categories and concepts, especially that concepts and categories may hold different meanings for different groups of people, in particular children and adults. It was with this particular problem in mind that the following definition of concept was adopted as a guiding influence for the present study: concepts are representations of the categories by which a person groups and divides the objects and events of the world in an adaptive, psychologically efficient way (Farah & Kosslyn, 1982). While this definition may not go far in clarifying for empirical purposes, when one is studying a concept or a category (and for the purpose of the present study a distinction is not made and the term, category, has been used throughout this paper to correspond to terminology used by prototype theorists), it does permit the idea that, although a child's knowledge and structure of a category may not be the same as that of an adult, that knowledge may still be sufficient for the child to operate efficiently in his/her environment. That is, just because a child's category knowledge is not the same as an adult's knowledge and structure of that same category, it does not mean that the child is in some way inferior to the adult. It is the ability to use knowledge in an efficient manner which is of importance.

## Theories of Categorization

Traditional category theory. The ways in which humans go about organizing the environment and the structure, mode of representation and acquisition of the resulting categories, have long been a source of interest and debate for observers of human behavior. According to the more traditional approaches to human categorization (i.e., those approaches based on an Aristotelian model), categories are defined by a set of criterial attributes (an attribute being something which helps to define a particular item or object, such as "is red," "has fur," but which is not necessarily synonymous with the stimulus, although an attribute label may also be a category label (Garner, 1978; Mervis & Rosch, 1981). This set of criterial attributes is both necessary and sufficient (i.e., all category members must possess the criterial attributes) for category membership. Once an object is determined to possess the criterial attributes for a given category, that object is given full and equal status with respect to all other category members. Furthermore, the set of criterial attributes which define a given category is determined arbitrarily. That is, the traditional approach to categorization assumes that there is no inherent structure to account for the occurrence of attributes in the world; all possible combinations of every attribute and its associated values are equally likely to occur in the environment. Thus, the way in which attributes occur in the world provides no basis for deciding how the objects and events of the world should be divided up into categories. Finally, the traditional approach also assumes that once a category is established, the category's boundaries are absolute (based on Bourne et al., 1979; Garner, 1978; Mervis, 1980; Mervis & Rosch, 1981; Smith & Medin, 1981).

Given these characteristics ascribed to categories by the traditional approaches to human categorization, the process of acquiring a category involves the abstraction of the set of criterial attributes by observing which attributes occur in all and only objects which belong to the category being acquired, i.e., items that are category members (Bourne et al., 1979; Bruner, Goodnow, & Austin, 1956; Farah & Kosslyn, 1982; Mervis, 1980; Mervis & Rosch, 1981; Smith & Medin, 1981). Accordingly, the research generated by the traditional approach has been concerned with the types of strategies that are used to discover the sets of criterial attributes for the categories studied (Bourne et al., 1979; Bruner et al., 1956). These strategies usually involve sophisticated hypothesis-testing abilities that are not fully developed until the elementary school years (Farah & Kosslyn, 1982). Thus, a traditional theory of categorization would predict that true categories would not be acquired until fairly late in childhood and that once acquired the categories of children and adults would be exactly the same (Bruner et al., 1956; Vygotsky, 1962). As will become evident in later sections of this paper, the situation seems to be quite different.

At the present time, the traditional approach to human categorization has come under attack by investigators of human cognition. While the traditional approaches have been quite useful in explaining the acquisition of certain types of artificially designed categories in laboratory settings (cf., Bruner et al., 1956, although prototype-like categories have been designed, cf. Posner & Keele), these approaches have difficulty with the results of studies of natural categories (e.g., those categories which may be named by the nouns of a language). Several investigators have reported that, rather than having absolute boundaries and category members which are ascribed full and equal status most natural categories are characterized by "fuzzy" boundaries and that category membership is ascribed on a "more or less" basis, with some category instances being considered as "better examples" or "most typical" instances of a particular category (Berlin & Kay, 1969; McCloskey & Glucksberg, 1978; Mervis, 1980; Mervis & Rosch, 1981; Miller, 1978; Rosch, 1973a, 1977b, 1975; Rosch et al., 1976; Smith et al., 1974; Solso & McCarthy, 1981). For example, it has been demonstrated that items which are judged to be typical members of a category are categorized in a more efficient manner (i.e., faster decision times and fewer categorization errors) than items which are judged to be less typical or representative of a category (e.g., Rips, Shoben, & Smith, 1973; Rosch, 1973b; Rosch et al., 1976; Smith & Medin, 1981). If natural categories are structured in the way suggested by the traditional approaches to categorization, there should have been no differences in the processing of category instances and subjects should not have found the task to rate category instances as to their typicality or representativeness a meaningful one.

Another area in which the appropriateness of traditional approaches to categorization has been questioned is the idea that the set of criterial attributes for a category is arbitrarily determined. Related to this idea is that a category _is_ defined by a set of necessary and sufficient attributes. If the first idea is true (i.e., arbitrariness), it would mean that all attributes in the world could be combined and would occur with equal probability in combination with each other (Garner, 1978; Mervis & Rosch, 1981). Yet, if one pays close attention to the environment, it can be observed that not all attributes in the environment are equally likely to occur with each other. Certain attributes are more likely to occur together than are combinations of other attributes. The environment therefore does appear to have some form of natural structure (Garner, 1978; Mervis, 1980; Mervis & Rosch, 1981; Rosch, 1973a, 1973b, 1975, 1977, 1978; Rosch et al., 1976). For example, a creature which has _feathers_ is unlikely to have four legs, fur, and bear its young live. The creature is very likely to have wings, a beak, and lay eggs (based on Mervis, 1980).

The second idea (that all members of a category possess a complete set of criterial attributes) has been criticized from several directions. The first area is that of the typicality studies mentioned above. If all category instances meet all of the criteria of that category, all category instances should be treated in the same manner. The typicality literature demonstrates that not all category instances are created equal. A second criticism centers around the idea of "family resemblance" (Wittgenstein, 1953). According to the notion of "family resemblance," each category member has one or more attributes in common with one or more other category members, but <u>no one attribite must be in common</u> to all category members, Rosch and Mervis (1975) have demonstrated that several categories are structured as family resemblances, and Rosch, Simpson, and Miller (1976) have shown with three different types of artificial categories, that the degree of "family resemblance" of category members determined the ease of item learning, speed of classification of items after learning of the category, ratings of the typicality of items, the order in which items were generated in a production task, and the facilitation or inhibition of responses to items in a priming task. In other words, those category items with the highest family resemblance (i.e., shared the greatest number of common attributes) were learned the fastest, rated as being highly typical of the category, were generated first in the production task, etc.

Third, if categories are defined by arbitrarily determined sets of criterial attributes, different levels of generality for the categorization of items would not be an issue and is not considered by traditional theory. As with the first two areas discussed with respect to the issue of arbitrariness in categorization, evidence has been gathered to indicate that there are different levels of generality for categorization which are not treated as equivalent and of which one level is considered to be more "basic" than other levels (Berlin & Kay, 1969; Mervis, 1980; Mervis & Pani, 1980; Mervis & Rosch, 1981; Murphy & Smith, 1982; Rosch, 1975, 1977, 1978; Rosch et al., 1976a). This basic level of categorization is the level at which the division of objects is the closest to the perceived structure of the environment and has been shown to be the most general level for which people are able to list a large number of attributes which the majority of category members share, the level at which a concrete image may be formed and the level at which people have similar motor programs for responding to category items. Category instances are also processed more efficiently (similar to the findings in the family resemblance studies) if they are members of a basic level category (Mervis, 1980; Mervis & Pani, 1980; Murphy & Smith, 1982; Rosch, 1975, 1977, 1978; Rosch et al., 1976a; Smith & Medin, 1981). This idea of levels of categorization is fundamental to an approach that has been developed to compete with the traditional approaches to categorization and will be described further when presenting the prototype theory of categorization.

Armstrong, Gleitman, and Gleitman (1982), Osherson and Smith (1982), and Smith and Medin (1981) have all suggested ways in which the traditional approach to categorization could be altered to provide an account for the types of results obtained in the studies cited above. The main characteristic of these modified approaches is that a category is described as having two parts--_a core description_ which relates the category with other categories and thoughts, and an _identification procedure_ which provides information necessary to make category membership decisions. Unfortunately, the modified versions of the traditional theories still run into the same types of problems as the original theories (e.g., use of non-necessary attributes, failure to specify how defining features are determined; Smith & Medin, 1981). In addition, the modified versions that appear to be the most promising are those modifications (such as the accessible identification procedures described by Smith & Medin, 1981) which move toward more probabilistic notions which are characteristic of the prototype approach and more featural approaches, such as that proposed by Smith et al. (1974) which describe categories as possessing _defining features_ which are present in all members of a category, but also as having characteristic features which are present in some, but not all members of a category which may be used in place of defining features when making category decisions (e.g., Keil & Batterman, in press).

_Prototype theory_. Given the types of findings and the criticisms of the traditional approach to categorization that were described previously, an alternative approach to the problem of human categorization has been proposed by Eleanor Rosch and her associates (Mervis, 1980; Mervis & Rosch, 1981; Rosch, 1973b, 1975, 1977, 1978).

The approach is often referred to as prototype or best example theory. According to this approach, categories are not necessarily delineated by a set of criterial attributes, but rather are related with respect to "family resemblance" ideas. Category membership is considered to be a matter of degree, with some category members being considered as being more representative or typical of the category than are other members. The category members judged to be the most typical of a category form the category's core or prototype and the rest of the category extends away from the core with the poorer examples of the category being further removed from the core than are category instances of more intermediate typicality. This idea of differences in category member representativeness is formally called by Rosch the category's horizontal dimension (or internal structure).

Rosch proposes a second dimension that is characteristic of category systems: the vertical dimension. According to the prototype approach, the vertical dimension is concerned with the level of inclusiveness of a category. That is, category systems are characterized by a hierarchical structure. While certain investigators have suggested that there are six different levels of categorization (e.g., Berlin, 1978), Rosch and her associates have proposed three major levels of category inclusion: (a) superordinate, (b) basic, and (c) subordinate. Of these different levels, it is the basic level which is considered to be the most important to human categorization, since it is believed that the basic level is that level which most closely reflects the correlational attribute structure of the world and for which the exemplars of this category level have the maximal within-category similarity and minimal between-category similarity. Thus, the basic level of categorization is the level which provides the most readily usable information about the category without sacrificing cognitive economy by overloading the system with unnecessary information (Mervis, 1980; Mervis & Rosch, 1981; Rosch, 1973b, 1977, 1978).

The prototype approach to categorization not only handles the adult literature well, but it also provides us with a means to account for current findings with respect to category development which are difficult to account for by using the more traditional theories. For example, the more traditional theories of category development (such as the theory proposed by Bruner et al. 1956), posit that categories are formed through a process of hypothesis generation and hypothesis testing of potential category attributes until the appropriate set of criterial attributes is discovered. Unfortunately, children are not able to perform this type of systematic hypothesis generation and testing in an efficient manner until adolescence (Farah & Kosslyn, 1982; Flavell, 1977). Yet, it is during the periods of infancy and young childhood that children are acquiring categories at a rapid rate. Even very young infants (10 to 24 months of age) are able to form a variety of basic categories and preschool children are able to produce and have their cognitive performance affected by categorical information (Anglin, 1977; Cohen & Strauss, 1978; Goldberg, Perlmutter, & Meyers, 1974; Lange, Kienapple, Sullivan, & Allen, 1981; Rescorla, 1980; Ross, 1981; Sherman, 1981; Sullivan, 1979; Younger & Cohen, 1983a, 1983b). Thus, at ages where traditional theories would predict little or no appropriate categorization, children do show the ability to form and use (in a limited way) categories, though their use may only be limited relative to adult competencies.

Given this state of affairs, how do young children acquire their knowledge of natural language categories? According to prototype theory, all that is needed for the formation of a category is for the child to detect the similarity among instances of a category so that a prototype can be selected or formed (Farah & Kosslyn, 1982; Mervis, 1980). A category may be formed on the basis of one instance of a category and as the child has more experience with instances and is able to note the similarities among those instances, the category will come to resemble the adult category (Bowerman, 1978; Farah & Kosslyn, 1982; Mervis, 1980; Mervis & Pani, 1980; Nelson, 1977a). If children do acquire categories by noting the similarities among category instances, then it would be logical to predict that the categories which a child first acquires should be those categories for which the intra-category similarity is at its maximum and the between-category similarity is at a minimum, that is, categories at the basic level.

Another prediction made by prototype theory is that the structure of children's early categories should be quite similar to the adult structure, in that typical instances should be included in both of their structures and discrepancies between those structures should be with respect to the poor category instances (Mervis, 1980). In other words, the categories of children should show evidence of an internal structure based on the representativeness of category instances.

Although the prototype theory does appear to handle the results of empirical studies very well, the theory is not without its critics. One serious criticism of the theory is its failure to specify the exact form of representation of a category's prototype (is the representation a template; an average form; a linguistic description?). This is quite similar to the failure of traditional theories of categorization to specify how criterial attributes are defined and agreed upon (Smith & Medin, 1981). Another criticism of the theory is that some categories which do appear to have well-defined boundaries with specific rules describing category membership (e.g., "odd numbers"), have been shown to exhibit a prototypical structure when people are asked to judge if one category instance is better than another instance (i.e., one "odd number" will be judged as being a better example of the category than another "odd number;" Armstrong et al., 1982; Osherson & Smith, 1981). These types of findings pose difficulties for prototype theory, since they indicate the possibility that the structure of categories shown by prototype studies is an artifact of the tasks used, rather than a true indication of the representation of a category item, if a person is asked to make a judgment of category typicality, they will make such judgments even though there is no underlying basis for such a judgment. Finally, Osherson and Smith (1981) have stated that a prototype theory cannot handle complex concepts, such as "truth" or "belief" (although such concepts are difficult for any theory of categorization).

Although these criticisms appear to hold some validity and should be taken under consideration when attempting to interpret data from categorization studies, a prototype theory still appears to be the most useful approach for describing the development of children's categories, especially object and perceptual categories. Even investigators of category development who would not specifically label their own approaches to the problem as being a prototype approach, have suggested that the idea of a prototype or category core (whether based on one or a number of category instances or being formed on the basis of perceptual or functional characteristics) is extremely useful in describing young children's category behavior (Anglin, 1977; Bowerman, 1978; Nelson, 1977a, 1977b).

## Research on Category Acquisition and Category Use

The following section will focus on current trends in the study of children's category development, especially as they relate to the prototype approach to categorization. Areas to be covered include infant categorization skills, semantic memory development, children's knowledge of category structure and the acquisition of different levels of categorization.

Infant categorization skills. According to prototype theory, all that is really necessary for a person to acquire a category is the ability to detect similarities among category instances and to note the naturally occurring breaks in the correlational attribute structure of the world. Since studies of infant perception have demonstrated that young infants can discriminate among various patterns (Hetherington & Parke, 1975) and that in order to show discrimination one must be able to detect differences and similarities among items, it is only logical to assume that infants have the basic abilities needed to form categories. The development of the habituation-dishabituation paradigm has provided a means to study the categorization skills of very young infants.

Younger and Cohen (1983a, 1983b) wanted to assess at what age infants would show sensitivity to correlational structures within a category and whether categories based on correlated features would be acquired differently than categories characteristized by uncorrelated features. In the first study, Younger and Cohen (1983a) found evidence to support a developmental progression in the ability of infants to process simple correlational information. At four months of age, infants appear to be able to process only specific features of category instances. By the age of seven months, infants are able to perceive relations among the features of a single category object. Finally, by the age of ten months, infants appear to be able to "abstract" invariant relations from a category.

In a second study, Younger and Cohen (1983B) habituated 24 ten-month olds to one of two sets of animal stimuli, one set which was designed with a number of attributes which were perfectly correlated and one set for which none of the attributes were correlated. Three novel animals were presented to each infant to assess if the infants were basing their responses with respect to specific features or correlated features. Younger and Cohen found that it was much easier for the infants to learn the stimulus set characterized by correlated features than for them to learn the uncorrelated stimulus set. Infants in the correlated set condition showed reliable habituation and responded on the basis of the correlation, whereas infants in the uncorrelated set condition showed little, if any, habituation (i.e., they responded as if all the stimuli, both previously experienced and novel, were different from each other). Thus, it does appear that young infants are able to detect and use to some extent the correlational attribute structure of the environment.

Ross (1980) wanted to assess if the within-category perceptual similarity would affect what categories are acquired early in life. Ross utilized a habituation-dishabituation paradigm. In her study, 12-, 18-, and 24-month olds were shown successive members of a category followed by the simultaneous progression of a previously unseen category member and a member of a novel category. It was demonstrated that, regardless of the child's age, children habituated to categories that were designed to be the most similar perceptually (i.e., "M"'s and "O"'s). The children did not show habituation to categories that were perceptually dissimilar (i.e., food and furniture). The findings were mixed for the categories at the intermediate levels of similarity (i.e., man and animals). The children showed habituation for the category, man, but not for the category, animals. Ross believes this is the result of the fact that the stimuli for the category man, were actually more physically alike than the animal stimuli. Ross concludes that by using a procedure which utilizes receptive behavior (attention) rather than productive activity, one can demonstrate the ability to recognize conventional categories at a very young age and that the types of categorization abilities that one can demonstrate in children is highly dependent upon the methodology used, the type of response required, and the types of categories being examined.

The issue of the role of methodology in the extent to which an infant's categorization skills may be underestimated was of concern to Tracy Sherman (1981). Of special concern to Sherman's work was that the traditional methodology did not allow enough time with each category instance so that the infant would be able to discriminate among instances. Thus, Sherman developed a modification of a paired comparison technique in conjunction with the habituation-dishabituation paradigm to test the categorization skills of ten-month olds for the category, male human faces. The procedure involved the presentation of two copies of a category instance simultaneously for a set number of trials, followed by a test trial which paired one of the copies with a new category instance. If the infant showed habituation to the old category instance (i.e., looked longer at the novel instance), the same procedure was then continued with the next category instance. If the infant did not show habituation to the old category instance, further familiarization trials were presented, followed by another test trial. This procedure constituted the study phase of her experiments.

The face stimuli used varied on three dimensions, involving nose width, nose position, and distance between the eyes. All other features of the faces were identical. Face stimuli which were used as "out of category" stimuli were also of males, but were composed of novel features. The actual category instances chosen represented the extreme values of all three dimensions and infants never saw category instances which represented the middle values of the dimensions.

In the first experiment, Sherman wanted to assess if an infant would regard a never-seen-before prototype as more familiar than an out-of-category instance and whether the infant would discriminate between seen and novel within-category instances. The results showed that the infants did treat the prototype as more familiar than an out-of-category instance, but failed to discriminate with respect to within-category instances. Sherman interpreted these results as suggesting that an infant's encoding of studied instances is close to some type of summary of the features rather than as a separate memory for each specific face.

The second experiment investigated whether a prototype based on the average value of the dimensions (mean prototype) or a prototype based on the most frequently observed features (modal prototype) would most closely correspond to the infant's representation of the face category. The major finding of this experiment was that the modal prototype was treated as significantly more familiar than was the mean prototype and Sherman reported that a similar finding has been observed for adults. Thus, it appears that an infant can abstract category information from observed instances when the discriminability of the category instances is guaranteed by the laboratory situation and that investigators must design their research so as to avoid the possibility of underestimating the infant's categorization abilities.

Rescorla (1980) approached the area of infant categorization from a different perspective than the investigators cited above.

Instead of using a habituation-dishabituation paradigm, Rescorla relied on diary data and comprehension/production test data to assess the category knowledge of six children between the ages of 12 and 20 months. The acquisition of three natural language categories was studied. In general, Rescorla found that during this age period, children develop some grasp of what the most focal items are of a given category, although their categories were less differentiated than the categories of adults. Rescorla believes that this finding is consistent with Rosch's view that certain category instances may be treated as focal instances of the category which serves as prototypical exemplars as part of the internal structure of the category.

With respect to the more specific findings concerning the category, animals (which are of special importance for the present study), Rescorla found that the category evidenced its greatest period of development between the ages of 12 and 17 months and that all of the children had dog and cat as part of their category. Further, the category, animals, was more differentiated than either of the two remaining categories studied, vehicles and fruit. Thus, by the time children are of preschool age (as in the present study) they should have a good working knowledge of the category, animals, based on their early emerging knowledge of the category found by Rescorla.

Summary. The studies cited above are a representative sample of the types of research being conducted with respect to categorization skills prior to the second year of life. The studies have all demonstrated that infants possess far greater categorization skills than can be accounted for by traditional theories of categorization, but which can be handled well by prototype theory and which provide support for the theory.

Semantic memory development. Researchers who have attempted to study the development of semantic memory have typically investigated the acquisition of natural language concepts. These studies have generally utilized one of two procedures in order to tap semantic memory. The first procedure involves the sorting of experimenter-selected exemplars of various natural language categories by subjects of different ages. The second procedure involves the use of an item production paradigm, in which subjects generate their own exemplars to given natural language category labels.

Saltz, Soller, and Sigel (1972) sought to examine developmental trends in the acquisition of natural language categories. Children at each of three age levels (5-6, 8-9, and 11-12 years of age) were asked to select which of 70 different pictures were instances of six different natural language categories. The categories included in this study were food, animals, transportation, clothes, toys, and furniture.

It was found that with age, the number of items identified per category increased and that the number of core items (a core item being defined as an item chosen 75% of the time) also increased with age. Younger children also tended to be least flexible in the multiple classification of items.

From their results, Saltz et al. suggested two major developmental trends. First, younger children tend to acquire fragmented subconcepts which are strongly tied to specific stimulus contexts. With age, category integration developed. Second, younger children are heavily dependent on perceptual attributes in identifying categories. With age, functional and abstract attributes become more important in identifying natural language categories. Thus, Saltz et al. suggested that the main drift of categorical development is most strongly characterized by a shift away from overdiscrimination and toward integration rather than a shift away from over generalization and toward differentiation.

