Episodic Specialization of Classification Rules

By

Scott William Allen, B.A.

A thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

1990

Episodic Specialization of Classification Rules

DOCTOR OF PHILOSOPHY (1990) McMASTER UNIVERSITY

(Psychology) Hamilton, Ontario

TITLE: Episodic Specialization of Classification Rules

AUTHOR: Scott William Allen, B. A. (University of Victoria)

SUPERVISOR: Professor Lee R. Brooks

NUMBER OF PAGES: (ix) 96

ABSTRACT

This thesis reports research investigating the effect of practice on the operation of an explicitly stated rule. Subjects were furnished with an easy, sufficient additive rule and given practice classifying pictures of imaginary animals into two categories. Following practice, subjects showed a strong similarity effect: They were faster and more accurate at classifying positive matches (new items in the same category as the most similar old item) than negative matches (new items in the opposite category to the most similar old item), suggesting that memory for specific prior episodes played an important role in applying the rule. Subjects were not always wrong on the negative matches, and were usually immediately aware of any mistakes they did make, suggesting that the rule had not been abandoned but may have served a monitoring function. Mnemonically distinctive stimuli are probably important for episode-based responding since these effects did not occur when the items were presented as written lists of features. When subjects were stressing accuracy, and paying a large premium in response times, the negative match effect remained, while fewer than half the subjects considered similarity to old items when questioned about the cause of their errors, suggesting that the episodic information can have effects without subjects' awareness. However, a similarity strategy can be under strategic control as shown when the rule was manipulated so that information from prior experience was in opposition to the present rule information for nearly all the items. Additionally, negative

matches were more disrupting if seen in the context of predominantly old items, these episodic effects were helpful when the rule was not perfectly predictive, and the classification decisions of third grade children were strikingly similar to those of the adults suggesting that children of this age group were employing a similar balance between the use of the abstract and episodic information. While some previous accounts of categorization suggest that episode-based responding is important only before subjects develop a rule or some other reliable basis of categorization the current results suggest that, for some common types of stimuli, similarity to prior episodes may become more important with practice.

ACKNOWLEDGEMENTS

Graduate training has often been described in terms of an apprenticeship. Lucky is the apprentice who is privileged to learn with a master rather than a journeyman, and doubly fortunate is the apprentice who studies with one who is a master both of his craft and of teaching it. I have been that apprentice and I am grateful to Lee Brooks for it. I am also grateful to the other members of my advisory committee, Larry Jacoby and Lorraine Allan, for their advice and support throughout my time at McMaster. Geoff Norman, although not directly related with the work on this thesis, has taught me a great deal in the time I have known him and I thank him for it. I also thank Jim Wagner for his help with the developmental experiments. This work, and my time at McMaster, have greatly benefitted from interactions I have had with my fellow graduate students and friends, especially Marianna Stark, April Takeda, Jane Collins, Mike Marriott, Jim MacRae, Fish, Glenn Regehr, and most especially, John Baker. None of this would have come about had it not been for Mike Masson at UVic who started me on this road through his infectious interest in cognition and his exemplary supervision. I thank the Natural Sciences and Engineering Research Council for financial support. Finally, I am grateful to Joanne whose love, support, encouragement, patience, and belief in me have kept me going through the rough times and have so much intensified the good times.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Learning to classify items into new categories represents an important part of learning new skills. Medical students must learn to diagnose disease categories in order to apply the proper treatment, mathematics students must learn to recognize categories of word problems in order to apply the appropriate formulae to solve them, and statistics students must learn to recognize different sorts of experimental designs in order to apply appropriate analyses. When a student is first starting out, a resource available to help with these decisions is the similarity between a new problem and one of the exemplar problems used in instruction. But a prevalent assumption in the concept learning field has been that as a person becomes increasingly familiar with a category, classification on the basis of similarity to specific items quickly becomes less important, while concentrating on the key diagnostic characteristics of the category take over the lion's share of the classification procedure. This thesis reports a research program that questions this assumption. The main question being asked is: "What happens when a person is given a rule to use and then given a chance to practice that rule?"

One can imagine two extreme answers to this question. The first is that with practice the idiosyncratic item information quickly drops out and the person simply becomes faster and more accurate at applying the rule. If this were the case, then apart from an initial training period, there should be no effect of specific items on classification performance. The opposite extreme answer is that after practice with a sufficient number of examples, a library of instances is built up in memory which is used in classifying novel examples on the basis of analogy to these known items. If this were the case, then once a sufficient storehouse of instances is built up, the specific prior episodes will be doing all the work and the original rule would have no part in classification performance. The use of specific episodes would be mimicking the effects of rule use.

The most interesting question, of course, is about the conditions under which one approaches either extreme. This thesis is an attempt to address some of the questions regarding what controls learning and processing in category learning situations in which one starts with a serviceable verbal statement of the classification criteria. Answers to these questions would be of interest in situations, such as medicine and other areas of formal expertise, in which learners receive explicit instruction in how to categorize stimuli.

The Background Literature

The task domain defined by this problem falls between two well researched areas. It is related to concept learning in that in both the type of rule learning explored here and in concept learning studies, people are learning to classify items into categories. However, in concept learning research, subjects are not given an adequate classification rule before they begin. The same is true in the problem solving literature. Seen from the perspective of the traditional concept learning and problem solving literatures, all the action has already taken place once the subjects have been told how to calculate the answer. Although in formal instruction in a problem solving domain subjects may be told how to calculate an answer, it is usually a difficult process which requires much practice to apply correctly to generate the proper solution. That is not true of the present studies.

However, many of the issues in concept research are still closely related to the current approach and should be addressed. Smith and Medin (1981) introduced a classification of theoretical approaches to concept learning that has become standard.

The classical view. Early research in concept formation took the form of requiring subjects to induce rules from presentations of items with feedback (e.g. Bruner, Goodnow, & Austin, 1956). The subject would be presented with a series of examples, usually consisting of a set of four discrete dimensions (e.g. shape of items, number of items, colour of items, number of borders) each with two or three discrete values (e.g. circle, triangle, or square; 1, 2, or 3; green, red, or black). The task facing the subject was to try to classify each of the items as it was presented. With practice, and feedback on every trial, the subject was meant to induce the underlying classification rule. This was seen as a reasonable model of how people learned to classify objects. The specific point of argument in their work was that people were accomplishing this task by a process of hypothesis testing rather than the process of accumulating differential habit strength to stimulus elements, as advocated by the behaviourists. But the underlying assumption common to both views was that learning a concept was mainly a process of learning what elements of the stimuli were essential for correct categorization. Bruner, Goodnow, and

Austin's 1956 book "A Study of Thinking" contended that both adults and children learned to apply categorization rules to specific examples and that through the presentation of many examples with feedback one could induce the underlying rules. Thus their research dealt with how people could learn different types of rules, what strategies were used, how difficult different logical forms of rules were and so on. The rules were defined by a set of criterial features that were necessary and sufficient for classification. This type of representation would allow for elegant, efficient representation of concepts. This has been called a classical view of concept learning and representation because the emphasis on discovering the necessary and sufficient conditions for membership in a concept has historically been a dominant concern in both philosophy and psychology (Smith & Medin. 1981).

There were a number of problems with the classical view (see Smith & Medin, 1981 for a detailed discussion). One of the most notable was the consistent failure to find necessary and sufficient features for many natural categories. If the process of classification is based on these features, they should be available. however that appears not to be the case. Further problems for the classical view come from the existence of a graded typicality structure (e.g. a robin is considered a more typical bird than an ostrich) and unclear cases (e.g. is a rug furniture?. Medin & Smith, 1984). According to a classical view all members are equally good members providing they possess the singly necessary and jointly sufficient features for membership. This feature of classically defined categories also excludes unclear cases -- if the necessary and sufficient features are present the exemplar is a member of the category, if they are not it is not a member.