In an attempt to replicate the findings of Saltz et al. (1972), Neimark (1974) had subjects in grades two, six and college sort 50 pictures with respect to the class labels "food," "things to eat," "clothing," and "things to wear." The second and fourth categories were used to test for the effect of label specificity.

Neimark's results revealed that the classes "food" and "things to eat" were functionally equivalent only for second graders. With age, there was a decrease in the size and composition of the category "things to wear." It is from the findings related to this category that Neimark concluded that her results do not support the earlier findings of Saltz et al. (1972) of increases in the size of natural language categories with age.

On the basis of her findings, Neimark also concluded that it appears that the natural categories of second graders are quite similar to the natural categories of college students. Neimark also concluded that the course of development of natural categories is affected not only by age, but also by the label and specific nature of the instances provided for categorization.

One criticism of the above stated research is that it has been based on stimulus sets of category instances presented and selected by the experimenter. Thus, the child has not been asked to produce instances that may not be tapped by experimenter provided instances. Nelson (1974) provided a remedy to this situation by having her five-year-old and eight-year-old children generate instances for nine natural language categories. Nelson believed that this procedure would reflect a different aspect of the contents and characteristics of young children's categories than those revealed by studies using experimenter-defined sets.

The results of Nelson's study revealed that eight-year-old children gave almost twice as many items on the average than did five-year-olds (8.04 and 4.64, respectively). The average number of responses per category were extremely variable. In terms of agreement as to the typical category members, five-year-olds appeared to agree among themselves to the same extent as did eight-year-olds. Also, though the categories of the five year olds were generally smaller, the categories showed more diversity of membership than did the categories of the eight-year-olds. Nelson also found that certain large categories, such as furniture, clothes, and tools, showed a tendency to overflow their boundaries, whereas categories which were actually subclasses of more general categories (e.g., insects, vegetables) produced the fewest responses, and the least agreed-upon core members.

Nelson states that in general, her findings do not appear to support the findings of Saltz, Soller, and Sigel (1972). The younger subjects in her study did not display a tendency toward fragmented or narrowly restricted concepts as would be predicted with the Saltz et al. results. In fact, in many cases, children appeared to have wide, unbounded categories. Nelson also did not find any evidence in her sample that younger children showed a greater reliance on perceptual attributes than did older children. Both groups of children appeared to rely on functional definitions rather than on perceptual or more abstract definitions of the categories. Finally, Nelson postulates that category growth seems to take place through the greater hierarchization and articulation of categories which takes place with age.

In order to obtain child norms and to investigate two alternative types of category bias reported to be shown by young children (i.e., the underinclusions of appropriate items and the overinclusion of inappropriate items), a category item production task was presented to preschool and grade school children by Rosner and Hayes (1977). In the study, the children produced verbal responses to four category labels: animals, food, furniture, and clothes. The children's responses to each of these categories were then judged as to their category appropriateness by a group of college students. Rosner and Hayes found that the mean number of responses per category increased with age and older children produced a greater number of distinct words per category than did younger children. With respect to the appropriateness of the category instances, younger children received significantly lower appropriateness scores than did older children, except for the category of animals.

Based on the above findings, Rosner and Hayes concluded that category membership is a matter of degree. They suggested that both underinclusion of appropriate items and overinclusion of inappropriate items are evident as developmental phenomena. The authors also put forth some cautions in interpreting production data. In particular, Rosner and Hayes caution that production data may be confounded by age differences in strategies, such as search, retrieval, and item censoring.

Rosner and Poole (1978) extended the work of Rosner and Hayes by examining longitudinally, children's judgments of category membership when the category strength of the to-be-judged terms (based on the Rosner-Hayes norms) was varied. The children were first tested in kindergarten and then again in first grade. Specifically, Rosner and Poole wanted to assess whether age changes in inclusion would be accompanied by increase in overinclusion. The children judged whether or not a word was a category member by simply saying "yes" or "no". Children were then presented with training and tested on their target category or on an irrelevant category. It was found that for the category, furniture, neither age or training affected the proportion of items judged to be category members. Both age and training did increase the proportion of items judged as belonging to the category, clothes. Also, it was found that the amount of underinclusion of category items was greater for the furniture category than for the clothes category. A second major finding of the study was that overinclusion did co-occur with the inclusion of moderate strength items, suggesting that there is great variation in the range of attributes characteristics of acceptable items and that what the young child may be doing is inferring that the occurrence of a single attribute or a few attributes is good enough to include an item as a category member.

Summary. The semantic memory literature does provide evidence that young children do have the ability to group category members together and to verbally produce category instances when presented with a category label. While the children's categories do differ from the categories of an adult (small, more diverse), they do have common knowledge of certain category instances and with age, the children's categories come to approximate those of adults.

Children's knowledge of the internal structure of categories. According to Mervis (1980) a prototype theory of categorization would predict that the structure of a child's natural language categories should be very similar to the internal structure of the adult categories. That is, both children and adults should include the most typical category instances in their structures and should be able to make a distinction between typical instances and those category instances which are less representative of that category. Several studies have been conducted to assess if children do exhibit the kind of internal category structure predicted by prototype theory and to determine on what basis the structuring is designed and whether the internal structure of a category affects cognitive performance.

Mervis, Catlin, and Rosch (1975) wanted to assess if there were any developmental differences in the structure of color categories. Previous research in the area, specifically the work of Berlin and Kay (1969), demonstrated that color categories were structured around a core of focal or best examples of a given color category which transcended language differences in the use of color terms but whose boundaries were dependent on a language's color terminology. Mervis et al. wanted to evaluate whether the focal cores of color categories remained invariant across age.

In their experiments, kindergarteners, third graders, and adults were presented color chips and a color label and were asked to select the best example for the color term presented. The results of this procedure indicated that the focal color chips for each color label was exactly the same for <u>all</u> age groups, except for the purple focus (although the difference between the adults' and children's focus was not significant). It was also found that the variance in judgments of the focal colors (across testing sessions) decreased with age on the dimensions of hue, brightness, and saturation. When color category boundaries were examined, it was found that it was not until the third grade that children learn that color categories must be mutually exclusive and that the boundary judgments of the children at both ages differed from the boundary judgments of the adults. Thus, at least for a perceptual category, color, children do have a category structure (with respect to the notion of best example) by at least the age of five years and probably earlier of one considers the results of the Sherman (1981) study.

Whereas Mervis et al. (1975) study concentrated on children's structuring of a perceptual category, Thompson and Bjorkland (1981) examined the ability of children to judge the typicality of instances of several natural categories (e.g., dogs, musical instruments, buildings). In their study, kindergarten, third, and sixth grade children and adults were presented with lists of words representing the categories and were asked to rate each category instance with respect to their category representativeness on a three-point scale. In order to assess if the children were actually basing their judgments on item typicality, an additional group of subjects (both children and adults) were asked to rate the same items on how much they liked each item.

First, it was demonstrated that there was no correlation between typicality judgments and liking judgments. The correlations did vary according to category and did increase with age, with some of the correlations approaching significance by adulthood. That is, children did appear to be making their typicality judgments on the basis of how representative an item was of a given category, but by adulthood, subjects were having a more difficult time in separating out preference from typicality.

Second, it was demonstrated that even the kindergarten children selected most of the items that adults defined as being typical of a category as being typical for them. It was for the less typical items that age differences occurred, with the inclusion of the less typical items decreasing with age. Mean typicality ratings did increase with age, as did the correlations between the child and adult ratings (although the ratings of even the kindergarten children were significantly correlated with the ratings of the adults).

An additional aspect of the Thompson and Bjorkland study was to examine if the judged typicality of an item would affect children's memory performance. When the to-be-remembered lists were constructed using age-appropriate normed lists, recall was greater than when adult appropriate normed lists were used. Typical items were recalled better than atypical items. Thus, while there is similarity in the ways in which children and adults rate category instances, the differences that do occur are especially important with respect to efficient cognitive performance for the children. Thompson and Bjorkland suggested that, given the type of results they have obtained, it may be extremely important to use child-generated norms when studying children's cognitive processing.

An important aspect of prototype theory is that a prototype may be formed through experience with a number of category examples. The work by Sherman (1981) cited earlier has demonstrated that infants did seem to be able to form a prototype for the category, male face, based on the frequency of occurrence of the various category attributes. Boswell and Green (1982) looked at children's ability to abstract and recognize prototypes based on polygon variations. The authors used a prototype plus transformations design. The result of the study showed that the performance of adults supported the view that category knowledge is structured around a central prototype with little retention of specific exemplar information regardless of whether categorization was required for only previously seen stimuli or for all related stimuli (i.e., targets and distractors). Preschool children, on the other hand, only showed prototype knowledge under the latter, more constraint free condition, but demonstrated a greater degree of exemplar specific knowledge than the adults when required to categorize only the previously experienced stimuli.

Boswell and Green interpreted their findings as being suggestive of the idea that children's knowledge of categories contains both prototype and exemplar specific knowledge, the activation of which depends on the demands of the task. Adults do not seem to be able to differentiate between the two types of information when faced with a situation in which a prototype must be distinguished from previously experienced stimuli. It may be that the types of tasks encountered by adults in their day-to-day environment require fast decisions based on some form of summary knowledge, while children, who are still very much in the active process of acquiring many categories, need to have on hand much more specific information concerning individual exemplars until an adequate knowledge base is built up.

Another feature of category structure which has been studied is the kinds of attributes or dimensions that are used when assessing category membership and if these attributes and dimensions differ with age. Howard and Howard (1977) conducted a multidimensional scaling analysis of animal names for children in first-, third-, and sixth-grade and for adults. Dissimilarity judgments were obtained for all possible pairs of ten animal names. For the children, the judgments were made as part of a zoo scenario. That is, children were told that their task was to place animals who were "very much alike" in cages next to each other and to place animals that were "very different" in cages that were further apart.

The multidimensional scaling procedure which was carried out showed that even the youngest children were not randomly responding to the stimuli, but provided judgments that yielded a systematic dimensional structure. The three dimensions which emerged from the analysis were size, domesticity, and predativity. From a developmental perspective, the youngest subjects had higher weights on the size dimension, while older subjects and adults had higher weights on the domesticity and predativity dimensions, suggesting a progression from the use of perceptual to more abstract features in assigning a structure to a natural category.

Melkman, Tversky, and Baratz (1981) investigated the preference for perceptual and conceptual attributes in the clustering of items for preschool and fourth-grade children. The children received a grouping task and a recall task. The results of the grouping task showed a clear preference for different types of attributes at different ages. For the four-year-old subjects, color and form were equally likely to determine a child's grouping behavior. By the age of five years, form dominated the groupings. The majority of the nine-year-olds formed their groups on the basis of conceptual attributes. This progression mirrors the progression from perceptual to abstract features obtained by Howard and Howard. The same progression was evident in the recall clustering scores for free recall in the Melkman et al., but when the task was cued recall, conceptual information was used by all ages. Thus, as in the Boswell and Green (1982) study, task demands determined the type of knowledge used by a child, in this case, perceptual versus conceptual information.

Rosner and Poole (1981) examined the issue of whether or not children show consistency between the attributes they ascribe to a category and the types of items which a child includes as members of a category. The investigators asked children to judge if an item was a member of a given category and then asked the children a series of attribute questions about the category (e.g., "Is some (all) furniture used to sleep on?"). The results of the inclusion task were the same as the results obtained by the Rosner and Poole (1978) study. For the attribute question task, children (first graders) made more errors than adults (i.e., responding "no" to a "some" statement or "yes" to an "all" statement). When the relation between attributes and item inclusion was assessed, it was found that attribute responses were related to underinclusion of items, but not to overinclusion. That is, children who agreed to a statement that <u>all</u> members of a category possessed a given attribute, would exclude certain category exemplars as category members (e.g., a child might agree that <u>all</u> clothes are worn on the body, but would not include <u>shoes</u> as a member of the category, clothes).

Although there was a lack of <u>total</u> consistency between attribute and membership judgments, statistically, Rosner and Poole's subjects did show some reliable consistency. This suggests that young children do have some coordination between their knowledge of category attributes and category instances. Rosner and Poole believe that their findings are consistent with Anglin's (1977) findings of little consistency between a child's verbal definition of a category and category item inclusion.

Non-prototype perspectives. Finally, two further studies will be presented which have examined the effects of category structure from two different perspectives than the studies cited above--information processing, as characterized by the use of semantic priming techniques and a Piagetian approach.

McCauley, Sperber, and their colleagues (McCauley, Weil, & Sperber, 1976; Sperber, Davies, Merrill, & McCauley, 1982) used a semantic priming technique to examine the development of memory structure and category relationships. McCauley et al. (1976) demonstrated that both kindergarten and second grade children responded faster to targets preceded by a high associate prime than when the prime was a low associate. With respect to category relatedness, only the second graders showed a significant priming effect when a high, categorically related prime preceded the target. The authors suggested that, at least for this type of task, associative and categorical relationships only become integrated after the age of seven.

Sperber et al. (1982) looked at the differences in the processing of perceptual categories (i.e., categories whose members are visually similar to one another) and non-perceptual categories (i.e., categories whose members are visually dissimilar to one another). Semantic priming and category verification tasks were used to assess both active and passive processing. The results indicated that knowledge of perceptual categories developed earlier (around the second grade) than for non-perceptual categories for both active and passive processing, except for the finding that second grade children did not show significantly faster verification speeds for perceptual over non-perceptual categories. By the fifth grade, knowledge of perceptual category relationships is well enough established to promote automatic activation of this type of category information, it is not until adolescence that automatic activation becomes evident for non-perceptual catgeories. Sperber et al. proposed that these results suggest that the mechanism underlying the acquisition of these types of categories may differ. That is, the acquisition of perceptual categories may occur early in development based on the child's ability to recognize similarities among category instances (ala prototype theory) and non-perceptual category acquisition occurred later in development because it involves more sophisticated abstraction processes to discover the rules by which category membership is defined.

A Piagetian perspective was used by Carson and Abrahamson (1976) to evaluate whether class inclusion performance would be affected when class inclusion problems included atypical category examples. Seven- and ten-year-olds were presented class inclusion problems for six natural language catgeories. The problems involved showing the subject a number of category instances which were subclasses of a major category (e.g., animals: horse, dog or bee, butterfly) and which were either typical or atypical examples of the category (e.g., dog = typical; bee = atypical). The ratings were obtained from fourth graders. The children were presented the problems in the following form: Category = food; examples = 5 hamburgers, 3 hot dogs; Inclusion question = "Is there more food or more hamburgers?" When the children's responses were analyzed, it was found that class inclusion performance was significantly better when the subclass examples represented typical category instances than when atypical exemplars were presented. Children who did display class inclusion behavior were often unwilling to extend the category label to cover items which were both atypical category instances, yet when a sorting task was given, children had no difficulty in grouping typical and atypical items together. Carson and Abrahamson stated that rather than showing a problem of conceptualization in the class inclusion task, the results of the sorting task, suggest that the problem is more semantic in nature (that is, children's difficulty with class inclusion problems).

Summary. Children do appear to have internal category structures which are similar to the category structures of adults, especially with respect to the ability to judge differences in category item typicality and the ability to form prototypes under certain circumstances. Differences occur with respect to the types of information used to structure a category (perceptual versus abstract information) and the extent to which categorical information can be used for efficient cognitive performance.

The acquisition of different category levels. Prototype theory not only proposes that categories are structured according to the representativeness of category instances (the horizontal dimension) but also that there are several levels of inclusion (the vertical dimension) that represents the degree to which a named category reflects the natural attribute structure of the environment (Mervis, 1980; Mervis & Rosch, 1981). Of special importance is the basic level of categorization, since it is the level of categorization at which categories form which most closely reflect large, naturally occurring attribute clusters. It is the level at which between-category differentiation is at its greatest; the level which seems to be the "most obvious way of dividing up the world" (Mervis, 1980). Since the basic level of categorization does appear to hold a special place in the theory, prototype theory predicts that a child's first categories and categorization skills should be at the basic level and then later at the superordinate and subordinate levels. It is also predicted that parents will recognize their children's perference for basic level information and will provide labels for objects and events at the basic level rather than at the more abstract levels (Mervis, 1980). The purpose of this section is to present empirical investigations of these predictions.

Do children acquire basic level categories prior to other levels of categorization? Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976a) did an analysis of the transcripts of Roger Brown's subject, Sarah, and found that almost all of Sarah's object terms were at the basic level. As part of the same study, three and four year old children were asked to sort objects into either basic or superordinate groups. While both the age groups were successful at sorting at the basic level, only the older children were able to successfully sort at the superordinate level. Anglin (1977) has also demonstrated that children begin learning categories at an intermediate level of abstraction and believes that these early categories are basically perceptual in nature (which is not that different from a prototype approach, since the basic level is the level at which category instances are the most perceptually similar).

Mervis and Crisafi (1982) used unfamiliar "nonsense" pictures to test the hypothesis that basic level categories are acquired prior to superordinate categories and that both of these category levels are acquired prior to the subordinate level. Children, ranging in age from two and one half to five and one half years, were presented with two nonsense stimuli and were asked to indicate which of the stimuli was most like a standard at each of the three category levels. The stimuli were designed to conform to the internal attribute structure of natural categories.

The results indicated that even children at the youngest age had better than chance performance for the basic level stimulus sets. At four years of age, children had better than chance performance on both basic and superordinate level sets. It was only after the age of five and one half years that children were also able to operate successfully at the subordinate level. Thus, the predicted progression of category level acquisition was supported.

Horton and Markman (1980) examined the effects of exemplar and linguistic information on the acquisition of basic and superordinate level categories. It was hypothesized that the ability to use linguistically specific information about a category in order to aid acquisition would be a function of a child's age and the type of category level being acquired. Since basic level categories have a high degree of perceptual similarity, it was predicted that these categories could be learned simply by exposing the child to category exemplars. In contrast, the dissimilarity among superordinate level category items would suggest that the child would need to be informed of the relevant criteria that are necessary for successful categorization. This was exactly the situation; linguistic information facilitated category acquisition only for the superordinate category level and for kindergarten and first grade children, but not for preschoolers. Horton and Markman stated that their results suggest that perceptual similarity among category members was enough to support the acquisition of basic level categories, but that the diversity that is characteristic of superordinate level categories make their acquisition difficult without further information about those categories, plus the ability to utilize this additional information in order to aid acquisition.

While these studies have clarified the status of the basic level in category acquisition, the studies have not gone into much detail as to the actual conditions which facilitate basic level category acquisition, Mervis and Pani (1980) hypothesized that categories acquired on the basis of exposure to typical category examples should be learned more accurately and easily than categories based on exposure to atypical category examples.

Five-year-olds and adults were taught names for six categories of artificial stimuli that were designed to mimic the structure of natural language categories. The subjects were asked to name the stimuli and later to point to the appropriate objects when presented with a category label. Artificial stimuli were used to control for familiarity effects. The results supported all of the hypotheses made by Mervis and Pani. That is, subjects acquiring categories on the basis of exposure to typical examples learned those categories more accurately and easily than subjects exposed to poor category examples. The only case in which this state of affairs did not occur was when categorization was expected on the basis of only one category example. In this situation, acquisition of a category based on one good example was not inherently any easier than learning a category based on one poor example. These results were further supported in a second, pseudo-naturalistic experiment. Thus, the ideal situation for learning a basic level category would involve the presentation of a number of good category examples for which the similarities among the examples are easily discernible.

What role do parents play in a child's acquisition of basic level categories? Do they encourage their children's use of basic level categories or do they provide input to their children to move beyond the basic level of categorization?

Mervis and Canada (1981) wanted to test the hypothesis that not only are children's first categories at the basic level, but that they are <u>child</u> basic categories, that mothers initially accept and encourage the use of by their children over adult basic categories. Three children and their mothers were observed longitudinally for nine months. The children's initial ages were 12½ months, 13 months, and 18 months. Three categories were studied: "kitty," "ball," and "car." The categories were represented by toys that included true (adult judged) category instances, related objects thought to be member of the child basic category, and unrelated objects.

Results were reported for the first two months of the study. These results indicated that the children's initial categories were indeed child basic, structured along the same principles as the adult categories. It was also shown that the children's mothers encouraged the use of the child basic categories. The only times in which a mother would re-label an object with the appropriate adult basic label and for which the child would accept and begin to use the adult appropriate label was when the child indicated a crucial difference between different basic level items or when the mother demonstrated a special use for an object and would provide the appropriate adult basic label.

Blewitt (1981) also found evidence that adults direct and encourage children to use the basic level of categorization. In her study, both teacher and maternal speech to preschool children was recorded and the nouns contained in the speech samples were rated for their level of category inclusion: subordinate, basic, or superordinate. Depending on the type of speech sample being rated and the individual rater, 70 to 89 percent of the nouns in the samples were at the basic level, 10 to 27 percent of the nouns were at the subordinate level and 1 to 5 percent of the nouns used were at the superordinate level. Thus, children were more likely to be given labels at the basic level than for any other level of categorization.

Blewitt was concerned that this extreme use of basic level terms by adults to young children might represent a simplication strategy designed to orient the child to the basic level of categorization. In order to assess this possibility, a group of college students were asked to write two versions of the same story, one version was intended for a child listener and the other version for an adult listener. When the stories were rated as to the levels of noun usage, it was found that the story intended for the child listener contained significantly more basic level nouns than did the story intended for the adult listener. The adult version of the story, on the other hand, tended to contain more subordinate level nouns.

Summary. It has been demonstrated that the basic level of categorization does hold a special place in the child's acquisition of category knowledge. It has also been shown that parents, and adults in general, recognize the preference of children for this level of categorization, encouraging its use by the child until the child requests more specific information or if the situation requires a different level of category knowledge.