The probabilistic view. In response to the problems of the classical view with natural categories, Rosch (1973, 1975a, 1975b, 1978) and her colleagues (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Rosch, Simpson, & Miller, 1976) developed a probabilistic view of natural categories. Instead of requiring adequate knowledge of a concept to be a set of necessary and sufficient defining features, people were held to normally learn a large set of features that were characteristic of, but not necessary for the concept. The basis for categorization was held to be the accumulation of a sufficient number of these characteristic features to distinguish the target concept from other plausible concepts. This form of knowledge would account for the fact that people normally acted as if some concept members were better than others, behaviour that was difficult to account for if membership were determined by singly necessary and jointly sufficient membership conditions. Although Rosch and her colleagues were careful to not specify a theory of representation, a popular form of model based on their work was that, through exposure to many items, some form of average or prototype was abstracted that consisted of all of the characteristic features. New items were assigned to a category on the basis of their relative similarity to the prototype of various categories. As with the classical view, subjects in learning experiments conducted under this view were never given classification rules when studying category learning, in this case because of the assumption that this form of learning was automatic and natural. This framework led to investigation of topics such as difference in classification performance on typical versus atypical category members, what level of category was easiest to learn, and what level was learned first by children.

For the purposes of this thesis the important notion in this view is that of implicit abstraction of the prototype and, as in the classical view, a loss of specific episodic information with continued learning. That is, as people become more familiar with a category, they rely more on the abstracted prototype and less on their memory for specific examples. For example, Homa, Sterling and Treppel (1981) took this position for dot patterns when they suggested that instances are used for generalization only early in a training sequence, before the learner has had an opportunity to abstract a prototype. In this case, the structure being learned was held to be ill-defined, rather than the explicitly defined rules thought to be learned according to the classical view, but again the assumption is that the influence of particular prior cases quickly falls out while the abstract information carries the weight of the decisions.

<u>The exemplar view</u>. More recently, as a response to the abstractive emphasis of the classical and probabilistic views, some investigators have emphasized the specific exemplars of categories (e.g. Brooks, 1978, 1987; Jacoby & Brooks, 1984; Medin & Ross, 1989; Medin & Schaffer, 1978; Smith & Medin, 1981). Although there are, as in the probabilistic view, a number of instantiations, (Estes, 1986a, b; Hintzman, 1984, 1986, 1988; Nosofsky, 1986; Nosofsky, Clark, & Shin, 1989) the common logic is that rather than abstracting the essence of a

category at the time of experience with the category exemplars, either by inducing a rule or by abstracting a prototype, one may simply make use of memory for specific prior examples to make a categorization decision for a new item. A new item is categorized on the basis of its similarity to individual prior cases rather than on its similarity to an abstracted prototype. According to this view there is no need for abstraction, either of a prototype or of defining rules. Most of the data that favored a prototype explanation over a rule explanation could also be explained by a nonanalytic use of prior exemplars for classification. The presumed benefit of having information in this form is that such knowledge would provide a more flexible response to complicated, interacting conditions in the world. Once again, the research in this framework did not tend to give subjects rules for classification when examining their behavior. A development of similar basic ideas has occurred in the artificial intelligence literature under the name of case-based reasoning (Schank, 1982; Kolodner, 1988).

Literature not centered on induction. One of the few researchers investigating categorization (loosely speaking) where subjects were provided with a rule followed by practice with examples has been Brian Ross (Ross, 1984, 1987, 1989a, 1989b; Ross & Kennedy, 1990; Ross, Perkins, & Tenpenny, in press). Teaching his subjects to solve probability word problems, Ross (1984) investigated their use of memory for prior examples when applying the rules they were learning to novel problems. He found that subjects' use of "remindings" of prior examples (where a "reminding" is an explicit retrieval of a prior case) was affected by the similarity of the prior example to the current problem and that remindings had a substantial effect on the successful solution of the problems. Subjects performed better on a novel problem when it had a story line similar to a problem in the training set that required a similar solution than when it had a neutral story line. Furthermore, subjects were worse than in the neutral condition when the test problem had a story line similar to a practice problem that required a different solution. Further studies have demonstrated different effects of superficial as opposed to structural similarity on the access and use of similar problems (Ross, 1989a) and that the exemplar comparison process involved may be a key aspect of a form of abstraction (Ross & Kennedy, 1990; Ross, Perkins, & Tenpenny, in press; see also Medin & Ross. 1989).

The major point of similarity with this thesis is that, as Ross (1989a) emphasizes, he finds the reminding effects even when the relevant abstract information has been presented to the subjects. This situation is an interesting case because it represents a common teaching situation through all levels of formal education right up through professional schools (e.g. Allen, Norman, & Brooks, 1990). At one level, this thesis can be viewed as taking Ross's work with remindings into a situation that is more clearly one of categorization. Ross focuses on people who are in the early stages of learning relatively difficult rules. Their use of remindings can be viewed as a way of coping with the difficulty of the application of the rule to the problem at hand. The rules they are working with are complex and their application to a particular problem is not at all obvious (as anyone teaching a course in probability quickly discovers). In this thesis the focus is on people who have been given a rule and practice with examples as in Ross's work but the difference in the rules and materials make the task quite different. Here the rules are meant to be very easy and, for the most part, perfectly sufficient. Their application to a particular example is meant to be very straightforward. The subjects will have had enough practice that the rules should be applicable to a new case with ease. This is meant to reflect more the end stage of the type of task at which Ross is looking. However, the current work is of independent interest, since everything that Ross asserts about learners struggling with initially difficult material could be true without determining whether prior episodes would be important for the current easier case. The unique focus of the current work is more on the type of knowledge that underlies performance on well practiced skills. As will be discussed more fully below, having expert performance largely based on separate episodes could be the basis of a considerable drift in the exact way in which the original rule is interpreted in cases not easily handled by the rule. The underlying theoretical issue is whether practice generally leads to a greater concentration in the form of underlying knowledge or to a dispersal of it.

The Purpose of this Thesis

The purpose of this thesis is to investigate the interplay between information about specific exemplars and abstract rule information in the learning of a relatively simple classification rule.

Experimental Materials

To illustrate the argument, consider the material presented in Figure 1.