# CHAPTER III

## METHOD

### Subjects

Children. The child sample was composed of children representing two grade levels: preschool and second grade. The preschool sample consisted of 20 children (11 males, 9 females; mean CA = 56 mos.) selected from children enrolled in classes at the Eastminister Child Development Center in East Lansing, Michigan who had received parental consent to participate in the study. The second grade sample consisted of 12 children (5 males, 7 females; mean CA = 91.25 mos.) selected through contact made with the parents of seven- and eight-year-olds through the use of ads in local newspapers in the Greater Lansing, Michigan metropolitan area. Parents were reimbursed for bringing their children to the laboratory. All of the children in the study had parental consent to participate and gave their own personal assent to participate prior to the start of any testing session. In order to assure confidentiality and anonymity, each child was assigned an identification number and parents were informed that data analysis would be conducted on a group basis.

Adults. The adult sample was composed of 22 college students (10 males, 12 females). The subjects were recruited from students enrolled in the introductory psychology courses at Michigan State

University and received credit for their participation which could be applied in meeting course requirements. Informed consent was obtained prior to the beginning of the testing session. Confidentiality and anonymity were assured in the same manner as for the child sample.

In addition to the adult sample described above, a second group of 34 college students was recruited to provide judgments of the physical similarity and category typicality of the stimuli used in the study. Of the 34 subjects in this group, the data of two subjects were not included in the analysis of the physical similarity judgments--one subject for failure to follow instructions and one subject because of an electrical power failure which occurred in the middle of the session.

Finally, all subjects who participated in any phase of the study were required to be native English speakers. Subjects were also required to have normal or corrected-to-normal vision and hearing.

## Stimulus Materials and Apparatus

Materials. The stimuli used in the study were color pictures representing the categories, "animals" and "dogs". The animal pictures used were selected from several books of the Golden Stamp Book series, specifically the Golden Stamp Book of Animals, Mammals, Horses, Birds and Snakes, Turtles, and Lizards. Pictures of dogs were selected from Simon and Schusters' Guide to Dogs (1980). The pictures used portray a single example of a given category (e.g., one cocker spaniel or one gorilla) against a natural outdoors background. The stimuli were chosen so that the exemplar in question was clearly distinguishable from the background and an attempt was made to equate the stimuli with respect to the orientation of the pictured category instance (e.g., profile vs. frontal views). Each picture was mounted on a 3 3/4" x 5 3/4" white card. The animal pictures measured 2 1/2" x 3 1/4"; the dog pictures measured 2 1/2" x 2 1/2".

For each category level, two sets of 10 pictures each were constructed representing items ranging in judged typicality from very typical to very atypical. One set of stimuli at each category level was designated as the target list (i.e., the set of stimuli with which subjects operated in all major tasks) and the other set of stimuli was used as the distractor list in the recognition memory task. In each stimulus set, five pictures represented high typical category instances and five pictures represented low typical category instances. The procedures for establishing the category typicality of each stimulus is discussed below. Finally, all subjects operating with a given category level were presented with the same target and distractor stimulus sets.

<u>Stimulus selection</u>. The exemplars representing the superordinate category level, "animals," were chosen from a pool of items selected from typicality norms provided by Rosner and her associates (Rosner & Hayes, 1977; Rosner & Poole, 1978) and constrained by the availability of pictures in the <u>Golden Stamp Books</u>. The Rosner norms were originally based on the percentage of college students who judged category exemplars generated by five-year-olds, ten-year-olds, and adults as to the appropriateness of the exemplars as instances of a given category. The norms have been subsequently re-scaled by asking college students to judge how typical they thought each

exemplar was of given category based on a seven-point Likert-type scale. The stimulus pool for the present study included items which received ratings of one or two in the Rosner norms (i.e., very typical) or ratings of four or five in the norms (i.e., medium to low typical). Use of the lowest end of the scale is not represented in the norms. A total stimulus pool of 51 stimulus items was constructed and the final group of 20 stimuli were selected on the basis of the typicality ratings (based on a five-point Likert-type scale, where 1 = very typical; 5 = very atypical) of the initial stimulus pool by a group of five pilot subjects. A listing of the stimuli used for the two category levels of their mean typicality rating based on the pilot data appears in Table 1.

The exemplars representing the basic level category, "dogs," were chosen from a pool of items based on norms provided by Blewitt (1981) and Thompson and Bjorkland (1981). The Blewitt norms are based on the typicality judgments of college students using a procedure similar to that used in the construction of the Rosner-Hayes typicality norms. Thompson and Bjorkland (1981) obtained norms for kindergarten children as well as for college students by using a three-point Likert-type scale (1 = good; 2 = OK; 3 = bad). Since the sum total of the number of items needed to construct the necessary stimulus sets was not met by the number of items represented by the two sets of norms, additional category exemplars were included in the stimulus pool based on the investigator's intuition as to their category typicality. The final stimulus selection pool contained 35 exemplars and the final stimulus lists used in the study were chosen in the same way as the final stimulus lists for the animal

Table 1

Stimulus Lists

I. Dogs

| A. Targets | Mean Typ.[1] | B. Distractors | Mean Typ. |
|---|---|---|---|
| German Shepherd | 2.0 | St. Bernard | 2.0 |
| Great Dane | 1.8 | Cocker Spaniel | 1.8 |
| Irish Setter | 1.8 | Bloodhound | 1.8 |
| Malamute | 1.4 | Brittany Spaniel | 1.4 |
| Collie | 1.0 | Harrier | 1.2 |
| Yorkshire Terrier | 4.0 | Komondor | 4.0 |
| Shi Tzu | 4.8 | Maltese | 4.4 |
| Greyhound | 3.6 | Little Lion Dog | 4.2 |
| Afghan Hound | 3.6 | Scottish Terrier | 3.2 |
| Pekingese | 3.4 | Belgian Sheepdog | 3.4 |

II. Animals

| A. Targets | Mean Typ. | B. Distractors | Mean Typ. |
|---|---|---|---|
| Horse | 1.0 | Lion | 1.4 |
| Rabbit | 1.8 | Raccoon | 1.2 |
| Fox | 1.4 | Squirrel | 1.4 |
| Tiger | 1.4 | Deer | 1.4 |
| Bear | 1.2 | Wolf | 1.6 |
| Cardinal | 4.0 | Owl | 4.0 |
| Anteater | 4.2 | Walrus | 4.0 |
| Llama | 4.6 | Lizard | 4.8 |
| Alligator | 4.6 | Tortoise | 4.4 |
| Frog | 4.2 | Boa Constrictor | 4.4 |

[1] typ. = typicality

category. The exemplars of the category, "dogs," and their respective category typicality ratings appear in Table 1.

In order to further verify the typicality of the stimulus sets used in the study and to provide additional information on the intracategory physical similarity of the stimulus items, a group of 34 (32 for the physical similarity ratings) college students performed two rating tasks. Half of the subjects performed the ratings tasks using the animal stimuli and half rated the dog stimuli. In the first rating task, subjects were given the appropriate category label and were asked to rate each stimulus on a five-point Likert-type scale as to how good an example (i.e., how representative) was of the category in question (1 = very good; 5 = very poor). Stimuli were presented one at a time for five seconds each, and the subjects were instructed to use the full range of the scale.

The second rating task involved having the subject make judgments concerning the physical similarity of pairs of category items. Subjects were instructed to make their judgments based on the overall physical similarity, especially overall shape, of the portrayed items. A five-point Likert-type scale was used where 1 = very similar; 5 = very dissimilar. Subjects were shown all possible pairs of the 20 stimulus items (190 pairs), for approximately ten seconds per pair. This data is of interest in that what may appear to be responding on the basis of category structure in the main experimental tasks, may actually reflect responding on the basis of the physical similarity of the stimuli. Thus, by having this type of information it is possible to address this issue.

In addition to the stimulus lists representing the categories, animals and dogs, a pool of items not representing either of the categories used in the main study, yet mimicking the category level factor, was developed to be used as the non-category contrasts in the match-to-sample task. The pictures chosen represent the categories, cars and vehicles. The car pictures were black and white photographs and were selected from Consumer Guide to 1982 Cars (1982). The vehicles were color pictures selected from illustrations by H. McNaught in 500 Words to Grow On (1973). As with the dog and animal stimuli, the non-category stimuli were mounted on white cards. Also the stimuli used in the practice tasks represented the categories of fruit and food and thus, mimic the superordinate and basic levels of a category hierarchy.

Apparatus. A portable stimulus presentation device was required which would allow the presentation of, and a measure of response to, a single stimulus (item recognition task), a pair of stimuli (paired comparisons task), and a triad of stimuli (match-to-sample task). A diagram of the apparatus designed to perform these functions, from the subject's viewpoint is shown in Figure 1.

The apparatus was constructed of wood painted in light blue enamel and measured 30 x 23 1/2" x 7 1/2" deep. It could be easily placed and operated on either a table or floor. When seated in front of the apparatus, the subject saw three openings or "windows" arranged in a triangle, with one "window" at the top of the apparatus and two "windows" at the bottom of the apparatus. The subject also viewed a sliding wood panel which covered the bottom two "windows" when closed, and two toggle switches. Each window measured

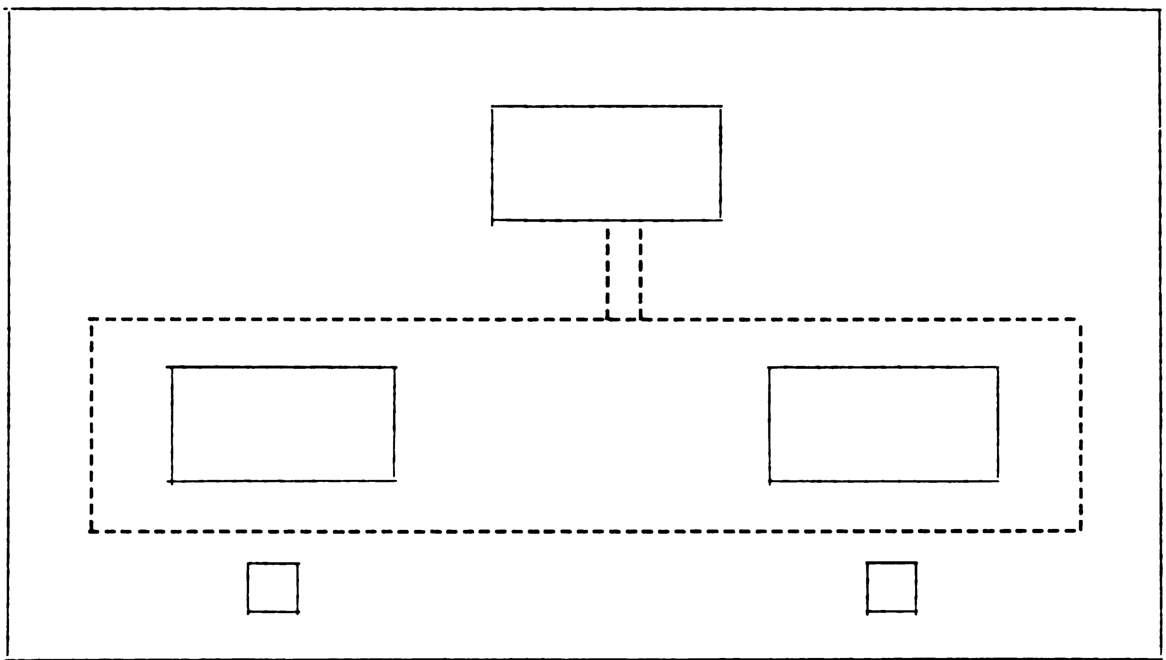

Figure 1. Diagram of Apparatus

3 1/2" x 5 1/2" and the lower two windows were spaced 7 3/4" apart. The sliding panel, which could be moved in an up/down manner and which was controlled from the rear of the apparatus by the investigator, measured 5 1/2" x 21 1/2". In order to conceal the top "window" when in use during the item recognition task, a removable wood panel measuring 7 1/2" x 8 1/2" was attached to the sliding panel. At other times when a given window was not being used for stimulus presentation, a blank white card would appear in that window. The toggle switches were located 7 1/2" below each of the two lower windows and 2" from the bottom of the apparatus. The switches were connected to a clock at the rear of the apparatus. In addition to the features mentioned above, removable signs used in the item recognition task to designate acceptance or rejection of a stimulus item could be positioned next to each of the switches.

When viewed from the investigator's vantage, several additional features of the apparatus were apparent. The size of each window measured 4" x 6" and included a ledge and clip to hold the stimulus cards securely in place. A sweephand clock, which could be read to hundredths of a second, was connected to a nine volt power supply and placed at the bottom of the apparatus. Two lights, one above each of the lower windows, were also attached to the back of the apparatus.

The apparatus operated in the following manner: With the sliding panel(s) in the ready position (i.e., concealing the appropriate windows), the investigator placed the stimulus card(s) in the appropriate window(s) for a given task. The investigator then dropped the sliding panel(s), revealing the stimuli, and depressing a microswitch which started the clock. When one of the switches was depressed by the subject at the front of the apparatus, the circuit was completed and the clock stopped. Depressing one of the switches also turned on the light indicating the subject's stimulus choice. Once a choice had been made, the investigator returned the sliding panel(s) to the ready position, recorded the subject's choice and response time, and reset the clock and switches (for the younger children). The older children and adults reset the switches themselves. The procedure was then repeated the required number of times for each specific task.

## Design

The format of the study was such that children at two grade levels (preschool and second grade) and college students were required to perform four tasks which were designed to reflect the ways in which the internal structure of natural categories affects cognitive processing with respect to those categories. The four tasks used were an item generation task, a paired comparisons task, a match-to-sample task, and a simple item recognition memory task. For half of the subjects at each grade level, the cognitive decisions made in each of the tasks were made for the superordinate level category, animals. The remaining subjects made their decisions for the basic level category, dogs. For each subject, the four tasks were distributed across two testing sessions, with the item generation, paired comparisons, and item recognition memory tasks comprising the first session and the match-to-sample task being conducted during the second session. This distribution of tasks across sessions allowed for the individual tailoring of the match-to-sample task based on each subject's performance in the Session 1 tasks, thus providing some degree of ecological validity.

In order to counterbalance for order of task presentation effects in the first session, approximately half of the subjects at each grade and category level received the item recognition task followed by the paired comparisons task. The remaining subjects received the paired comparisons task followed by the item recognition task. For all subjects, the item generation task was the first task of Session 1.

Each of the tasks yields several measures of interest to the study. First, the item generation task yields information which is thought to reflect a subject's verbalizable knowledge of the category levels under investigation and information as to if this type of knowledge varies according to category level and grade level. Three different measures, one based on individual data and two based on group data, speak to the above issue. The first measure is the number of items produced by a subject (not including repetitions or obvious miscategorizations) for a given category level. The measure has been considered to be a reflection of a subject's verbalizable and <u>personal</u> knowledge of category (Nelson, 1974; Rosner & Hayes, 1977). The second and third measures are calculated based on overall group performance on the task. The second measure to be used is a type/token ratio, which is defined as the number of different responses generated divided by the total number of responses given to a specific category label for a particular group of subjects (Nelson, 1974). The type/token ratio reflects differences between groups of subjects in the diversity of choices for a given category. The final item generation measure is referred to by Nelson (1974) as the Index of Commonality. The index is defined as the frequency of the three most popular responses divided by the total number of responses for the first three choices given by each subject at a particular grade level for each category being examined. This measure is an index of the degree of agreement within a group of subjects with respect to category membership.

In the recognition memory task, the main dependent variable of interest is the time taken by a subject to accept or reject a test stimulus as a target based on the judged category typicality of a given stimulus. Also of interest to the study is the effect the prototypical structure of a category level has on a subject's ability to accept or reject test stimuli approximately and a measure of sensitivity (<u>d'</u>). Thus, the effects of grade level, category level, and prototypical structure was assessed for such measures as percent correct acceptances, percent false alarms, and <u>d'</u>.

The paired comparisons and match-to-sample tasks provide information as to a subject's own prototypical structure for a given category level as reflected in the rank orderings of the stimuli which are the result of these tasks. Thus, it is assumed that the rank ordering obtained for each subject reflects that subject's organization of the stimulus items with respect to how representative the stimuli are to that subject of the category level under examination. The rank orderings are based on the frequency of choice for each stimulus with respect to all other stimuli representing a specific category level (or in the match-to-sample task, for all stimuli in a high or low typicality group) with items having a greater frequency of choice being considered as being more representative of a category level for a given subject than items having a lower frequency of choice. Once obtained these rank orderings can be used to assess consistency of category structuring both across and within grade levels for a given category level. Errors made in selecting pair members as being similar to a standard in the match-to-sample task (e.g., selecting a low ranked exemplar as being more like the standard than its high ranked exemplar pair mate) may also be considered to reflect a subject's ability to use his own internal category structure in a consistent manner. Finally, the effect of internal category structure on cognitive processing was assessed by looking at the times taken by subjects to respond for each of five pair types (based on combinations of high ranked, low ranked, and unrelated category instances) presented in the match-to-sample task.

## Procedure

Each subject was tested individually by the investigator. Subjects were randomly assigned to each of the category levels of interest with the constraint that there were approximately an equal number of subjects per category level. Subjects were also randomly assigned to each of the different task orders used for the first testing session. The task orders used were either item generation-paired comparisons-recogniton or item generation-recognition-paired comparisons task. It was decided to test these two task orders to control for any possible effects of intrusiveness of a given task with respect to the cognitive decisions to be made. The different task orders may also be viewed as a control for the effect of repeated presentation of stimuli on cognitive decisions. The item generation task was always presented first to avoid the possible contamination of the subject's generation that could result from seeing the stimulus pictures for a particular category.

Session I. As stated above, each subject was tested individually in a room in which the subject and the investigator were alone. The adult and second grade subjects sat at a table beside the investigator. The apparatus was set up on the table facing the subject. The apparatus was set up on the floor of the testing room for the preschool subjects and the children sat on the floor in front of the apparatus. The subject then began the item generation task in the following manner (instructions to be presented below for all tasks will be in the child's form--appropriate adjustments were made for the adults) and complete instructions appear in Appendix A:

> "Let's see how good a remember you are. I'm going to say the names of some things. After I say the name, I want you to tell me all of the things you can think of that belong to that name. For instance, I might say "flowers" and you might think of rose, daisy, tulip, and lots of others. Let's try it. I think I'll say the word, "toys". Can you tell me the names of all of the toys you can think of? (Child responds for one minute.) That was good. The next time I say a name

all you have to remember is to tell me the names of all of the things that you can think of that go with the name I say. Do you know what you are to do? Good. Let's begin."

The investigator then presented the subject with each of the category level names (i.e., "animals" and "dogs"), one at a time. For each label, the subject was allowed two minutes to respond. After 15 seconds of silence, the investigator prompted the subject ("Can you think of anything else that is a __________?") and when the subject indicated that he/she could think of nothing else or at the end of the two minute period, the task ended. The category, animals, was always the first label generated to by the subject.

Upon completion of the item generation task, the investigator proceeded either directly to the paired comparisons task or to the item recognition task. For the purpose of illustration, the situation in which the paired comparisons task occurs first, followed by the recognition task will be described below.

In order to begin the paired comparisons task, a set of four practice pictures (fruits or food, depending on the category level being tested) were used to demonstrate the task. The subject was told to pretend that a man from outer space had just landed on Earth. The subjects were told that the spaceman knew nothing about Earth and that it was their task to help the spaceman to learn about our world. The subjects were told that they would see some pictures, two pictures at a time, and that they were to pick the picture that they thought would help the spaceman to learn the most about the category in question by pressing the button below that picture. The experimenter stressed that the subject was to determine which picture would be most helpful to the spaceman in trying to understand the catgeory and not to pick the picture that they liked the most. The investigator then presented four practice trials (at the appropriate category level, superordinate or basic), stressing the task instructions and correcting the subject is necessary. Once it appeared that the subject understood the task, the investigator presented all possible pairs of the target stimuli (45 pairs). Two predetermined random presentation orders were constructed, with half of the subjects receiving each order. The investigator recorded the subject's stimulus choice and time to make a response on each trial. The above procedure was based on similar procedures used by Horton and Markman (1981) and Thompson and Bjorkland (1981).

Immediately following the completion of the paired comparisons task, the investigator began the item recognition memory task. The subject was told that the task was meant to see "how good a rememberer" he/she was. As with the previous task, a short practice task was presented. The child was told that he/she would see three pictures, one picture at a time. It was emphasized to the child that he/she watch the pictures very carefully so that he/she would be able to remember them. The investigator then presented the practice stimuli at a rate of five seconds per picture. The practice stimuli for the recognition task were the same as for the paired comparisons task. Following presentation of the practice stimuli, the investigator told the subject that he/she would see three more pictures, some of which would be exactly the same as the pictures he/she had just viewed and some which would be new. The subject was told that the pictures would appear in the upper window of the apparatus. The switches were designated as "yes" and "no" by black and white line drawings of a "smiley face" for "yes" and a "sad face" for "no". The child was instructed to depress the switch with the "smiley face" if a picture was <u>exactly</u> the same as one he/she had previously been shown and to depress the switch with the "sad face" if a picture had not been viewed before. In order to confirm that the subject understood the task, the investigator said, "Which button do you push if a picture is exactly the same as one of the pictures I just showed to you?" and "Which button do you push if the picture is a new one?" The investigator observed the child's responses and made corrections when necessary. Once it appeared that the subject understood the task, the investigator presented the three practice stimuli, one of which was the same as one of the stimuli in the practice study list.

Upon completion of the practice task, the investigator said, "Now I am going to show you a lot more pictures, one picture at a time. Make sure you watch each picture very carefully so that you will be able to remember it later." The appropriate target stimuli were then presented, one picture at a time for five seconds each. After presentation of the target list, a buffer task (color naming) was administered for approximately 30 seconds. The subject was then told that he/she would see some more pictures, some of which would be exactly the same as the ones just viewed and some of which would be new. The subject was then reminded that the task was to depress the switch with the "smiley face" if a picture was the same as one in the target list and to depress the switch with the "sad face" if the picture was not present in the target list. The test list (targets plus distractors) was then presented. A single random order of the test stimuli was used for all subjects. The investigators recorded the subject's choice and response time for each stimulus. Upon completion of the item recognition task (or the paired comparisons task, depending on task order), the first session was ended. On the average, first sessions lasted from 30 minutes for the adult subjects and second graders to 40 minutes for the preschoolers.

Session II. The second session was conducted for the majority of the subjects within one week of the first testing session. Some of the second graders and adults were tested on the same day as the first session because of scheduling problems. Those subjects who received all four tasks on the same day were given a half hour break between the third task and the last task, the match-to-sample task. None of the preschool subjects were tested completely on the same day.

The room set ups were the same as for Session I. Prior to the second session, each subject's data for the paired comparisons task was tallied and the stimuli were rank ordered with respect to their frequency of choice and grouped so that five stimuli fell into a high typical group and five stimuli fell into a low typical group. The assignment of stimuli with tied ranks to either the high or low groups was made through the use of a random numbers table. This situation occurred fairly frequently in all grade levels with 100% of all preschoolers having at least one tied rank, 83% of all second graders, and 68% of the college students. Only nine subjects (two second graders and seven college students) demonstrated perfect rank ordering of the stimuli.

In addition to using frequency of choice to assign stimuli to the high and low groupings, the stimulus which received the greatest frequency of choice in the paired comparison task for a given subject was designated as the standard to be used in the match-to-sample task for that subject. If tied ranks occurred for the stimuli with respect to the greatest frequency of choice, one of the tied stimuli was chosen at random to be the match-to-sample task standard. This situation occurred for 30% of the preschool subjects, 33% of the second grade subjects, and 14% of the college students.

Once the standard and the high and low typicality ratings were established for a specific subject, the stimuli were then assigned to five different types of stimulus pairs, with 10 pairs being generated for each stimulus pair type. The stimulus pair types represented all combinations of high typical stimuli, low typical stimuli, and stimuli representing a category (e.g., vehicles or cars) unrelated to the categories of main interest. Thus, the stimulus pair types used were high/high pairs, low/low pairs, high/low pairs, high/unrelated pairs, and low/unrelated pairs. Two randomized pair presentation orders were generated with approximately half of the subjects at each age and category level receiving each of the presentation orders.

In order to introduce the subject to the new task situation (as in the description of the first session, the child instructions will be presented for simplicity), the investigator reminded the subject of how he/she had helped the spaceman to learn about earth things during the previous session. The subject was then told that the spaceman was in need of some additional help since he was not sure he knew enough about the category in question. Once it was established that the subject did remember something about the first session, the investigator began the practice task in the following manner:

> "In order to help Zoltar this time, I am going to show you some pictures, three pictures at a time. The picture that will be in this window will always be the same. (The investigator pointed to the top opening.) The pictures you will see in these windows (pointed to bottom openings) will change a lot. Let's see how it works. Zoltar knows that this is a picture of FOOD (revealed practice standard). Zoltar isn't quite sure about these pictures (revealed practice exemplars). You can help Zoltar to learn more about FOOD by pressing the button under the picture (pointed to exemplars) that you think is the <u>most like</u> the picture in the top window (pointed to standard). Let's look at two more pictures. Pick the picture that is <u>most like</u> the top picture (pointed to standard and provided necessary feedback). How about these pictures? (Continue with practice making sure child understands task and that at least one stimulus pair contains a non-exemplar.)"

The subject was presented with four practice trails, which included one trial where one of the pair members was an exact duplicate of the standard.

Following the practice trials, the subject was told that the spaceman really needed extra help in learning about the category which had been used in the first session (i.e., animals or dogs).

The investigator then revealed the subject's standard and told the subject that the spaceman was certain that the standard was a category member, but that he was not sure about the category membership of some of the other pictures. The subject was reminded of the instructions given during the practice task. It was emphasized that he/she select the picture which he/she thought was most like the standard even though most of the pictures would not be exactly the same as the standard. The subject was asked to tell the investigator what the task was and once it was certain that the child understood the task, the investigator began the test pairs. The subject's stimulus choice and time to respond was recorded for each stimulus pair. In addition to this information, justifications were elicited (or it was attempted to elicit them) if and when a subject selected an unrelated category exemplar over a category exemplar. Upon completion of the match-to-sample task, the second session was ended. The average time to complete the match-to-sample task was 15 minutes for the adults and second graders and 25 minutes for the preschoolers.

While the instructions for the tasks described above were basically the same for both child and adult subjects, there were some differences. First, the adult subjects did not receive the practice tasks, although clarifying examples were given. Second, the tasks were not presented in the context of the spaceman story. The adult subjects were told that one way to put the tasks into a context was to imagine they were trying to teach someone about a category of which that person had little or no knowledge.

## CHAPTER IV

## RESULTS AND DISCUSSION

### Age and Category Differences in Category Knowledge

The first set of analyses refer to age level and category level differences in performance on the item generation task. Table 2 presents the mean number of items generated (not including repetitions or obvious miscategorizations, such as "water" as an example of a dog) to the category labels "animals" and "dogs". All subjects were asked to provide responses for both of the category labels.

Differences in overall category knowledge. An Age Level (3) by Category Level (2) analysis of variance was performed on the generation scores for each of the category labels presented. Since sex differences have not been found in previous studies using an item generation procedure, sex of subject was not entered as a factor in this set of analyses. For the category label, "animals," the analysis yielded a significant main effect of age, $F$ (2, 48) = 42.35, $p < .001$ (mean number of items generated being 8.2, 15.83, and 21.73 for preschoolers, second graders, and adults, respectively. The remaining main effect and interaction were non-significant, $p > .05$. Individual two-tailed $t$-test comparisons showed that the preschool children generated significantly fewer animal names than

Table 2

Mean Number Items Generated in the Item Generation Task for Each Age and Category Level

| | CATEGORY LEVEL | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Superordinate | | | | Basic | | | |
| | Animals | | Dogs | | Animals | | Dogs | |
| Age Group | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Preschool | 9.20 | 3.52 | 1.90 | 1.37 | 7.20 | 2.20 | 1.90 | 1.37 |
| Second Grade | 15.00 | 8.69 | 3.83 | 3.97 | 16.67 | 5.16 | 3.83 | 2.79 |
| Adults | 20.36 | 4.70 | 10.09 | 4.93 | 23.09 | 4.48 | 12.55 | 4.25 |

did either the second grade children or the adults, $t$ (30) = -4.38, $p$ < .001, and $t$ (40) = -11.27, $p$ < .001, respectively. Second grade children also provided fewer animal names than did the adult subjects, $t$ (32) = -2.95, $p$ < .01.

A significant main effect of age was also obtained for the generation scores for the category label, "dogs," $F$ (2, 48) = 42.28, $p$ < .001, (mean number of items produced being 1.9, 3.83, and 11.32 by preschoolers, second graders, and adults, respectively. As was the case in the analysis for the category label, "animals," all other main effects and interactions were non-significant, $p$ > .05. Individual two-tailed $t$-test comparisons also showed that the preschool children produced significantly fewer dog names than did the second graders, $t$ (30) = -2.24, $p$ < .05, or than the adults, $t$ (40) = 5.78, $p$ < .001; and second grade children produced significantly fewer dog names than did the adults, $t$ (32) = -4.89, $p$ < .001.

Additional $t$-test comparisons were made to assess whether, at each age level, there were differences in the ability of subjects to produce appropriate items for each of the category labels. Since there were no significant differences in category item production scores for the two different category levels (thus indicating similar levels of categorical knowledge for subjects in both the basic level and superordinate level category conditions at each age level), the item generation scores were summed across the category level factor for each of the three age groups. At each age level, subjects were able to produce a greater number of appropriate category instances when presented with the superordinate category label, "animals," than when they were asked to provide instances for the basic level category label, "dogs"; $\underline{t}$ (19) = 9.36, $\underline{p} < .001$ for preschoolers; $\underline{t}$ (11) = 8.95, $\underline{p} < .001$ for second graders; and $\underline{t}$ (21) = 13.47, $\underline{p} < .001$ for the adult subjects.

Thus, as in previous studies using an item generation task (Nelson, 1974; Rosner & Hayes, 1977; Sullivan, 1979) older children and adults were able to produce more category items from their semantic knowledge base than younger children. It was also demonstrated that, regardless of age, subjects were able to produce more category instances when given a superordinate level category label to provide category examples for, in this case "animals," than when given a basic level category label for item production, "dogs". While this finding may at first appear to be contrary to the general theoretical position of the present study (i.e., that performance should be better when working with basic level categories than when working with superordinate level or subordinate level categories), the finding is not altogether unexpected given the nature of the task at hand. That is, when a person is presented with a category label for item generation, the instances (names) that the person produces represent the next, more specific level of category inclusion in the categorical hierarchy being assessed. In other words, asking for item generation to a superordinate level label requires knowledge of basic level category terms or names, whereas the ability to produce appropriate category instances for a basic level category label requires knowledge of more specific, subordinate level category terms or names. Subordinate level category knowledge does not appear to be present to any great extent nor does it appear to be used by very young children. It is therefore easier to provide basic level names for items generated than if the names required for generation are at the subordinate level. Similar findings have been reported by Anglin (1977) and Nelson (1974).

Measures of category membership agreement. Two related indices of category membership agreement (based on Nelson, 1974) were also calculated for the item generation scores obtained for the two category labels, "animals" and "dogs". As in the above mentioned *t*-test analyses, the scores used to calculate each index were summed across the category level condition. The first index calculated was a type/token ratio (TTR). This index reflects the diversity of responses given for a given category label. The TTR is defined as the total number of different items produced, divided by the total number of responses given for a specific category label. In general, the higher the proportion as represented by the TTR, the greater the category's diversity.

The second index used, the index of commonality (IC), reflects the degree of agreement among subjects on the strongest or most popular items produced for each category label. The IC is defined as the frequency of the three most popular responses divided by the total number of responses given for the first three responses given by each subject for a specific category label (Nelson, 1974). The higher the proportion as represented by the IC, the greater the agreement among subjects as to what are the strongest category items for that specific category. The results of the calculation of the two indices for each age group and category label are presented in Table 3.

Table 3

Type/Token Ratio and Index of Commonality Measures Based on the Item Generation Task for Each Age and Category Label

| | Category | | | |
|---|---|---|---|---|
| | Animals | | Dogs | |
| Age Group | T/T Ratio* | IC** | T/T Ratio | IC |
| Preschool | .33 | .57 | .72 | .29 |
| Second Grade | .47 | .75 | .50 | .64 |
| Adults | .24 | .98 | .21 | .77 |

$$\text{*Type/Token Ratio (T/T)} = \frac{\text{Total number of different items}}{\text{Total number of responses}}$$

$$\text{**Index of Commonality (IC)} = \frac{\text{Frequency of the 3 most popular responses}}{\text{Total number of responses given for the first 3 choices generated by each subject}}$$

As can be seen in Table 3, there is little difference in the amount of diversity present in the categories, "animals" and "dogs" among the second graders and adults. However, a large difference in TTR scores was found at the preschool level. There also appears to be a decrease in the amount of category diversity with age, especially with respect to responses given for the category label, "dogs".

The difference in the measure of category diversity shown by the preschoolers may be the result of the type of responses which the children gave at this age level. In general, when preschool children were asked to name all of the "animals" they could think of, they would provide typical animal names in the same manner as the older children and adults (e.g., "cat," "dog," or "horse"). On the other hand, unlike the older children and adults, preschool children seldom produced typical names for the category instances for the "dog" category (e.g., "collie, " "sheepdog"). The responses of the preschoolers to this category label often took the form of "adjective plus dog". For example, "big dog," "baby dog," or "mamma dog" would be produced by the preschool-age child. While this type of responding may not appear at first to be technically correct (since an assumption underlying the task is that the common nouns of the language will be produced as category instance names), this type of responding does represent a form of category knowledge, although it represents a type of knowledge that is less specific (perhaps less well structured; see Chi & Koeske, 1983) and refined with respect to subordinate terminology than is the knowledge demonstrated by the older children and adults. This type of responding has also been reported and scored as correct in an item generation task in the protocols of Rosner and Hayes (1977) and in the same study, this type of responding was found to be judged by adults as being representative of appropriate category responses.

With respect to the IC measure, it can be seen in Table 3 that the overall level of agreement among the subject as to which category instances are the strongest or most popular examples of their respective categories is, in general, greater for the instances generated for the superordinate level category label than was the degree of agreement obtained for the instances generated for the basic level category label. The degree of agreement among subjects for each of the category labels also showed an increase with age. The finding of greater agreement for responses given to a superordinate level category label corresponds to the finding that it is easier to provide instances for a superordinate level category label since the appropriate response will be basic level in nature, a category level for which a knowledge base appears to be fairly well established even at a young age. The lack of agreement obtained for the responses given to the basic level category label most probably represents a greater degree of idiosyncratic knowledge in the semantic knowledge of individuals, especially for the younger children (who also have less experience interacting with the environment than either the school-aged children or the adults). Even for the adult subjects, the amount of agreement obtained for instances of the basic level category label was not perfect (unlike the high amount of agreement in responses for the instances given for the superordinate level category label), since the knowledge required to provide appropriate responses to a basic level category label is quite specific and specialized in nature. Thus, the difference in agreement in the responses given for superordinate level and basic level category labels may reflect differences in subject area expertise among the subjects. College students do not have much time or reason to have a large knowledge base for breeds of dogs, unless they are dog lovers or perhaps veterinary students interested in specializing in small animal practice. When the knowledge necessary for appropriate responding is at a more general level, the degree of agreement as to what the most popular category items are, increases.

Summary. The results of the item generation task demonstrated that even young children are able to verbalize their knowledge of a specific category. The results also showed that the amount of knowledge demonstrated by a subject (as measured by the number of appropriate category items produced) increases with age (and perhaps also with experience). The category level of the response terms to be produced also had an influence on the number of items produced for a given category label. Responses representing basic level category terms (i.e., the responses required for appropriate responding to a superordinate level category label) were easier to produce and more abundant than were the responses to the category label which required knowledge of a more specific and specialized nature (i.e., responses given for a basic level category label, which represent subordinate level category terms). Despite the fact that the number of category instances produced increased with age, the diversity shown in those catgeory responses decreased with age, especially with respect to the responses given to the basic level category label. Also, agreement as to which category instances are the strongest (most typical?) of a given category label increased with age. The reciprocal relationship between category diversity and agreement on category membership may reflect decreases in the use of idiosyncratic category knowledge, plus a growing, consensual category knowledge base based on increased experience with the environment and through various cultural institutions, such as the formal educational system.

The replication by the item generation task used in this study of several findings (especially the finding that young children have verbalizable category knowledge) reported by previous investigators in the area (e.g., Nelson, 1974; Rosner & Hays, 1977), now sets the stage for the assessment of the ways in which people of different ages utilize category knowledge in different task settings. Also, the use of an item generation task in the present study has demonstrated that the amount of verbalizable category knowledge and the degree to which people agree on the contents of that knowledge does vary not only with age, but also as a function of the level of specific category occupied in a categorical hierarchy.

## Age and Category Level Influences on Cognitive Performance

The following sets of statistical analyses were conducted to assess the possible ways in which age, category level, and category item typicality would affect different types of cognitive performance, in particular, the rank ordering of categorical stimuli, recognition memory, and decisions concerning category membership. Since the results of the paired comparisons task (PC) are directly related to the way in which the match-to-sample (MTS) task was structured for each subject and to the way in which certain measures which were obtained in the recognition memory task were analyzed, the results of the PC task will be presented first, followed by the results for the recognition memory task, and then those results obtained for the MTS task. Finally, the relationship between results obtained in the PC task and those obtained in the MTS task will be explored.

Paired Comparisons Task Performance

As may be recalled from previous descriptions of the PC task, the PC task was designed to provide a rank ordering of the experimental stimuli for each subject by forcing the subject to make a choice between the members of item pairs composed of category exemplars. The frequency of choice for each individual category exemplar was then used to construct a rank ordering for each subject. The five most frequently chosen instances for each subject were considered to represent the high typical category instances for that subject, with the remaining five category items being considered as representing category items of low typicality. If ties occurred in the frequency with which particular items were chosen and when those ties did not allow for a clear assignment of those items to either the high or low typical groups, the tied items were then randomly assigned, using a table of random numbers, to one or the other of the different typicality groupings. The frequency of choice of individual category items by each subject was also used as a means of designating the standard or "prototype" that was to be used in the MTS task. This standard or "prototype" was defined as that category item selected most frequently by a particular subject. Mervis and Rosch (1981) have previously put forth the notion that frequency of choice may be used as an indication of the category typicality of a category exemplar.

The first area of interest is concerned with the ability of subjects to provide a rank ordering for the experimental stimuli. Of the 54 subjects in this study, only nine subjects demonstrated perfect rank ordering of the experimental stimuli based on the frequency of item choice, two second grade children and seven adults. In other words, only nine subjects had protocols for which each stimulus received a different frequency of choice. Both of the second grade subjects, and four of the adult subjects were in the superordinate level category condition, with the remaining three adults being in the basic level category condition. The majority of the subjects (100% of the preschool children, 83% of the second grade children and 68% of the adults) had at least two category instances with tied ranks. Preschool children averaged 6.25 tied items (range = four to nine tied items), second grade children averaged 6.25 tied items (range = zero to nine tied items), and adults averaged 3.64 tied items (range = zero to eight tied items). No distinct pattern in the distribution of where the tied ranks occurred within the orderings was evident at any age or category level. There were no differences in performance in any of the remaining experimental tasks based on the number of ties in rank ordering shown by a given subject.

An analysis of variance for ranked data with ties, was performed at each age/category level combination. This type of analysis yields a chi-square statistic and provides an assessment of whether there are any differences in the way subjects rank a particular set of stimuli (Winer, 1971). Of the six analyses conducted, only the adult subjects demonstrated the ability to provide a differentiation among stimulus items: $\chi^2_{.99}(9) = 50.76$ for adults in the superordinate level condition and $\chi^2_{.99}(9) = 69.70$ for adults in the basic level condition. The mean rank obtained for each stimulus with this procedure showed that the adults ranked the stimuli in such a way that two clusters of five stimuli were formed (i.e., five stimuli received mean ranks indicative of high category typicality and five stimuli received mean ranks indicative of low category typicality). The stimuli that made up the high typical cluster and the stimuli that made up the low typical cluster for the adults at both category levels corresponded exactly to the experimenter-defined stimulus groupings that appear for the target items in Table 1. On the other hand, children at both age and category levels tended to make little distinction among the individual stimulus items (i.e., the mean rank for each stimulus was approximately the same). The mean ranks of the target stimuli of the second grade children did begin to resemble the pattern evidenced by the adults, but the number of subjects in each category level condition was too small for the analyses to reach significance.

A Kendall coefficient of concordance (W) was also calculated based on the rank orderings obtained for the subjects at each age/category level combination. The Kendall coefficient of concordance is an index which is used to assess the extent to which a group of subjects agree in their preferences for various stimuli in a stimulus grouping. The index corresponds to a correlation ratio (Winer, 1971). In accordance with the ANOVA results cited above, preschool children at both category levels ($\underline{W}$ = .15 for superordinate level category subjects and $\underline{W}$ = .14 for basic level category subjects) and second graders at both category levels ($\underline{W}$ = .26 for superordinate level category subjects and $\underline{W}$ = .21 for basic level category subjects) showed little agreement in their preference in choosing the stimulus items. In contrast, the adult subjects at both category levels showed a greater degree of agreement with each other ($\underline{W}$ = .51 for the superordinate level category subjects and $\underline{W}$ = .70 for subjects in the basic level category condition); with a $\underline{W}$ = 1.00 denoting perfect agreement, especially with respect to their preferences at the basic level or categories.

A second area of interest is concerned with the ability of subjects to establish a standard or "prototype" using a paired comparisons technique. Unlike the results with respect to providing a complete rank ordering of the experimental stimuli, the majority of subjects at each age level, were able to designate one category exemplar as a standard or "prototype" (i.e., a stimulus receiving the greatest frequency of choice and not being tied in rank with any other stimulus). The percentage of subjects at each age level who did designate a unique prototype was 70% for the preschool children, 67% for the second grade children, and 86% for the adult subjects. Thus, even the youngest children were able to select one of the stimulus items as being the best example of a catgeory, even though they were not able to make further differentiations among the remaining stimuli.

Table 4 presents the number of subjects at each age and category level who chose a specific target stimulus as the standard or "prototype," which was then used in the MTS task. It can be seen from the table that, with age, the number of different target items chosen as a "prototype" decreased, possibly indicating a greater consensus among people as to what a "good" or "best" example of a catgeory is. This appeared to be especially true for adults in the basic level category condition. In this group, the majority of subjects were split between only two of the stimuli as their "prototype" choice. It should also be noted that, in the distribution of the stimuli chosen as "prototypes," all of the choices made by the adults at both category levels and by second grade children in the superordinate level category condition, represented category items which were rated as having high typicality as established in the experimenter-assigned stimulus groupings. Second grade children in the basic level category condition and preschool children at both category levels tended to have "prototypes" which were distributed among both high typical and low typical category items. These differences in the type of stimulus designated as a "prototype" may be indicative of true differences in perceived item typicality from the type of ratings of item typicality established by the experimenter or they may represent idiosyncratic responding on the part of the subjects, based on some other dimension than category representativeness, such as personal preference or stimulus novelty. For example, the preschool child whose standard was "anteater" indicated never having seen a picture like that before, yet the child insisted that it was

Table 4

Frequency of Choice for Each Target Stimulus as Prototype in the Match-to-Sample Task for Age and Category Level

| | P | S | A |
|---|---|---|---|
| I. Dogs | | | |
| German Shepherd | 2 | 1 | 0 |
| Great Dane | 1 | 2 | 0 |
| Irish Setter | 2 | 0 | 5 |
| Malamute | 0 | 1 | 5 |
| Collie | 1 | 0 | 1 |
| Yorkshire Terrier | 2 | 1 | 0 |
| Shi Tzu | 1 | 0 | 0 |
| Greyhound | 1 | 0 | 0 |
| Afghan Hound | 0 | 0 | 0 |
| Pekingese | 0 | 1 | 0 |
| Total number different prototypes | 7 | 5 | 3 |
| II. Animals | | | |
| Horse | 1 | 2 | 4 |
| Rabbit | 3 | 1 | 2 |
| Fox | 0 | 0 | 0 |
| Tiger | 1 | 2 | 3 |
| Bear | 1 | 1 | 2 |
| Cardinal | 0 | 0 | 0 |
| Anteater | 1 | 0 | 0 |
| Llama | 0 | 0 | 0 |
| Alligator | 2 | 0 | 0 |
| Frog | 1 | 0 | 0 |
| Total number different prototypes | 7 | 4 | 4 |

P = Preschool
S = Second Grade
A = Adults

the stimulus which would help the "spaceman" to learn the most about the category, "animals".