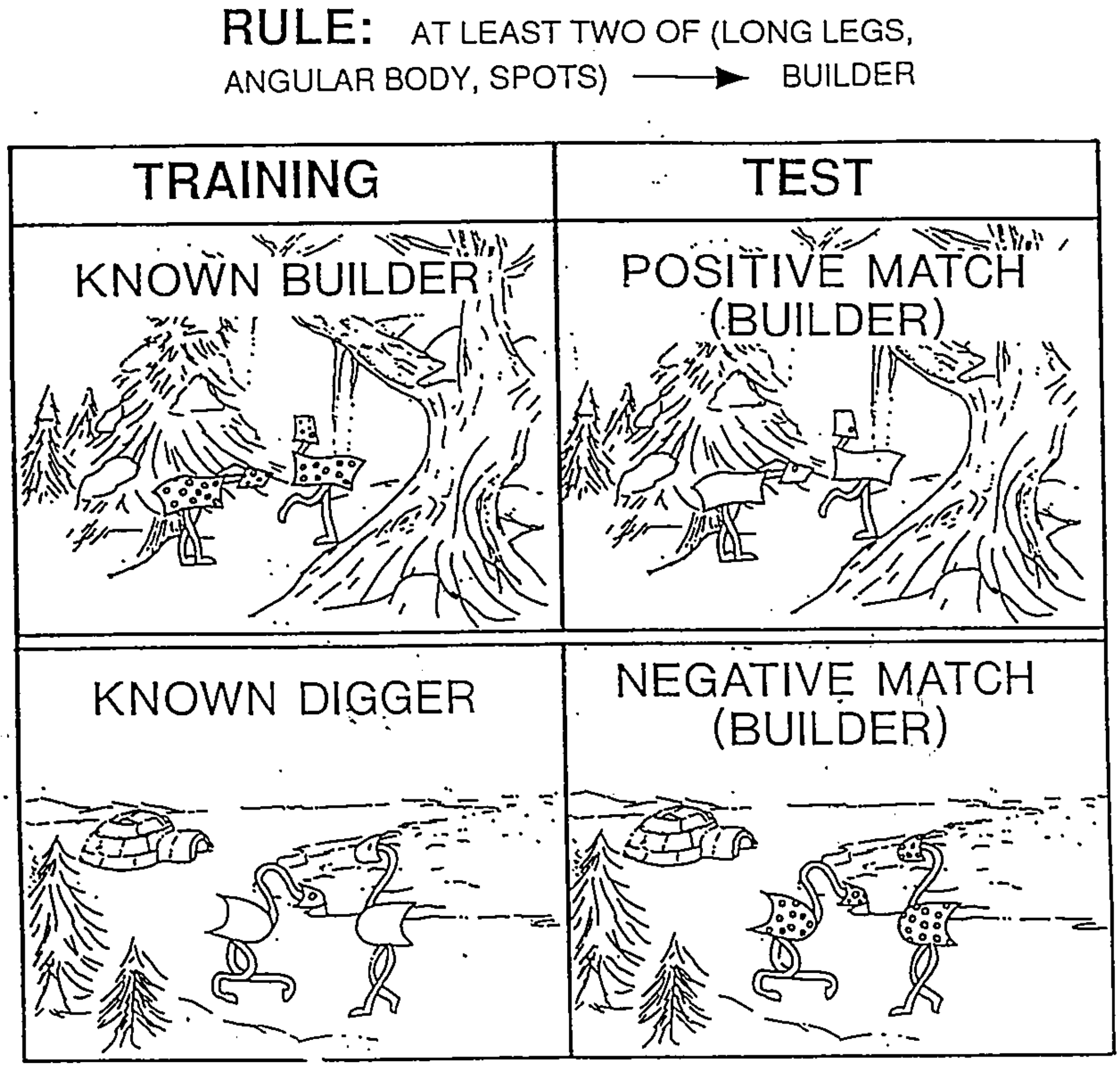


Figure 1: An example of the experimental materials. According to the rule given, the training items shown in the left column are a "builder" on top (long legs, angular body, spots) and a "digger" on the bottom (only long legs). The "positive match" appears only in test and according to the rule is in the same category as the most similar training item. The "negative match" is similar to one of the training items, but according to the rule is in the opposite category to that item (spots in addition to long legs makes it a builder).

The imaginary animals shown are divided into two categories: Those who live in shelters they build from materials available in their environment (Builders) and those who live in holes they dig (Diggers). They can be placed in these categories by a simple (simplest possible) additive rule, involving three of the five binary dimensions used to construct them. In the experiments to be reported, subjects were given a categorization rule that was sufficient to correctly categorize all of the animals in the experiment. They were then given training in categorizing eight of these items (e.g. the "known builders" and "known diggers" in Figure 1). The question is whether this training resulted only in a faster application of the rule or (also) established special cases that would influence the categorization of subsequent new items. To determine this, positively and negatively matching items were introduced into the test sequence. According to the rule, a "positive match" is in the same category as the most similar training item. A "negative match" is also similar to one of the training items, but according to the rule is in the opposite category to that item. For example, spots in addition to long legs makes the negative match of Figure 1 a builder despite its resemblance to the "known digger." If the subjects produce more errors and longer reaction times in the categorization of the negative matches, one might conclude that for these conditions the similar prior processing episodes had an effect on the processing of new similar items. The general procedure of categorization would have partly specialized around the particular cases experienced in training, such that new items similar to known items would not be categorized solely on the basis of the original general rule.