Summary. The results of the paired comparisons task indicate that young children do not appear to be able to rank items according to how representative the items are of a given category, either at the superordinate or the basic level of category inclusion. Except for the finding that even the youngest children are able to select one category example as a "prototype" or "best" example" or a category, preschool and second grade children tended to treat the remaining category items as if they were equally representative of the category, although by the second grade, the children did begin to show an adult-like ordering of the stimuli (i.e., forming two distinct clusters of stimuli, one cluster consisting of items judged to be highly typical of a category and one cluster consisting of items judged to be less typical or atypical of a category). This finding appears to be in contradiction to the hypothesis that even very young children should be able to provide an ordering or consistent categorical structure for a given set of category examples, especially if the examples are representative of a basic level category, and thus, would appear to support a more traditional Aristotelian approach to categorization. The finding is also in conflict with previous studies, such as those of Thompson and Bjorkland (1981) and Mervis and Pani (1980), which have demonstrated that young children are able to judge and provide a category structure based on category item representativeness for both superordinate level and basic level categories. Mervis and Pani (1980) did report one finding which is similar to the findings obtained in the present study. These investigators found that for the youngest children in their sample, the characteristic category structure for a basic level category tended to have well defined "very good" and "very poor" examples for the category, but the structure was less well defined for category items that were representative of moderate levels of item typicality. In fact, the authors reported that the children often refused to make judgments with respect to the typicality of the moderately typical category items.

Thus, while the results of the paired comparisons task do not support a strong form of the hypothesis concerning the ability of young children to provide an internal category structure for categories at different levels of inclusion, a less constrained form of the hypothesis, based on the results of the Mervis and Pani (1980) study, may be put forth and supported. That is, young children are able to select the item they perceive as being the most typical example of a category, but unless the remaining category items are radically different from the "prototype," or atypical of the category in some way (e.g., size, shape), the remaining category items will not be judged as being different from each other with respect to their category representativeness. A great amount of experience with the members of a category may be what is necessary before fine distinctions based on category representativeness can be detected and used (i.e., internal structural differentiation of a category).

With respect to the type of category item that is selected by a subject as a "prototype," it was demonstrated that preschool children were as likely to select exemplars that were rated by adults as being low in item typicality as they were to select items rated by adults as highly typical of a category. On the other hand, second grade children and adults, in general, selected as their "prototype" those category instances previously rated as being of high typicality, again demonstrating an increasing ability to detect finer degrees of category representativeness and a decrease in the use of idiosyncratic knowledge. The transition from an undifferentiated to a differentiated category structure is consistent with the finding of McCauley et al. (1976) that category membership produced priming in a stimulus naming task for second grade children, but not for kindergarten children. Finally, it appeared that with age people tend to be in greater agreement as to what a "good" example of a category should look like.

## Recognition Memory Performance

Before proceeding to a presentation of the results of the item recognition task, a word must be said about the possible effects of task presentation order on cognitive performance. It was decided to counterbalance the order of task presentation of the PC task and the recognition memory task in order to assess whether the repeated presentation of the target stimuli which occurred in the PC task would affect performance on the recognition memory task in contrast to the situation in which the subject did not have the benefit of repeated presentation of the target stimuli prior to being asked to recognize the stimuli in the recognition memory task.

An Age Level (3) x Category Level (2) x Task Presentation Order (2) analysis of variance was performed for each of the following recognition memory measures: Percent correct acceptance of target items (or % Hits), percent of false alarms, and *d'*. The results of the analyses showed that the order of task presentation did not have a significant effect on recognition memory performance for any of the three measures. Thus, the factor of task presentation order was dropped from the rest of the analyses conducted.

The first set of analyses for the recognition memory task to be presented involve three measures of task performance: % Hits, % False Alarms, and *d'* (defined for the purposes of these analyses as the % Hits minus the % False Alarms). Table 5 presents the mean performance score for each of these measures at each age and category level. An Age Level (3) x Category Level (2) x Sex (2) between-subjects analysis of variance was performed for each of the recognition measures. The analysis for the dependent variable, % Hits, produced a significant main effect of age, $F\ (2, 42) = 5.54$, $p < .01$. All other main effects were non-significant, $p > .05$. Thus, it does appear that the ability to recognize target stimuli correctly in a simple recognition memory task increases with age, although the possibility does exist that this result may reflect an increase in response bias to say yes to anything. This issue will be discussed later in this paper.

The % Hits analysis also produced a significant Category Level x Sex interaction, $F\ (1, 42) = 4.44$, $p < .05$; and a significant Age Level x Category Level x Sex interaction, $F\ (2, 42) = 4.88$, $p < .02$. The interactions are shown in Figures 2 and 3. With respect to the

Table 5

Recognition Memory Task Performance for Age and Category Levels

| | Category Level | | | | | |
|---|---|---|---|---|---|---|
| | Superordinate | | | Basic | | |
| Age Group | % Hits* | % FA** | d' | % Hits | % FA | d' |
| Preschool | .81 | .24 | .58 | .80 | .22 | .58 |
| Second Grade | .82 | .03 | .78 | .83 | .02 | .82 |
| Adults | .95 | .01 | .95 | .91 | .02 | .89 |

* Hits = correct acceptance of target stimulus

** FA = false alarms

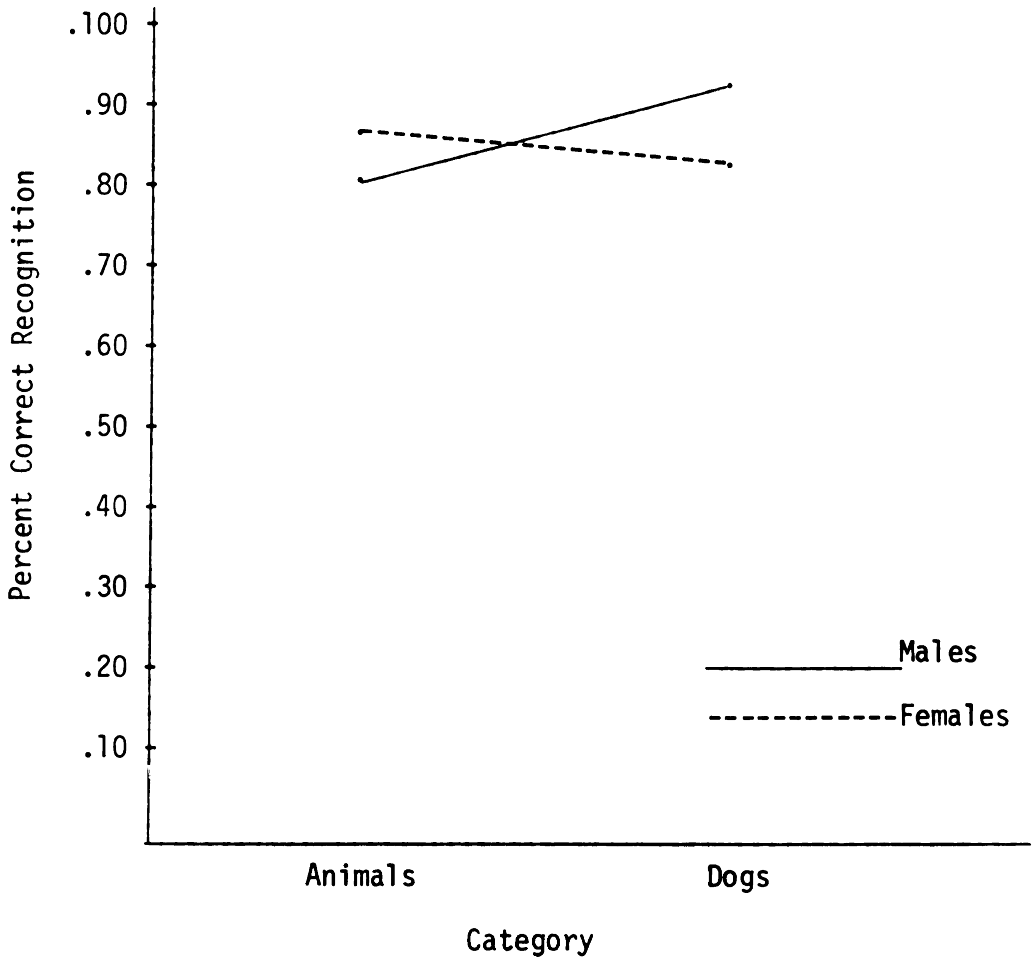


Figure 2. Category x Sex Interaction for Recognition

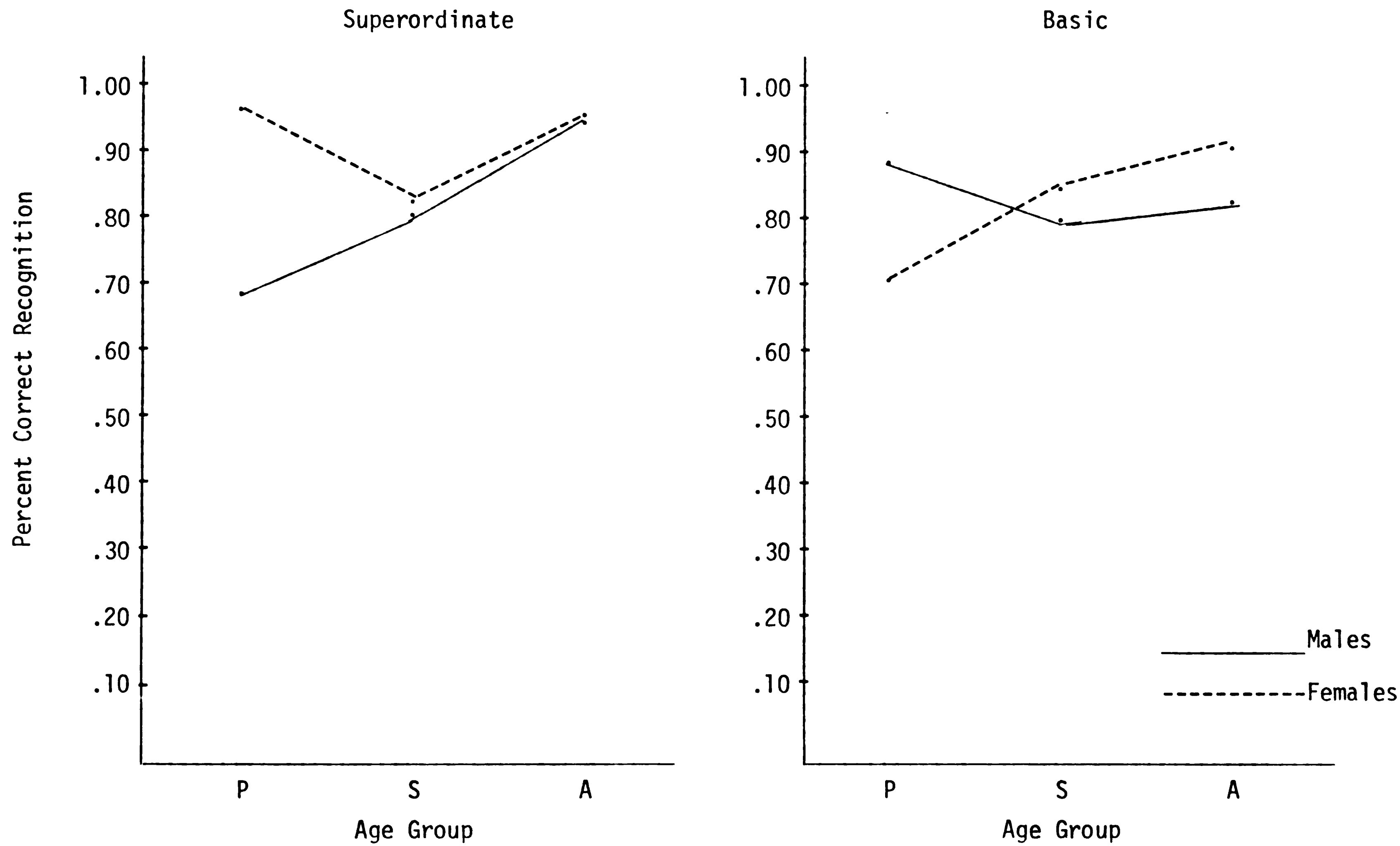


Figure 3. Age x Category x Sex Interaction for Recognition

two-way interactions, males and females appear to recognize items from a superordinate level category approximately equally well, with males possessing better recognition for items in the basic level category. The three-way interaction provides some clarification for the relationship between the differences in recognition memory between sex and category levels. It can be seen that only at the preschool level are there sex differences in the ability to recognize items differing in their level of category inclusiveness, and the difference in recognition ability only reaches significance at the superordinate category level, $\underline{t}$ (8) = 2.53, $\underline{p} < .05$ (in favor of females), although the difference in recognition ability at the basic level of categorization is only marginally close to significance, $\underline{p} < .10$, in favor of the males.

The next two analyses, using the percent of false alarms and $\underline{d}'$ as dependent variables, both showed significant main effects of age, $\underline{F}$ (2, 42) = 11.60, $\underline{p} < .001$ for false alarms and $\underline{F}$ (2, 42),= 15.20, $\underline{p} < .001$ for $\underline{d}'$, with the percent of false alarms decreasing with age level and the sensitivity for accepting target stimuli and rejecting distractor stimuli increasing with age. Thus, a response bias interpretation of the Hit data is unwarranted. The $\underline{d}'$ analysis also produced a marginally significant Age Level x Category Level x Sex interaction, $\underline{F}$ (2, 42) = 3.18, $\underline{p} < .06$ which is presented in Figure 4. This effect is much the same as that seen in the Hit data. Preschool males showed a trend toward greater sensitivity in distinguishing target stimuli from distractor stimuli when presented with items representing a basic level category than did preschool

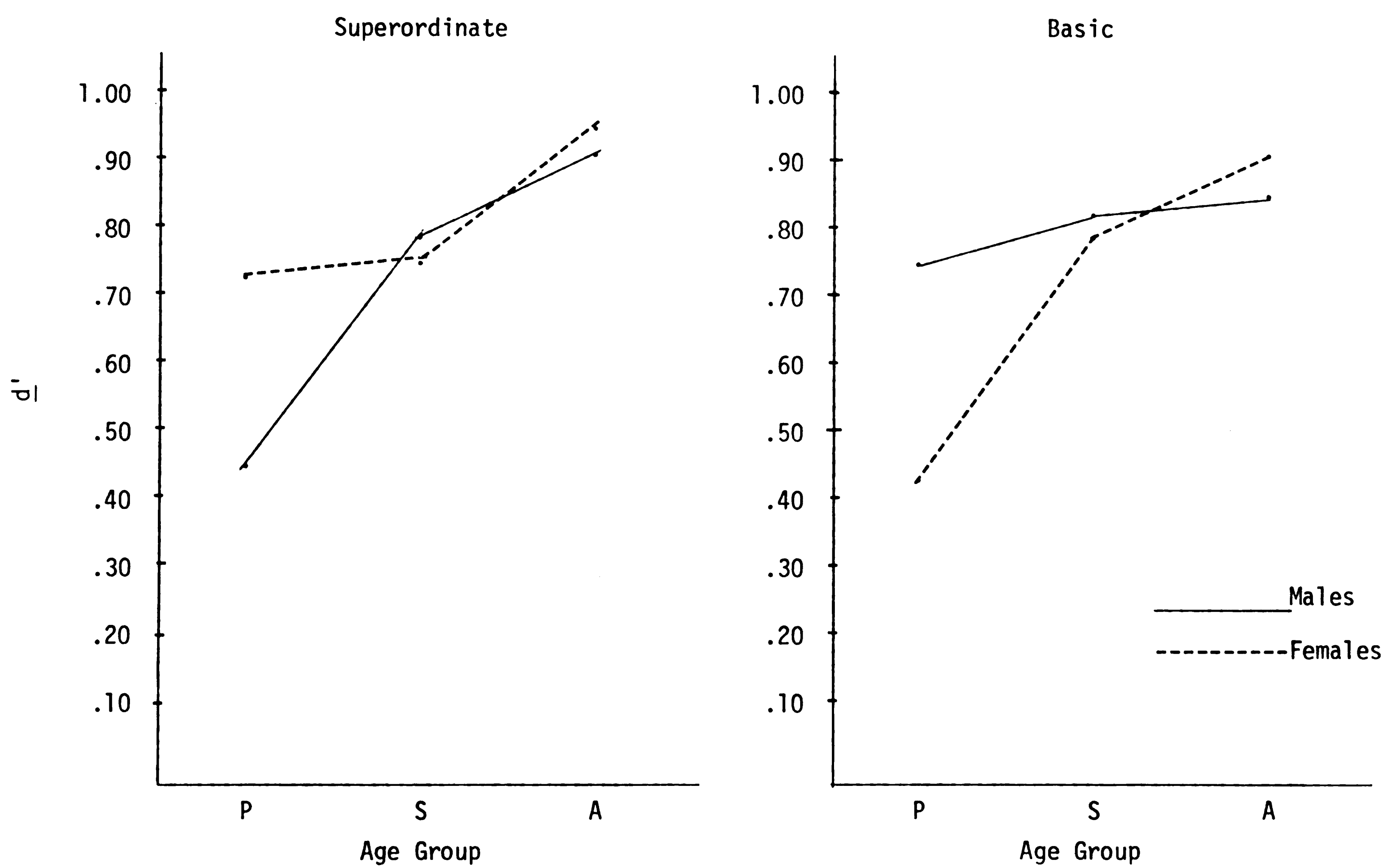


Figure 4. Age x Category x Sex Interaction for Recognition Sensitivity

females, $t$ (8) = 2.11, $p < .07$. The difference in $d'$ scores between preschool males and females for the superordinate level category was not significant, $p > .10$.

It was also of interest to assess the effect of category item typicality on a subject's ability to accept or reject a stimulus item as being a member of the target list in the recognition memory task. For example, would a subject be more likely to correctly accept a target stimulus that had received a high typicality rating than a target item that had received a low typicality rating, since the stimulus with the high typicality rating could be assumed to have a greater category strength and might also be more familiar to the subject than would be the low typical stimulus. Also, would a subject be more likely to accept a distractor item as a member of the target list (i.e., commit a false alarm) if the distractor item is rated as high on the typicality scale than if the distractor item would be assumed to have more in common (e.g., features, attributes) with the high typical target items from the same category and would thus be more likely to be confused with the high typical target items than would the low typical distractor items.

Individual two-tailed $t$-test comparisons were conducted to assess possible differences in the mean percentage of high typical target items and low typical target items correctly accepted by subjects at each age/category level combination who did not show perfect recognition memory performance. Individual $t$-tests were also conducted to assess the difference in the mean percentage of high typical distractor items and low typical distractor items that were false alarms only for the preschool children, since the percentage of false alarms committed by the rest of the sample was very small. The results of these comparisons show that, regardless of the age/category level group of a subject, there were no significant differences in the ability to correctly accept high typical targets over low typical targets as members of the test list or in the tendency to produce a false alarm for a high typical over a low typical distractor item, all $\underline{p} > .10$. Thus, for overall measures of recognition memory performance, unlike the previous findings for recall memory performance (Thompson & Bjorkland, 1981), the internal structure of a category (as reflected by category typicality ratings) did not have an influence on recognition memory ability.

A possible explanation for the inability to find an influence of a category's internal structure on recognition memory performance may be that the measures used in the analyses were not subtle enough in nature to detect any potential influences. That is, while the internal structure of a category does not appear to affect overall recognition memory performance, the internal structure of a category <u>might</u> have an influence on the amount of cognitive processing (as measured by response time) necessary to make a decision concerning the acceptability of a stimulus item as a member of the target list. For example, since the information concerning high typical category items may be assumed to be more readily available in the internal structure of a category for a subject than for low typical items, subjects should be able to respond faster to the high typical targets than to the low typical targets. On the other hand, it should take longer to respond to distractor items of high typicality than to distractor items of low typicality because of the greater degree of category similarity existing among the high typicality distractor items and the high typical target items. For each subject, the mean time to respond to each stimulus type (high typical target, high typical distractor, low typical target, and low typical distractor) was calculated for both the experimenter assigned typicality groupings of the target stimuli and for the typicality groupings of the target items based on each subject's responses in the PC task. Subjects never provided rankings for the distractor items. The times are presented in Tables 6a and 6b.

A four factor analysis of variance, with between-subjects factors of age level (preschool, second grade, and adult), category level (basic or superordinate) and sex and with the repeated measures factor of stimulus type (high typical target, high typical distractor, low typical target, or low typical distractor) was performed on the mean response times obtained in the recognition memory task. The results of the analysis produced a significant main effect of age, $\underline{F}$ (2, 42) = 52.29, $\underline{p} < .001$, indicating a decrease in response time with age, regardless of stimulus type (mean times being 5576 msec. for preschool children, 2500 msec, for second grade children, and 1047 msec. for adult subjects). All remaining main effects, including that of stimulus type, and all interactions were nonsignificant, $\underline{p} > .05$.

Since the standard deviations obtained seemed to be unusually large for this type of data, the possibility that the amount of variation present in the data might be masking potential category level and stimulus type effects was considered. In order to try to reduce the amount of variation, a speed transformation was performed on the response times and the resulting values were submitted to the

Table 6a

Mean Response Times for Recognition Memory Target and Distractor Stimuli Varying in Typicality for the Category, Animals

| | | Animals Age Group | | | | | |
|---|---|---|---|---|---|---|---|
| | | Preschool | | Second Grade | | Adults | |
| Typicality | Stimulus Type | Mean | SD | Mean | SD | Mean | SD |
| High | Target* (experimenter) | 5571*** msec. | 2882 | 2013 | 564 | 1031 | 529 |
| | Target** (subject) | 5580 | 3002 | 1903 | 654 | 1033 | 493 |
| | Distractor | 6514 | 2806 | 1996 | 874 | 1039 | 379 |
| Low | Target (experimenter) | 5639 | 2404 | 2188 | 751 | 1006 | 372 |
| | Target (subject) | 5754 | 2376 | 2312 | 648 | 1005 | 374 |
| | Distractor | 6554 | 2962 | 1702 | 663 | 952 | 366. |

* Typicality based on experimenter-defined ratings

** Typicality based on subject's rank orderings

*** All times expressed in milliseconds

Table 6b

Mean Response Times for Recognition Memory Target and Distractor Stimuli Varying in Typicality for the Category, Dogs

| | | Dogs Age Group | | | | | |
|---|---|---|---|---|---|---|---|
| | | Preschool | | Second Grade | | Adults | |
| Typicality | Stimulus Type | Mean | SD | Mean | SD | Mean | SD |
| High | Target* (experimenter) | 4583*** msec. | 1990 | 3174 | 1757 | 1047 | 541 |
| | Target** (subject) | 4595 | 1515 | 3021 | 2238 | 1035 | 532 |
| | Distractor | 5389 | 2063 | 2616 | 1068 | 1027 | 374 |
| Low | Target (experimenter) | 5010 | 2005 | 2626 | 1376 | 1371 | 840 |
| | Target (subject) | 5059 | 2576 | 2678 | 1163 | 1392 | 834 |
| | Distractor | 4911 | 1657 | 3744 | 2345 | 965 | 386 |

* Typicality based on experimenter-defined ratings

** Typicality based on subject's rank orderings

*** All times expressed in milliseconds

same type of four factor repeated measures ANOVA as were the non-transformed response times. As in the response time ANOVA, the only significant effect was that of age, $\underline{F}$ (2, 42) = 33.78, $\underline{p} < .001$. Thus, as in the analyses of the more general measures of recognition memory performance, the internal structure of a category and that category's level of inclusiveness in the categorical hierarchy did not have an influence on the amount of cognitive processing necessary for efficient recognition memory performance at any age level.

Summary. The results of the recognition memory task provided additional evidence in support of the findings of Meyers and Perlmutter (1978) that even preschool children have extremely good overall recognition memory abilities, although these abilities do continue to improve with age, especially with respect to the sensitivity in distinguishing distractor items from target items in a simple recognition memory task. There also appeared to be some difference in the preschool children's ability to remember items representing levels of categorization. Girls tended to have better recognition scores than did boys when the items to be remembered represented a superordinate level category, whereas boys showed a trend toward better recognition of items representing a basic level category. Boys also displayed greater sensitivity in distinguishing distractor items from target items at the basic level of categorization, with the reverse being shown by the girls at the superordinate level (i.e., girls showed a greater degree of sensitivity than the boys) although the difference was not statistically significant. This difference in recognition memory performance in the preschool children may be due to differences in the children's ability to use verbal versus perceptual types of stimulus information as an aid to memory. The one sex effect which appears to be the most consistent in the literature is that females tend to be verbally superior to males and that males tend to have better visual-spatial abilities than females, (cf., Maccoby & Jacklin, 1974). Thus, in the case of the present study, preschool girls performed the best under conditions for which they had a fairly large and well developed verbal knowledge base which could be used as an aid to memory and for which perceptual discrimination skills were minimized. The boys tended to show superior performance when the decisions to be made depended on the ability to detect small differences in perceptual features in order to perform efficiently (i.e., a basic level category, the members of which are considered to be more perceptually similar than members of a superordinate level category). Further research needs to be conducted with a variety of superordinate and basic level categories before applying this explanation to other types of cognitive performance.

The hypothesis predicting superior performance under basic level category conditions and the ability to be able to utilize a category's internal structure to facilitate memory performance was not supported. Except for the preschool findings discussed above, recognition memory performance (whether for overall or response time measures) was approximately equivalent for both of the category levels studied. The internal structure of a category (as represented by item typicality) also did not appear to exert an influence over recognition memory performance. These findings are in apparent conflict with the results of previous studies such as that of

Thompson and Bjorkland (1981) who found better recall memory for high typical category items than for low typical category items, and that of Rosch et al. (1976) who found faster response times for high typical stimuli over low typical stimuli.

An explanation for these results may reside in the different types of processing necessary for recognition versus recall memory tasks. The processing required in recognition is more passive in nature, requiring, in general, (at least in the case of simple recognition) only the ability to distinguish previously experienced items from novel stimulus items (Bransford, 1979). Since recognition memory performance is good at all ages studied (approaching ceiling performance for adults), the additional information provided by a category's internal structure or its level of inclusion in a categorical hierarchy, may not be necessary for, nor detrimental to, efficient memory performance.

Recall memory tasks, in contrast, require more active processing on the subject's part, since the subject must produce the to-be-remembered items from memory without the aid of strong external cues (i.e., the to-be-remembered items themselves; Bransford, 1979). In this type of situation, it is often helpful to have some additional type of information about the to-be-remembered items that can be used to organize the items prior to and during retrieval or which can be used to integrate the to-be-remembered items into already existing organizational schemes. Knowledge of the different levels of category inclusion and of categorical structure may be useful in pointing to potential organizational schemes that could be useful for remembering and which may even be necessary for efficient memory performance in a recall task.

Match-to-Sample Task Performance

The match-to-sample task (MTS) was designed to provide information concerning within-subject consistency in the rank ordering of the experimental stimuli in comparison to the orderings that we obtained in the paired comparisons task and to provide information about the way in which the internal structure of a category may effect category membership decisions when a subject is presented with a "good" or "prototype" category instance as a standard for making judgments. The first set of analyses to be discussed will refer to measures that are pertinent to the MTS task only. The second set of analyses that are to be presented in this section will deal with the relationship between the PC task performance and the MTS task performance.

The MTS task involved presenting the subject with triads of stimuli, one member of which (the standard or "prototype") remained constant across trials. The remaining types of stimulus pairs which were presented were as follows: high typical-high typical pairs, low typical-low typical pairs, high typical-low typical pairs, high typical-noncategory item pairs, and low typical-noncategory item pairs. The last two pair types were included to determine, especially for the youngest children, whether responses were being made on the basis of the category represented by the standard. While the order of presentation of the stimulus pair types was the same (half of the subjects received one randomly ordered list, half received the same list only in reverse order), the actual members of a given stimulus pair was dependent on the high typical and low typical category groupings obtained for each subject in the PC task. Two types of response measures were obtained for each subject: mean response time to select one member of a stimulus pair as being "most like" the standard for each different pair type and the percent of response errors made for the high typical-low typical pairs and for the two mixed category-noncategory pair types.

An Age Level (5) x Category Level (2) x Sex (2) x Stimulus Pair Type (5) repeated measures analysis of variance, with the repeated measures factor of Stimulus Pair Type, was performed on the subjects' mean response times. The mean response times obtained for each stimulus pair type at each age and category level are presented in Tables 7a and 7b. The analysis yielded a significant main effect of age, $\underline{F}$ (2, 42) = 12.61, $\underline{p} < .001$, indicating the typical pattern of decreasing response time with age (mean response times being 4857 for preschool children, 1596 for second grade children, and 1059 for adults, all times are expressed in milliseconds). All remaining main effects and all of the interactions were nonsignificant, $\underline{p} > .05$. Thus, as in the response time analyses conducted for the recognition memory task data, category level and category item typicality did not have a significant influence on how long it took a subject to select a stimulus pair member as a match for the standard. If the typicality of the pair members had had an influence on response time, high typical-high typical catgeory pairs should have shown the longest response times (since both pair members would be perceived as being very similar to the standard) followed by low typical-low typical stimulus pairs (since both pair members would be assumed to be equally unlike the standard). High typical-low typical category pairs would

Table 7a

Mean Response Times for Match-to-Sample Stimulus Pairs for the Category, Animals

| | Animals Age Group | | | | | |
|---|---|---|---|---|---|---|
| | Preschool | | Second Grade | | Adults | |
| Pair Type | Mean | SD | Mean | SD | Mean | SD |
| High/Low | 4592.60* | 1895.11 | 1900.67 | 990.26 | 1087.91 | 465.78 |
| Low/Low | 4805.70 | 2039.61 | 2183.83 | 1376.10 | 1348.73 | 709.29 |
| High/Low | 4764.20 | 1971.80 | 1490.00 | 525.06 | 1026.45 | 458.20 |
| High/Noncat | 4340.00 | 2819.86 | 1355.83 | 543.94 | 968.82 | 468.39 |
| Low/Noncat | 4464.40 | 2700.81 | 1695.50 | 683.23 | 1311.54 | 948.38 |

* Times expressed in milliseconds

Table 7b

Mean Response Times for Match-to-Sample Stimulus Pairs for the Category, Dogs

| | Dogs Age Group | | | | | |
|---|---|---|---|---|---|---|
| | Preschool | | Second Grade | | Adults | |
| Pair Type | Mean | SD | Mean | SD | Mean | SD |
| High/High | 4786.90* | 1746.70 | 2656.50 | 1850.12 | 993.64 | 343.45 |
| Low/Low | 5018.60 | 1902.31 | 2963.17 | 2052.19 | 1149.27 | 640.43 |
| High/Low | 4257.10 | 1549.81 | 3149.83 | 2168.12 | 980.18 | 344.00 |
| High/Noncat | 4389.80 | 1805.62 | 1772.83 | 907.35 | 813.55 | 389.82 |
| Low/Noncat | 4396.20 | 2127.60 | 2096.67 | 1112.82 | 960.18 | 477.83 |

* Times expressed in milliseconds

have been expected to show the fastest response times of the category pairs if the subject responded on the basis of the category's internal structure. Finally, the mixed category-noncategory pairs would have been expected to show the fastest response rates overall, since the subject would be making a simple category membership decision, (assuming that the subjects did in fact use category information to make their decisions). Times for the basic level category would have been expected to be faster than response times for the superordinate level category pairs. In general, the results did not show even a trend toward this pattern, although the adult subjects did come the closest to such a pattern. Also, speech transformations that were performed on the data yielded the same pattern of results as those obtained using mean response time as a measure.

An Age Level (3) x Category Level (2) x Sex (2) x Stimulus Pair Type (3), repeated measures ANOVA with the repeated measures factor of Stimulus Pair Type, was performed on the percentage of errors made for each of the three mixed stimulus pair types (i.e., high typical-low typical category pairs, high typical-noncategory pairs, and low typical-noncategory pairs). For the high typical-low typical category pairs, an error was defined as selecting the low typical category pair member as being the match for the standard. For the category-noncategory pairs, an error was defined as selecting the noncategory pair members as the match for the standard, thus making a cross category selection. Table 8 presented the mean percentage of errors made by subjects at each age and category level for each type of stimulus pair.

Table 8

Mean Percent Response Errors for Match-to-Sample Stimulus Pairs for Age and Category Level

| | Category Level | | | | | |
|---|---|---|---|---|---|---|
| | Superordinate | | | Basic | | |
| Age Group | HL* | HN** | LN*** | HL | HN | LN |
| Preschool | .40 | .49 | .44 | .24 | .14 | .24 |
| Second Grade | .38 | .00 | .00 | .35 | .00 | .00 |
| Adults | .17 | .01 | .06 | .12 | .05 | .05 |

* HL = High/Low stimulus pairs

** HN = High/Noncategory stimulus pairs

*** LN = Low/Noncategory stimulus pairs

The results of the analysis provided to be more fruitful than the results of the previous analyses reported for the present study. The analysis yielded significant main effects of age, $\underline{F}$ (2, 42) = 20.98, $\underline{p}$ < .001; and stimulus pair type, $\underline{F}$ (2, 84) = 15.10, $\underline{p}$ < .001. The analysis, unlike the previously reported analyses, also yielded a marginally significant effect of category level, $\underline{F}$, (2, 42) = 3.89, $\underline{p}$ < .06. The main effect of sex was not significant. Thus, not only did the tendency to make error decrease with age (the mean percent of errors being 32% for the preschool children, 11% for the second graders, and 7% for the adults), but also, as was hypothesized, the subjects were more accurate in responding correctly to the different stimulus pair types at the basic level of categorization than when the stimulus pair types represented a superordinate level category. The analysis also yielded two significant two-way interactions and a significant three-way interaction. The significant two-way interactions were Age Level x Category Level, $\underline{F}$ (2, 42) = 4.68, $\underline{p}$ < .05; and Stimulus Pair Type x Age Level, $\underline{F}$ (2, 84) = 5.34, $\underline{p}$ < .01. The significant three-way interaction was Stimulus Pair Type x Category Level x Sex, $\underline{F}$ (2, 84) = 3.31, $\underline{p}$ < .05.

Figure 5 depicts the Age Level x Category Level interaction. Simple effects ANOVAs showed that at each category level, the mean percent of errors decreased with age, $\underline{F}$ (2, 24) = 5.80, $\underline{p}$ < .01 for the superordinate level category and $\underline{F}$ (2, 24) = 4.09, $\underline{p}$ < .05 for the basic level category. At the preschool level, children working with stimulus items representing the superordinate level category made significantly more errors in responding than did preschool children at the basic level or categorization,

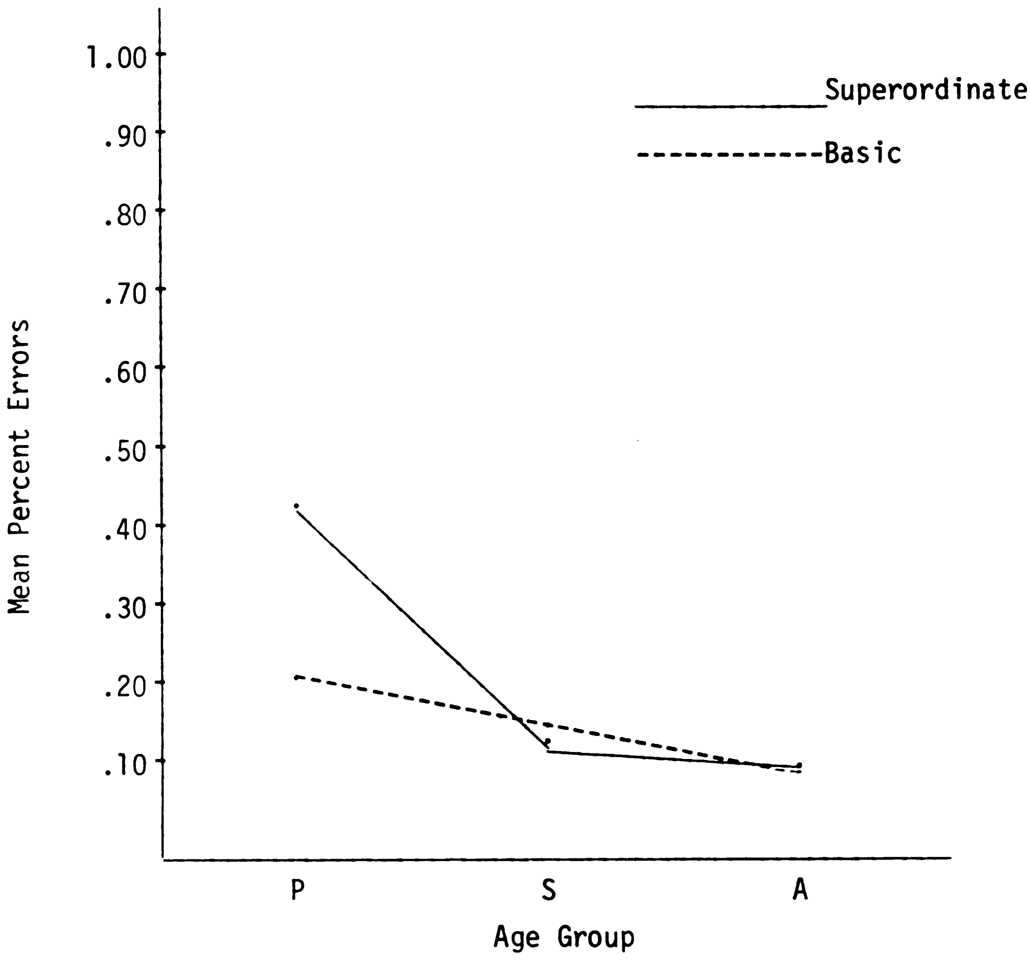


Figure 5. Age x Category Interaction for Match-to-Sample Errors (percent of errors summed across error types)

$\underline{t}$ (18) = 2.95, $\underline{p} < .01$. Second grade children and adults made approximately the same number of errors at each of the two different category levels.

Figure 6 shows the significant Stimulus Pair Type x Age Level interaction for the mean percent of response errors. Adults made fewer errors on the high typical-low typical category item pairs than did either the preschool children or the second grade children (who showed approximately equal error rates), $\underline{F}$ (2, 51) = 8.33, $\underline{p} < .01$. For the two category-noncategory stimulus pair types, the preschool children had a higher percentage of errors on both high typical-noncategory pairs and low typical-noncategory pairs than did the second graders (who had no errors in performance) or the adults (who did not differ in the amount of errors made for each of these two stimulus pair types), $\underline{F}$ (2, 51) = 14.00, $\underline{p} < .01$ for the high typical category-noncategory stimulus pairs and $\underline{F}$ (2, 51) = 23.00, $\underline{p} < .001$ for the low typical category-noncategory stimulus pairs. Adults did not differ significantly from the second grade children with respect to the percent of errors made for the two category-noncategory stimulus pair types. The adults did have a significantly greater percentage of errors made on the high typical-low typical pair types, $\underline{F}$ (2, 42) = 8.00, $\underline{p} < .01$.

The significant Stimulus Pair Type x Category Level x Sex interaction is presented in Figure 7. Simple effects ANOVAs with repeated measures demonstrated that, for the superordinate level category, there were no significant differences in the percentage of errors made for each of the stimulus pair types by males $F < 1$, but that for females there was a significant difference in the