Ecological considerations

However, the major part of the argument is <u>not</u> that such effects are likely to occur in general, but rather only under particular conditions that turn out to be of some ecological interest. An attempt was made to mimic some of these conditions in the experimental conditions that were chosen. (a) The stimuli were chosen to have some of the properties of natural objects. Although one can analyze any of the animals, breaking it into its component parts to determine their values, the parts do form an integrated, recognizable whole. This is in contrast to the stimuli often used in the concept learning literature that were made up of discrete, nonintegrated dimensions. (b) When we experience an exemplar of a category in "real life" we generally encounter an item repeatedly in the same surroundings, interacting with those surroundings. This is modelled in the current experiments by providing different background scenes on which the animals occur. Each background is home to one builder and one digger; although the background scenes are not predictive with respect to the categories, they are predictive with respect to individual items. (c) The form of the classification rule (e.g. see Figure 1) is at least similar to the type of rule often used in verbal communications to beginners regarding a new visual discrimination (e.g. Lepidoptera identification manuals -- Butterflies: thin body, "club" at end of antennae, chrysalis seldom enclosed in a cocoon, eyes of adult generally more prominent, hold wings vertically when at rest, diurnal. Moths: fat body, can have "furry" antennae, cocoon, wings horizontal at rest, nocturnal, phototropic -- but with exceptions on all individual features). Comparable rules are given in areas of medicine that require learning visual categorizations, such as dermatology, histological pathology, radiology, and endoscopy. In addition, this "additive rule of thumb" is a verbal form that would generate the family resemblance structures that have been held to be useful for characterizing ill-defined, natural-kinds categories. (d) Finally, and possibly most important, the test items are presented in generally familiar sets of items. With the exception of excursions to the zoo or the museum, few areas of life have the kind of concentrated, direct confrontation with strange exemplars that occur in many concept learning experiments. One could easily anticipate that previous knowledge resources will be used differently if the subjects' problem solving abilities are being challenged than if things seem pretty much like normal.

Collectively, the conditions just listed define an interesting set of circumstances that are sufficiently common, and sufficiently likely to interact with presumptions about the effects of practice, that they deserve investigation. Several of the subsequent experiments are designed to evaluate the importance of these conditions. Clearly these materials and procedures, are not meant to mimic closely any particular situation in the natural world. Rather, the purpose is to develop an idealized situation in which the stimuli have enough natural characteristics to be interesting while at the same time providing a sufficiently clear classification rule that there is no need to rely on anything else. The ecological considerations just discussed are meant to reflect what may be important about everyday classification of visually rich stimuli and, to the extent that our experimental situation is not "true-to-life", it should favor simple application of the rule in the subject's possession.

Overview of the experiments

The first experiment demonstrates a strong tendency for subjects to use analogy to previously seen items as a classification strategy even when they are in possession of a perfectly predictive and simple classification rule. That is, they use analogy to old items when there is no logical necessity to do so. Experiment 2 suggests that configural, memorable units are important to the use of an analogy strategy. This would be expected because use of an analogy strategy requires first of all that the items be available in memory and secondly that they be amenable to an overall similarity basis of comparison. This was lent support by the fact that analogy effects were seen when the item information was presented to the subjects in the form of drawings of animals, but were not seen when the same information was presented in the form of written lists of features. Apart from the form of stimulus representation, the evidence of nonanalytic responding seems robust across a number of factors. It is not reliant on having additional individuating information other than the pictures of the items themselves: Subjects showed evidence of analogy-based responding with or without the presentation of stories containing idiosyncratic information about each item coincident with the item presentation in the training phase.