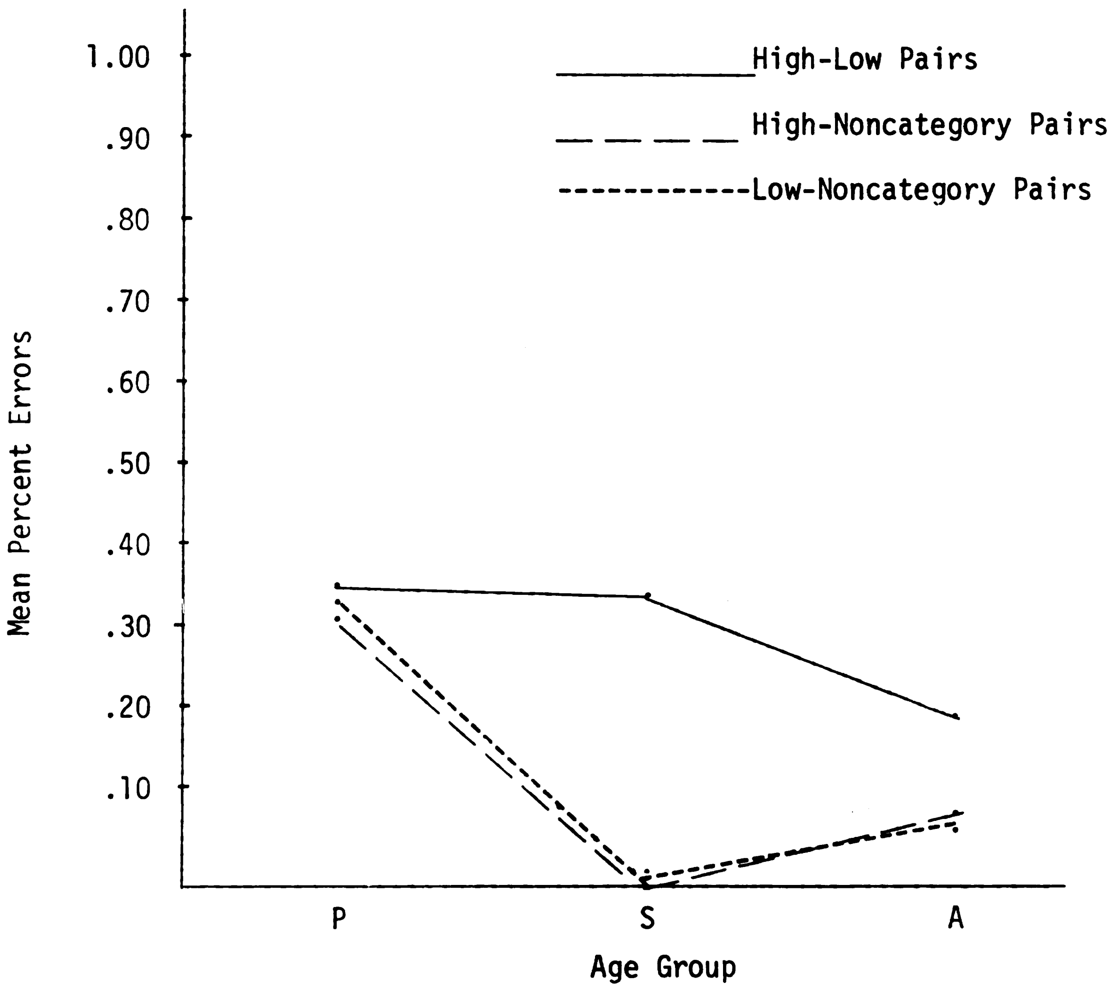


Figure 6. Stimulus Pair Type x Age Interaction for Match-to-Sample Errors

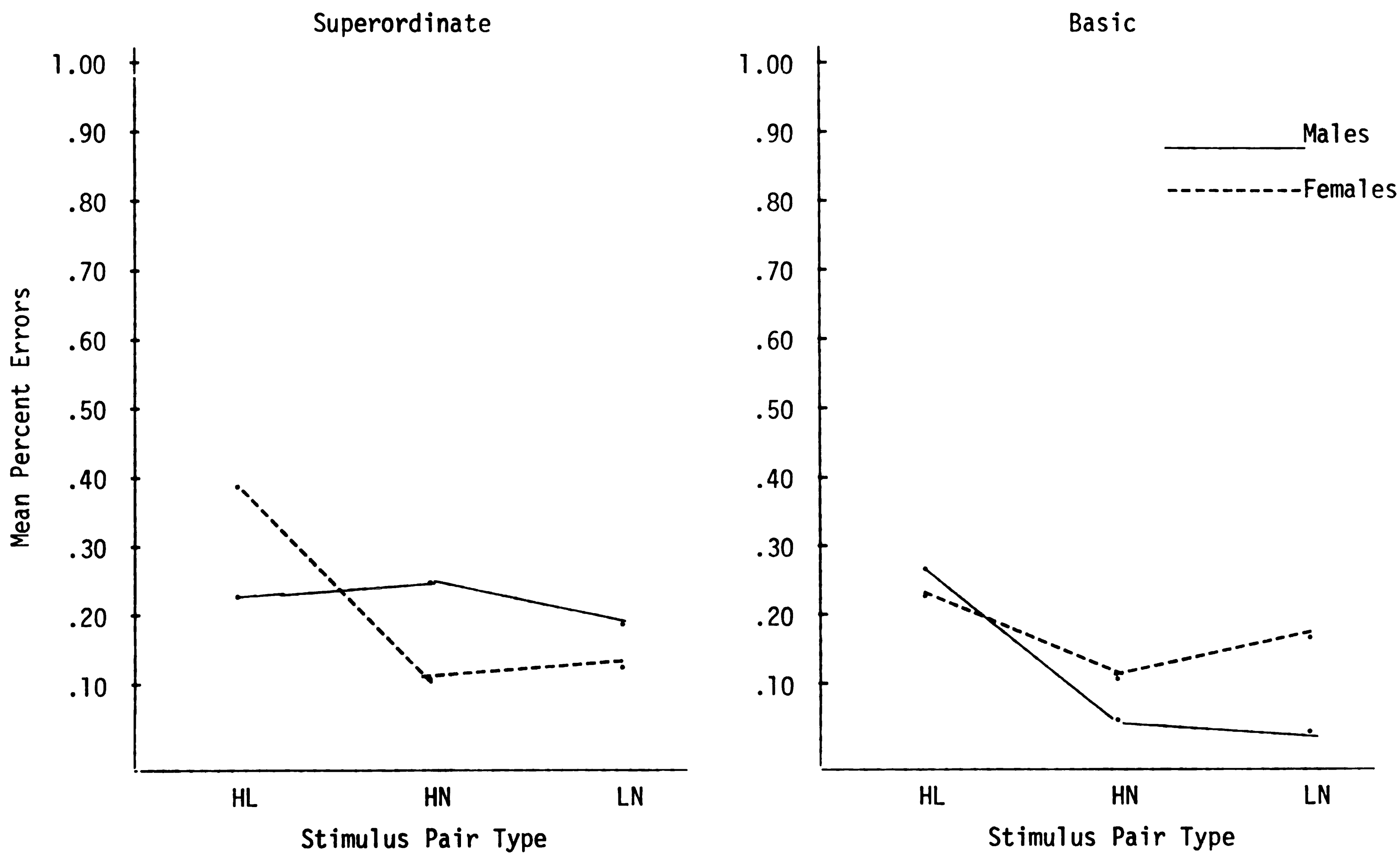


Figure 7. Category x Sex x Stimulus Pair Type Interaction for Match-to-Sample Errors

percentage of errors made for each of the different stimulus pair types, $F$ (2, 26) = 6.00, $p < .01$, with females making more errors on the high typical-low typical category pairs. With respect to performance in the basic level category condition, the opposite held true. That is, there were no significance differences in the percentage of errors made by females on each of the different stimulus pairs types, $F$ (2, 26) = 2.00, $p > .05$. Male subjects, on the other hand, made significantly more errors in responding to the high typical-low typical category pairs than when responding to either of the category-noncategory pairs, $F$ (2, 24) = 8.00, $p < .01$.

The results of the MTS task presented so far have demonstrated that, as was evident in the recognition memory task, that neither the internal structure of a category nor that category's level of inclusiveness in the categorical hierarchy significantly affected the amount of time needed by a subject to make a matching response to a standard. Response time did increase with age and an examination of the mean response times shows that by adulthood, the response times do begin to resemble the expected pattern, but in general, response time for this type of task is not affected by category item typicality or level of inclusiveness.

The response measures reflecting accuracy in responding provided quite a different state of affairs. Not only did response errors decrease with age, but errors were also less frequent at the basic level or categorization than they were at the superordinate level or categorization as would be predicted by a prototype approach to category behavior. This was especially so for the preschool children. Errors were more frequent for the category pairs for both the second grade children and the adults than for the mixed category-noncategory pair types, indicating that their responses were made on the basis of the category knowledge necessary for the category level presented. When an error was made by one of the older children or by an adult for a category-noncategory pair, a justification for the choice was requested. The majority of the justifications were based on a failure to pay close attention to the members of the stimulus pair when making their selection. A few of the justifications indicated a momentary shift in the category for which the decision was based. For example, one of the adult subjects whose standard was "horse," selected the stimulus "car" over the stimulus "alligator," (a low typical category instance for this subject). The justification given for this selection was that the "car" was more similar to the "horse" than the "alligator" because the "car" and the "horse" were both <u>modes of transportation</u> to a greater extent than the "horse" and "alligator" were to being animals.

The preschool children tended to make just as many errors on the category-noncategory stimulus pairs as they did for the category pairs. This might be taken to indicate that the preschool children did not have a good working knowledge of the categories presented. When these children were asked to provide justification for their choices, their justifications were more often than not, nonexistent. When justifications were given, the preschool children indicated that they made their choice on the basis of personal preference. This occurred even though it had been established in the practice task that the children did understand the requirements of the MTS task. It may have been the case that the children's familiarity with the target stimuli (having seen each of the stimuli several times in both the recognition memory task and the PC task) and the novelty of the noncategory stimuli may have also influenced the children's stimulus selection pattern for the category-noncategory stimulus pairs.

It may be the case that males may be more sensitive to visual differences among stimuli than are females. This would be especially important for discriminating among stimuli representing a basic level category, since one must attend to a larger number of features than just those features correlated with category membership, thus accounting for the tendency of males to make some choices that might violate the typicality structure of their basic level categories. In contrast, females may be relatively more sensitive to a more verbalizable, all-or-none type of category knowledge or structure, and would thus, be little affected by subtle visual differences among stimuli.

Relationships Between Paired Comparisons and Match-to-Sample Task Performance

In addition to the response time and error data obtained in the MTS task, the manner in which a subject responded to the stimulus items in the high typical-high typical category pairs and to the low typical-low typical category pairs provided information with respect to the within-subject consistency in the use of the internal structure of a specific category. The frequency with which each of the high typical stimuli was selected as a match for the standard over all other high typical stimuli and the frequency with which each of the low typical stimuli was selected over all other low typical stimuli as a match for the standard were used to construct a rank ordering of the stimuli which could then be compared to the rank ordering of the high and low typical stimulus items obtained in the PC task. The types of response patterns obtained in the PC task made the job of assessing within-subject category structural consistency much more difficult than had been expected, since only nine of the subjects provided a perfect rank ordering of the stimuli in the PC task. Thus, the results to be presented should be considered as being extremely tentative until more sophisticated and sensitive methods are developed to assess structural consistency.

One index of the consistency in the use of a categorical structure is whether a subject's "prototype" remains constant across task situations. Since in the MTS task the subject's "prototype" was always available for inspection and since the subject's "prototype" could also appear in the MTS task as a member of one of the different stimulus pair types, the type of response that would indicate consistency across tasks would be an identity match. In general, identity matches did predominate in the MTS task, but only one group of subjects, adults in the superordinate level category condition, demonstrated perfect consistency in their "prototype" selection across the two tasks (i.e., PC and MTS tasks). Seventy percent of the preschool children at both category levels had the same "prototype" across tasks. For the second grade children, 83% of the subjects in the superordinate level category condition and 67% of the subjects in the basic level category condition showed "prototype" consistency, as did 82% of the adult subjects in the basic level category condition.

What could account for these types of results? At least with respect to the adult subjects, and to the majority of the second grade children, the failure may have been the result of not paying close attention to all of the members of the stimulus triad prior to making their selection. With this type of task, a single error in making the necessary identity match could result in the person having a different "prototype" for the MTS task, even if that person responded correctly on all other trials. It was informally observed that subjects often caught their error in stimulus selection after it had been made and could comment on their mistake. An attentional explanation might also be pertinent for the preschool children, since they often had to be reminded of the task at hand. If an error in making an identity match was pointed out to the child, most of the preschool children would stick to their selection and offer no justification for their selection. It could also be the situation that, even though they were reminded of the requirements of the MTS task, and had demonstrated an understanding of the task during the practice task, the children were responding in a manner more appropriate to the paired comparisons task (that is, making a selection based on the pairs of changing stimuli, rather than making selections based on an item's similarity to the standard) and thereby provided a new ordering for the stimuli.

The high percentage of subjects having tied ranks in the PC task made it extremely difficult to decide on a method to compare that task to the MTS task. Also, since some items in the PC task had to be randomly assigned to the high typical and low typical groupings of the subjects in the PC task because of the presence of tied ranks, a degree of artificiality is present in the data. It was decided to use the number of items placed in a different position in the rank ordering of the MTS task from the position they were in in the PC task ordering as an index of structural consistency. The index was easy to calculate for all of the subjects and could easily handle tied ranks. One third of the sample showed perfect consistency in their ordering of the high typical stimulus items (15% of the preschool children, 8% of the second grade children and 23% of the adult subjects), while only 13% of the total sample provided the same ordering of the stimuli they had ranked as having low category item typicality (0% of the preschool children, 8% of the second grade children, and 27% of the adults). Twenty-four percent of the sample had stimulus orders totally different (all items being in a different position) from that obtained in the PC task for their high typical items (50% of the preschoolers, 42% of the second graders, and 9% of the adults), and 57% of the total sample provided totally different orders for the low typical stimuli (75% of the preschoolers, 67% of the second graders, and 36% of the adults).

Individual two-tailed $\underline{t}$-tests were performed on the mean number of differently ordered items for each age level to assess if any differences existed with respect to category level. Only one test reached significance--second grade children had significantly more items with the same order in both the PC and MTS tasks for the superordinate level category than at the basic level of categorization, $\underline{t}$ (10) = 12.57, $\underline{p} < .05$. When the differences in the rankings of high versus low typical category items were assessed, it was found that only the preschool children (at both the category level conditions) were more consistent in ranking high typical than low typical stimuli, $t$ (18) = -6.21, $p < .01$ for the superordinate level category and $t$ (18) = -2.03, $p < .05$ for the basic level category. Age did not appear to play a significant role in the consistency in ordering of the high typical stimuli at either of the category levels, except that adults were more consistent than the second graders in ordering these stimuli, $t$ (15) = 3.47, $p < .01$. With respect to the consistency in ordering low typical stimuli, adults were more consistent than the preschool children at the superordinate level, $t$ (19) = 2.43, $p < .05$, and were also more consistent than either the second grade children or the preschool children at the basic level of categories, $t$ (15) = 2.15, $p < .05$ for second graders and $t$ (19) = 3.35, $p < .01$ for preschoolers.

While the results of these analyses may be considered as only suggestive in nature (given the difficulties with the ranking measures in the two tasks), they do tend to follow the pattern that a prototype theory of categorization would predict. That is, one would expect to find greater consistency in the ranking of items rated high in typicality than if the items being ranked are of low category item typicality, since the low typical items have fewer attributes in common than do the high typical items and that judgments with respect to the low typical items may be more idiosyncratic in nature. The consistency found in the ranking of the high typical category items should also be stable across age, since there should be a greater consensus among people with respect to the "good" examples of a category and that differences in category structure that do occur should be the result of differences in the ranking of low typical items (Mervis & Rosch, 1981).

Judgments of Category Item Typicality and Physical Similarity

A second group of college students provided additional ratings of category typicality for all of the experimental stimuli (i.e., targets and distractors) and also provided judgments of the physical similarity of pairs of same category stimulus items. Half of the subjects rated items from the superordinate level category and half of the subjects rated items from the basic level category. Both of the scales used were five-point Likert-type scales, with a rating of one designating high typicality or physical similarity and a rating of five designating low typicality or minimal physical similarity.

Tables 9a and 9b show the results of the category typicality ratings task for each of the different category levels for both target and distractor items. The ratings produced category clusters (high typical and low typical category clusters) at both of the category levels which replicated the experimenter-assigned stimulus clusters (based on the ratings obtained from pilot subjects). The subjects were able to make different ratings for each stimulus at each category level, $\chi^2_{.99}(19) = 190.89$ at the basic level of categorization and $\chi^2_{.99}(19) = 195.94$ at the superordinate level of categorization. The subjects were also in agreement with respect to their preferred stimuli, $\underline{W} = .59$ for basic level stimuli and $\underline{W} = .61$ for superordinate level stimuli.

The mean physical similarity ratings based on the high typical-high typical stimulus pairs and the low typical-low typical

Table 9a

Mean and Modal Typicality Ratings Based on Adult Typicality Judgments

| Animals | Mean | Mode | SD |
|---|---|---|---|
| Targets | | | |
| Horse | 1.24 | 1.00 | .44 |
| Rabbit | 1.17 | 2.00 | .69 |
| Fox | 2.35 | 2.00 | 1.00 |
| Tiger | 2.35 | 1.00 | 1.37 |
| Bear | 1.94 | 1.00 | .83 |
| Cardinal | 3.47 | 3.00 | 1.01 |
| Anteater | 3.94 | 5.00 | 1.20 |
| Llama | 3.88 | 5.00 | 1.36 |
| Alligator | 3.94 | 4.00 | .90 |
| Frog | 3.76 | 5.00 | 1.20 |
| Distractors | | | |
| Lion | 1.94 | 1.00 | .90 |
| Raccoon | 2.24 | 3.00 | 1.10 |
| Squirrel | 2.18 | 2.00 | .88 |
| Deer | 1.82 | 2.00 | .81 |
| Wolf | 2.88 | 1.00 | 1.54 |
| Owl | 3.76 | 3.00 | .97 |
| Walrus | 3.94 | 4.00 | 1.03 |
| Lizard | 4.41 | 5.00 | .87 |
| Tortoise | 3.76 | 4.00 | .83 |
| Boa Constrictor | 3.59 | 3.00 | 1.28 |

Table 9b

Mean and Modal Typicality Ratings Based on Adult Typicality Judgments

| Dogs | Mean | Mode | SD |
|---|---|---|---|
| Targets | | | |
| German Shepherd | 2.29 | 3.00 | .85 |
| Great Dane | 2.53 | 3.00 | 1.13 |
| Irish Setter | 2.06 | 1.00 | .97 |
| Malamute | 1.88 | 1.00 | .93 |
| Collie | 2.30 | 1.00 | 1.11 |
| Yorkshire Terrier | 3.94 | 4.00 | 1.09 |
| Shi Tzu | 4.29 | 5.00 | 1.16 |
| Greyhound | 4.24 | 5.00 | .90 |
| Afghan Hound | 3.71 | 4.00 | 1.05 |
| Pekingese | 3.82 | 5.00 | 1.07 |
| Distractors | | | |
| St. Bernard | 2.41 | 3.00 | 1.12 |
| Cocker Spaniel | 2.24 | 2.00 | .90 |
| Bloodhound | 2.24 | 2.00 | .75 |
| Brittany Spaniel | 2.18 | 3.00 | .95 |
| Harrier | 2.06 | 1.00 | 1.03 |
| Komondor | 4.71 | 5.00 | .59 |
| Maltese | 4.00 | 4.00 | .94 |
| Little Lion Dog | 4.35 | 5.00 | 1.00 |
| Scottish Terrier | 3.52 | 4.00 | .87 |
| Belgian Sheepdog | 2.65 | 3.00 | 1.00 |

stimulus pairs for both target and distractor items are presented in Table 10. Both high typical target and distractor stimuli at the basic level of categorization were rated as being more physically similar to each other than were the high typical targets and distractors at the superordinate level of categorization, $\underline{t}$ (18) = 2.17, $\underline{p} < .05$ for target items, and $\underline{t}$ (18) = 2.45, $\underline{p} < .05$ for distractor items. There were no significant differences in the mean physical similarity ratings obtained for target and distractor items on the basis of level of categorization. Thus, as would be predicted by a prototype theory of categorization, items from a basic level of category are judged to be more similar to each other than category instances which represent a superordinate level category.

Table 10

Mean Physical Similarity Ratings Based on High/High and Low/Low Stimulus Typicality Pairings for the Physical Similarity Judgment Task

| | | Category Level | | | |
|---|---|---|---|---|---|
| | | Superordinate | | Basic | |
| Stimulus Type | Pair Type | Mean | SD | Mean | SD |
| Targets | High/High | 3.22 | .65 | 2.57 | .71 |
| | Low/Low | 4.26 | .56 | 3.55 | 1.44 |
| Distractors | High/High | 3.12 | .52 | 2.36 | .83 |
| | Low/Low | 4.19 | .70 | 3.59 | .79 |

# CHAPTER V

# CONCLUSIONS

## Structure and Consistency of Use of Categorical Knowledge

What conclusions can be drawn from the pattern of results that was obtained in the present study? First, with respect to the main focus of the study concerning the ability to provide and consistently use an internal structure for a natural language category at different age and category levels, it was not until adulthood that subjects were able to provide an internal structure for the categories that were examined. In general, the children tended to treat the category exemplars as being equally good instances of the category under study, with the one exception being that the majority of children were able to select one category exemplar as representing the "best example" of a set of stimuli.

One possible explanation for this pattern of results is that the general form which the young child's categorical structure takes is one of "best example plus other category examples," rather than a more differentiated structural form. A similar type of categorical structure has been obtained for the basic level categories of young children by Mervis and Pani (1980). The one difference between the categorical structures described by Mervis and Pani and the structures obtained in the present study is that the Mervis and Pani structures had more differentiation at the low typical end of the structure than was evident for the children observed here. The general reason why young children display such an undifferentiated categorical structure may be that there is an inability on the part of the young child to make judgments based on fine distinctions among category instances on the basis of category item representativeness because of their lack of experience with large numbers of category items. A larger number of subjects at the second grade level may have provided more information as to the point in cognitive development at which the structural form of a child's natural language categories assume the form that most resembles the structural form shown by adults (i.e., differentiation of category instances according to different degrees of category item representativeness.

A second possible reason for the failure to obtain the type of internal category structure that would be predicted by a prototype theory of categorization, even for young children, is that the paired comparisons procedure may not have been a direct enough means for tapping the child's categorical structure. That this may be possibly the case is pointed to by previous research that has shown that when children are asked directly to rate category instances on the basis of category representativeness, children (even as young as four and five years of age) are able to provide a category structure that is similar to, although not exactly the same as, the structure shown by adults (Mervis & Pani, 1980; Thompson & Bjorkland, 1981). It may be the case that, at least with respect to children, a paired comparisons task may be used to assess the bases on which children make their category decisions (e.g., size, shape of category items;

Howard & Howard, 1977), but that the bases may vary with different cognitive tasks and that there may not be one overall category structure that is automatically activated by all types of cognitive tasks (Mandler, 1983), thus accounting for the lack of consistency of the use of categorical knowledge in the remaining experimental tasks.

The adult subjects were also the only group to show between-subjects consistency in the way in which the examples of a category were rated (i.e., the adults ranked the category examples in a similar way). The adults were also the only group studied that demonstrated within-subjects consistency across tasks in the structuring of their categories (i.e., providing similar orderings of the stimulus items in both the paired comparisons task and in the match-to-sample task). These findings are not surprising if the response patterns shown by the children are taken as reflecting a general lack of categorical structure. If, on the other hand, the response patterns of the children are taken as reflecting a "best example plus other examples" category structure, the children may also be considered as having demonstrated a type of structural consistency, since the majority of the children did maintain the same "prototype" or "best example" across tasks. Again, deciding this matter must await a more detailed examination of the apparently transitional category structure of the second grade children. One possible way of assessing such a transitional stage would be to conduct a study in which much more detailed mappings of a child's knowledge are obtained than was possible in the present study and then to compare the mappings across different age levels and perhaps, longitudinally with the same child. The procedure used by Chi and Koeske (1983) to study a young child's knowledge of dinosaurs would be quite appropriate for such a study.

## Effects of Category Structure on Cognitive Performance

A second major focus of the present study was to assess the possible effects of category item typicality and a category's level of inclusion in a categorical hierarchy on cognitive task performance. The major conclusion which may be drawn from the information provided from the different tasks examined is that recognition memory performance is unaffected by the category item typicality of the to-be-remembered items, and that, except for the preschool children, is also unaffected by the level of inclusion in a categorical hierarchy represented by the to-be-remembered items. The difference in preschool recognition ability (i.e., females having better recognition for superordinate level category items and males having better recognition for basic level category items) may be the result of their performance being more influenced by differences in verbal and spatial-perceptual skills than is the performance of older children and adults.

It can also be concluded from the data that the amount of cognitive processing required to make a decision (whether the decision is one of acceptability of an item as one of a to-be-remembered list or is one of matching a stimulus item to a standard on the basis of category similarity) is not affected by the category item typicality of the category's level of inclusion of the item for which the decision is being made. The one area for which the level of inclusion of a category does play a role in performance is that of accurately making a match to a standard. That is, subjects matching stimuli representing a basic level of categorization produced fewer errors in responding than did subjects responding to stimuli representing a superordinate level category. The finding of better matching performance for basic level over superordinate level stimuli may be a result of the stimuli at the basic level of categorization being more physically similar to each other than was true for the stimuli at the superordinate level of categorization. This possibility, which is suggested by the redundancy gains due to physical identity that are widely observed in name and category level matching task (Posner & Rogers, 1978), is consistent with the phsyical similarity ratings obtained from the college students.

Thus, in general, the results of the present study appear to point out that category item typicality and level of inclusion of a category in a categorical hierarchy is not as important for efficient cognitive performance when the cognitive processing required by a task is of a relatively passive nature than when a task requires cognitive processing that is more active or demanding of the person (e.g., recall versus recognition tasks). Level of inclusion in a categorical hierarchy does become important when the decisions being made involve accurately determining if an item is a member of a particular category under study, with accuracy being greatest for decisions being made for a basic level category.

Developmentally, using the types of procedures described in the present study, one sees the development of a more differentiated categorical internal structure with age. The category structures of the young child are relatively undifferentiated, only distinguishing between the "best example" of a category and other examples of a category. Children of school age seem to have categorical structures that are transitional in nature, approaching the differentiation among category items based on item representativeness that is seen in the category structures of adults, but still retaining some of the characteristics of the structures of the preschool child. The transitional category structures of the school-age child may represent the child's growing experience with category items and their learning about category items (as imposed by formal educational procedures). More research needs to be conducted before the actual form of the school-age child's categories can be expanded upon and the possible reasons for the developmental differentiation seen in category structure are brought to light.

Finally, with respect to consistency of use of categorical knowledge, only the adults were able to show any across-task consistency in the way they used their knowledge of categories and category structure. Although the second grade children were beginning to have category structures that resembled the structures of adults, they were not able to use that structure in a consistent fashion except for the ability to establish and maintain a prototypical example. Preschool children also did not show consistency of the use of categorical knowledge except for that of the use of the prototypical example. What may be necessary for consistency in the use of categorical knowledge to occur is both a well differentiated category structure and a good deal of experience in using such a structure (Chi & Koeske, 1983; Mandler, 1983). The data here seem to be pointing toward a type of production deficiency in the use of categorical knowledge (Flavell, 1977). That is, as appears to be the situation with mnemonic strategy development, having a category structure which resembles to some extent the adult model does not guarantee that the child will be able to utilize the information contained in that structure in either a spontaneous or consistent fashion. Again, further research on the nature of the school-age child's category structures (such as the work done by Chi & Koeske (1983) and Mandler (1983) that have shown similarity in category structure across age not evidenced in the present study) needs to be conducted before the validity of a production deficiency explanation for the lack of consistency in the use of categorical knowledge by children can be assessed.

Further Directions

Given that the results of the present study were, in general, contradictory to the findings of previous research conducted in the area of human categorization, what direction can these findings give to future research and theory building in the area? First, caution should be taken in making broad sweeping statements about the universality of positive cognitive processing gains when the information being processed is representative of a basic level category. As demonstrated in the present study, basic level category information does not automatically result in superior cognitive performance. The type of cognitive processing required by a task (active versus passive processing) may play a more important role in category knowledge/performance effects than the specific hierarchical level of the category information per se. This caution also applies to statements regarding the effects of prototypicality on cognitive processing and performance.

A second recommendation for future investigation in the area is for further exploration of the modification of prototype theory. The theory that has been put forth by Armstrong, Gleitman, and Gleitman (1982) may provide a starting point for such exploration. Also, a theory of categorization which places a greater emphasis on the role of individual expertise or knowledge of category information may be helpful in explaining the developmental findings. The child studied by Chi and Koeske (1983) clearly was an "expert" on dinosaurs and showed excellent cognitive category performance and would have most probably have outperformed an adult who was not an "expert" in the area of dinosaurs if the comparison would have been made. Since, in general, adults have more expertise with all types of categories than do children, category performance effects may include a large expertise component. Thus, future investigations might compare not only different types of category knowledge on cognitive performance across age, but also might compare how different degrees of knowledge of a specific type of category (regardless of age) affects cognitive performance.

In summary, the area of human categorization is not as clear cut as either the prototype theorists or the more traditional categorization theorists would have us believe. The future of the area lies not only in the recognition that a single theory of categorization may not be sufficient to understand the processes that allow us to simplify the information in our environment, but also that categorization is not an isolated process, but is intimately involved with other cognitive processes, such as memory and perception.

# APPENDIX A

## Task Instructions

Item Generation Task:

Today, we are going to look at some pictures and also do some other things. You are going to see a lot of different pictures and it should be a lot of fun. Before we look at the pictures, I want to see how good a rememberer you are. I bet you are a good rememberer aren't you? I am going to say the names of some things. After I say a name I want you to tell me all of the things you can think of that belong to that name. For instance, I might say FLOWERS and you might think of rose, daisy, tulip, and lots of others. Let's try it. I think I'll say the word TOYS. Can you tell me the names of all the TOYS you can think of? (Child responds for one minute.) That was good. The next time I say a name all you have to remember is to tell me the names of all the things that you can think of that go with the name I say. Do you know what you are to do? Good. Let's begin. Tell me all the ANIMALS you can think of (allow up to two minutes; after 15 seconds of silence, prompt--Can you think of any more ANIMALS? After another 15 seconds of silence go to the next category). That was very good. You know a lot about ANIMALS. How about this one? Tell me all the DOGS you can think of (use same procedure as for ANIMALS). Very good. You are a good rememberer. Let's look at some pictures now.

Paired Comparisons Task:

Let's play a pretending game. Do you like to pretend? I want you to pretend that a man from outer space has just landed on Earth. The spaceman's name is Zoltar and he is from the planet Zebnon, which is far, far away. Zoltar is a friendly spaceman, but he doesn't know anything about Earth. Zoltar wants to learn all he can about earth things and he needs some help. Your job is to help Zoltar to learn about earth things. In order to help Zoltar learn more about Earth, I am going to say a name and then show you some pictures that go with that name. You will see two pictures at a time in these windows (demonstrate apparatus). When I show you the pictures, I want you to pick the picture that you think will help Zoltar to learn the most about the name. Push the button under the picture you pick (demonstrate procedures). When you pick a picture, pick the picture that will help Zoltar to learn the most about the name I say, not the picture that you like the most. OK, what do you do when I show you the pictures (wait for child's response and give appropriate feedback). Let's try it.

Let's pretend that Zoltar wants to learn about FOOD (show first picture pair). Pick the picture that you think will help Zoltar to learn the most about FOOD. (Repeat with remaining practice pairs and assist where necessary.) What picture would help Zoltar the most? (Show pair composed of "ice cream cone" and "bread".) Remember, pick the picture that will help Zoltar the most to learn about FOOD, not the picture that you might like the best (observe child's choice and give appropriate feedback).

Zoltar is really interested in learning about (category name--ANIMALS or DOGS). There are a lot of different kinds of (category name) on Earth, so Zoltar is going to need a lot of help this time. I'm going to show you a lot of pictures of (category name), so you can help Zoltar. Each time I show you some pictures, push the button under the picture you think will help Zoltar to learn the most about (category name), just like you did for FOOD. Remember pick the picture that will help Zoltar to learn the most about (category name), not the picture that you like the most. Ready? (Present stimulus pairs; prompt where necessary; record choice and time, give appropriate feedback to keep child motivated.) That was very good, I think you helped Zoltar to learn a lot about (category name).

Recognition Memory Task:

I want to see how well you can remember pictures of things that you see. There are going to be a lot of pictures, so I want you to listen very carefully to what I tell you about the pictures. Your job is to try to remember some of the pictures you see. Let's try it once.

Practice Task

I am going to show you three pictures. Watch them carefully because I'm going to ask you later if you remember any of them (show picture stimuli). Now I'm going to show you some more pictures. I want you to tell me if any of the pictures are exactly the same as the pictures you just saw. If a picture is exactly the same as one you've seen, press the button by the happy face (demonstrate). If the picture is new, press the button by the sad face (demonstrate). Let's try it. Is this picture exactly the same as any of the pictures you just saw? (Show picture and help where necessary.) That was good. Now we're ready for the real task.

Actual Task

Now I'm going to show you some more pictures, one picture at a time. Watch each picture very carefully so that you will be able to remember them. Ready? (Present stimuli--five seconds per picture.) Now I want you to name all of the colors that you know (approximately 30 seconds).

I'm going to show you a lot more pictures, some of the pictures will be exactly the same as the ones you just saw, but some of the pictures will be new. If you see a picture that is exactly the same as one of the pictures you just saw, press the happy face button. If the picture is new, press the sad face button (demonstrate). Remember, only press the happy face button if a picture is exactly the same as one of the pictures you just saw. What do you do if the picture is exactly the same as one you saw before? (Wait for the child's response and give appropriate feedback.) What do you do if the picture is new? (Wait for the child's response and give appropriate feedback.) Ready? (Present test stimuli; prompt if necessary; record choice time and prompt.) That was very good. I new you were a good rememberer.

Match-to-Sample Task:

Remember the last time you were here with me and you looked at all of the different pictures? Remember how you helped Zoltar, the spaceman, to learn all about (category name)? Well, Zoltar needs your help again. You see, Zoltar isn't sure he knows enough about (category name). He's a little confused about things, since he has so much to learn about Earth before he returns to his home on Zebnon. Let's see if you can help Zoltar some more.

In order to help Zoltar this time, I am going to show you some more pictures, three pictures at a time. The picture that will be in this window will always be the same (point to top opening). The pictures you will see in these windows (point to bottom openings) will change a lot. Let's see how it works. Zoltar knows that it is a picture of FOOD (reveal practice standard). Zoltar isn't quite sure about these pictures (reveal practice exemplars). You can help Zoltar to learn more about FOOD by pressing the button under the picture (point to exemplars) that you think is the most like the picture in the top window (point to standard). Let's look at two more pictures. Pick the picture that is most like the top picture (point to standard and provide necessary feedback). How about these pictures? (Continue with practice making sure child understands task and that at least one stimulus pair contains a non-exemplar.)

Let's see if you can help Zoltar to learn more about (category name). Zoltar knows that this is a picture of a (category name) (reveal child's standard). I'm going to show you a lot of pictures now. Remember, each time I show you the pictures, press the button under the picture that is <u>most like</u> this picture (point to standard). What do you do when you see the pictures? (Wait for child's response and provide appropriate feedback.) Good. Let's help Zoltar some more. Ready? (Present test pairs; prompt and provide feedback when necessary; record choice and time.)

Physical Similarity Ratings Task:

This part of the study has to do with how physically similar we perceive instances of a given category to be. I am going to show you pairs of items which represent the same category and I want you to judge how physically similar the two items are. That is, I want you to judge how similar the two items are with respect to their overall shape, not necessarily that they are the same color, etc. For example, (show SWEATER and SHIRT), these two items of clothing would probably be considered to be very similar while these two items (show RAINCOAT and SOCKS) would not be considered to be very physically similar, yet they are both instances of the category, CLOTHING. If you saw these two items (show RAINCOAT and BOOTS) you might be tempted to say they were quite similar since they are from the same category and are often associated with each other, yet physically they are quite different.

On the following forms, I want you to rate the physical similarity of a large number of pairs of items. All of the items you see will be from the same category. You will judge the pairs on a five-point scale. A one means that you feel the items are very similar to each other. A five means that the items are very dissimilar from each other physically. A three means moderate similarity. For example, if I presented the following pair (show APPLE and PEAR for the category, FOOD), you would probably rate them a one or two. If I presented this pair (show GRAPES and BREAD) you would probably rate them a five. A three would probably be assigned to this pair (show APPLE and GRAPES).

Typicality Ratings Task:

This part of the study has to do with what we have in mind when we use words which refer to categories. I am going to show you three examples of a category and I want you to select the example which you feel best represents the category label I give you. (Present color cards--true red, orangish red, and purple red.) Select the card which you think best represents the color RED. All of the cards represent the color category, RED, yet the orange red and the purple red may not seem to be as good examples of RED (as clear cases of what RED refers to) as the clear "true" red. In short, some reds are redder than others. The same is true for other kinds of categories. Think of VEGETABLES. We all have some notion of what a "true" or "real" vegetable is. To me, corn or peas are very good examples of vegetables, while rutabaga is not a good vegetable example for me.

Notice that this kind of judgment has nothing to do with how well you like the thing; you can like a purple red better than a true red but still recognize that the color you like is not a true red. You may love rutabagas, but hate peas and carrots, without thinking about what people usually mean by the word VEGETABLES.

On this form you are asked to judge how good an example of a category various instances of the category are (not how much you like a particular instance). At the top of the form is the name of a category. I will then show you pictures of some members of the category. For each picture, you are to rate it on a five-point scale. A one means you feel that the pictured member is a very good example of your idea of what the category is. A five means you feel the picture fits very poorly with your idea or image of the category (or is not a category member at all). A three means you feel the member fits moderately well. For example, one of the members of the category FRUIT is APPLE. If APPLE fits well your idea or image of FRUIT, you would mark a one on your form; if APPLE fits your idea of FRUIT very poorly, you would mark a five; a three would indicate moderate fit. Use the other numbers on the five-point scale to indicate intermediate judgments.

# REFERENCES

## REFERENCES


Anglin, J.M. (1977). Word, object and conceptual development. New York: Norton.

Anglin, J.M. (1978). From reference to meaning. Child Development, 49, 969-976.

Armstrong, S., Gleitman, H., & Gleitman, L.R. (1982). On what some concepts might not be. Manuscript submitted for publication.

Berlin, B. (1978). Ethnobiological classification. In E. Rosch & B.B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berlin, B., & Kay, P. (1969). Basic color terms: Their universality and evolution. Berkeley, CA: University of California Press.

Blewitt, P. (1981). "Dogs" vs. "Collie": Vocabulary in speech to young children. Paper presented at the meeting of the Society for Research in Child Development, Boston.

Boswell, D.A., & Green, H.F. (1982). The abstraction and recognition of prototypes by children and adults. Child Development, 53, 1028-1037.

Bourne, L.E., Jr., Dominowski, R.L., & Loftus, E.F. (1979). Cognitive processes. Englewood Cliffs, NJ: Prentice-Hall.

Bowerman, M. (1978). The acquisition of word meaning: An investigation into some current conflicts. In N. Waterson & C. Snow (Eds.), Development of communication: Social and pragmatic factors is language acquisition. New York: Wiley.

Bransford, J.D. (1979). Human cognition: Learning, understanding and remembering. Belmont, CA: Wadsworth.

Bruner, J.S., Goodnow, J.J., & Austin, J.G. (1956). A study of thinking. New York: Wiley.

Carey, S. (1978). The child's concept of animal. Paper presented at the meetings of the Psychonomic Society, San Antonio, TX.

Carson, M.T., & Abrahamson, A. (1976). Some members are more equal than others: The effect of semantic typicality on class-inclusion performance. Child Development, 47, 1186-1190.

Chi, M.T.H., & Koeske, R.D. (1983). Network representation of a child's dinosaur knowledge. Developmental Psychology, 19, 29-39.

Cohen, L.B., & Strauss, M.S. (1979). Concept acquisition in the human infant. Child Development, 50, 419-424.

Consumer guide to 1982 cars. (1982). Publication International, Ltd.

Farah, M.J., & Kosslyn, S.M. (1982). Concept development. In H.W. Reese & L.P. Lipsitt (Eds.), Advances in child development and behavior (Vol. 16). New York: Academic Press.

500 words to grow on (1973). (Illustrated by H. McNaught). New York: Random House.

Flavell, J.H. (1977). Cognitive development. Englewood Cliffs, NJ: Prentice-Hall.

Garner, W.R. (1978). Aspects of a stimulus: Features, dimensions, and configurations. In H. Rosch & B.B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldberg, S., Perlmutter, M., & Meyers, N.A. (1974). Recall of related and unrelated lists by two-year-olds. Journal of Experimental Child Psychology, 18, 1-8.

Golden book on animal stamps. (1966). Racine, WI: Golden Press.

Golden book on bird stamps. (1966). Racine, WI: Golden Press.

Golden book on horse stamps. (1966). Racine, WI: Golden Press.

Golden book of north american mannals stamps. (1966). Racine, WI: Golden Press.

Golden book of snakes, turtles and lizard stamps. (1966). Racine, WI: Golden Press.

Horton, M.S., & Markman, E.M. (1981). Developmental differences in the acquisition of basic and superordinate categories. Child Development, 51, 708-719.

Howard, D.V., & Howard, J.H., Jr. (1977). A multidimensional scaling analysis of the development of animal names. Developmental Psychology, 13, 108-113.

Keil, F.C., & Batterman, N. (In press). A characteristic-to-defining shift in the development of word meaning. Journal of Verbal Learning and Verbal Behavior.

Lange, G., Kienapple, K., Sullivan, R.L., & Allen, J. (1981). Relationships between study behavior, object knowledge and recall memory proficiency in preschool children. Journal of Psychology, 109, 11-26.

Maccoby, E.E., & Jacklin, C.N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.

McCauley, C., Weil, C.M., & Sperber, R.D. (1976). The development of memory structure as reflected by semantic priming effects. Journal of Experimental Child Psychology, 22, 511-518.

McCloskey, M.E., & Glucksberg, D. (1979). Natural categories: Well defined or fuzzy sets? Memory and Cognition, 11, 1-37.

Mandler, J.M. (1983). Representation. In P.H. Mussen (Ed.), Handbook of child psychology (4th ed.), Vol. III: Cognitive development (J.H. Flavell & E.M. Markman, Eds). New York: John Wiley & Sons.

Melkman, R., Tversky, B., & Baratz, D. (1981). Developmental trends in the use of perceptual and conceptual attributes in grouping, clustering, and retrieval. Journal of Experimental Child Psychology, 31, 470-486.

Mervis, C.B. (1980). Category structure and the development of categorization. In R. Shapiro, B.C. Bruce, & W.F. Brewer (Eds), Theoretical issues in reading comprehension. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mervis, C.B., & Canada, K. (1981). Child-basic categories and early lexical development. Paper presented at the meetings of the Society for Research in Child Development, Boston.

Mervis, C.B., & Crisafi, M.A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. Child Development, 53, 258-266.

Mervis, C.B., & Pani, J.R. (1980). Acquisitions of basic object categories. Cognitive Psychology, 12, 496-522.

Mervis, C.B., & Rosch, E. (1981). Categorization of natural objects. Annual Review of Psychology, 32, 89-115.

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

Meyers, N.A., & Perlmutter, M. (1978). Memory in the years two to five. In P.A. Ornstein (Ed.), Memory development in children. Hillsdale, NJ: Lawrence Erlbaum Associates.

Miller, G.A. (1978). Practical and lexical knowledge. In E. Rosch & B.B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Lawrence Erlbaum Associates.

Murphy, G.L., & Smith, E.E. (1982). Basic-level superiority in picture categorization. Journal of Verbal Learning and Verbal Behavior, 21, 1-20.

Neimark, E. (1974). Natural language concepts: Additional evidence. Child Development, 45, 508-511.

Nelson, K. (1974). Variations in children's concepts by age and category. Child Development, 45, 577-584.

Nelson, K. (1977a). Cognitive development and the acquisition of concepts. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nelson, K. (1977b). The conceptual basis of naming. In J. MacNamara (Ed.), Language learning and thought. New York: Academic Press.

Nelson, K., & Brown, A.L. (1978). The semantic-episodic distinction in memory development. In P.A. Ornstein (Ed.), Memory development in children. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Rescorla, L.A. (1980). Category development in early language. Journal of Child Language, 8, 225-238.

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

Rosch, E.H. (1973a). Natural categories. Cognitive Psychology, 4, 328-350.

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

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

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), Studies in cross-cultural psychology. London: Academic Press.

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

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

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

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

Rosner, S.R., & Hayes, D.S. (1977). A development study of category item production. Child Development, 48, 1062-1065.

Rosner, S.R., & Poole, D.A. (1978). A longitudinal study of category item judgments. Paper presented at the meetings of the Midwestern Psychological Association, Chicago.

Rosner, S.R., & Poole, D.A. (1981). Category attributes and item inclusion. Paper presented at the meetings of the Society for Research in Child Development, Boston.

Ross, G. (1980). Categorization in one and two year olds. Developmental Psychology, 16, 391-396.

Saltz, E., Soller, E., & Sigel, I.E. (1972). The development of natural language concepts. Child Development, 43, 1191-1202.

Schuler, E.M. (Ed.). (1980). Simon and Schuster's guide to dogs. New York: Simon Schuster.

Sherman, T. (1981). Categorization skills in infants. Paper presented at the meeting of the Society for Research in Child Development, Boston.

Smith, E.E., & Medin, D.L. (1981). Concepts and categories. Cambridge, MA: Harvard University Press.

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

Solso, R.L., & McCarthy, J.E. (1981). Prototype formation: Central tendency vs. attribute frequency model. Bulletin of the Psychonomic Society, 17, 10-11.

Sperber, R.D., Davies, D., Merrill, E.C., & McCauley, C. (1982). Cross-category differences in the processing of subordinate-superordinate relationships. Child Development, 53, 1249-1253.

Sullivan, R.L. (1979). The relationship between episodic and semantic recall - memory proficiency in three- and four-year old children. Unpublished masters thesis, Purdue University.

Thompson, B.E., & Bjorkland, D.F. (1981). Children's judgments and processing of typical and atypical category exemplars. Paper presented at the meetings of the Society for Research in Child Development, Boston.

Vygotsky, L.S. (1962). Thought and language. Cambridge, MA: MIT Press.

White, T.G. (1982). Naming practices, typicality and underextension in child language. Journal of Experimental Child Psychology, 33, 324-346.

Winer, B.J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Wittgenstein, L. (1953). Philosophical investigations. New York: Macmillan.

Younger, B.A., & Cohen, L.B. (1983a). Development change in infants' perception of correlated attributes. Paper presented at the meetings of the Society for Research in Child Development, Detroit.

Younger, B.A., & Cohen, L.B. (1983b). Infants' acquisition of correlated vs. uncorrelated categories. Paper presented at the meetings of the Society for Research in Child Development, Detroit.