In Experiment 3, subjects appear to use analogy to prior episodes not only when instructions stress speed but also when accuracy is stressed and they are paying a large premium in response times to avoid errors. They continue to make a higher proportion of errors on negative match items, that is items where the most similar previously seen items is in the opposite category to the item in question. A reanalysis of conditions from Experiments 1 and 3 suggests that analogy-based responding is not critically dependent on a very high proportion of old items in the test set: A comparison of two different test lists showed more analogy-based responding in the case where the test set was made up almost entirely of precisely old items, but there was also evidence of analogy based responding when this was not the case.

It is apparent from the results of Experiment 4 that analogy to prior episodes continues to be used in a situation where the episodic information is at odds with the rule classification the majority of the time: This situation, which one might expect to cause exclusive reliance on the perfectly predictive rule, seemed instead to cause subjects to use a strategy that incorporated use of memory for prior episodes and switching the response to make up for the deficit in the episodic information that was available to them. The results of Experiment 5 suggest that the stimuli that were earlier demonstrated to be more amenable to an analogy strategy (integrated, memorable items) led to superior performance under conditions where the rule was no longer perfectly predictive, rather, more like a rule of thumb.

Finally, in Experiment 6 the developmental priority of a similarity strategy over a rule based strategy was investigated by presenting the basic experiment to grade 3 children. Contrary to studies where rule induction was the focus, the children in this rule learning situation appear to balance the rule and episodic information in the same way as do the adults.

Chapter 2

THE BASIC FINDING

Experiment 1 is designed to be a demonstration of the effect of prior episodes on classification performance in the presence of a simple sufficient rule. Two groups were run: the rule group was given a completely adequate rule at the beginning of the session, the no rule group was not. Both groups were given the same initial training session in which they were given feedback about the correctness of their categorizations. In the subsequent test phase if the rule group is influenced by similarity to specific training items, then they should make more errors on negative matches than on positive matches despite the rule applying equally well to both. If the no rule group is responding on the basis of similarity to old items rather than inducing a rule, then a very large number of errors are expected on the negative matches because this group would have no other basis of response. To the extent that these no rule subjects do induce a rule during the training, then they should behave like the rule group (and consequently be less useful for our purposes).

As mentioned in the introduction, our initial interest is in conditions that are favorable for making specific memories highly available. Two experimental arrangements not mentioned in the introduction were designed to further this end: the subjects were required to learn some idiosyncratic information about each of the training items, and the test items were preceded by context-setting background-only slides. Subsequent experiments will demonstrate that these two conditions are not essential for the current task, although one could easily imagine that they would be important for producing episodic effects in other tasks.

Experiment 1: The Initial Demonstration

Method

Subjects. The subjects were 80 students from the introductory cognition course at McMaster university, participating for course credit.

Materials. The stimuli were line drawings of imaginary animals (also used by Brooks, 1978) made up from five two-valued dimensions: Body shape (angular or curved), spots (present or absent), leg length (long or short), neck length (long or short), and number of legs (two or six). The animals were divided into builders and diggers by a three feature additive rule that used body shape, leg length, and spots as relevant features. An item was defined as a builder if it had builder values on at least two of the three relevant features and a digger otherwise. A logical description of the stimuli is given in the Appendix. As can be seen, the value of each relevant feature was consistent with the classification of the animal in 75% of the cases. The values of the irrelevant features appeared equally often in animals of each category and thus were, in themselves, nondiagnostic. Each animal differs from each of the others on at least two attributes, with the exception of its matching item (as defined below) which differs only on the spots dimension. It is important that the feature that is varied to generate the corresponding items is one of the features relevant to the rule. If this were not the case then any transfer to new items might be regarded as simply a failure of discrimination between the old and new items. Making the discriminating feature one of the features relevant to the rule makes this possibility much less likely.

Four different rules were used in order to counterbalance the items across the conditions. The values associated with builders for each of the four rules were 1) Long legs, angular body, and spots; 2) Short legs, angular body, and no spots; 3) Short legs, curved body, and spots; and 4) Long legs, curved body, and no spots. This manipulation of rules, along with varying which items appeared in the training set and which in the test set, ensured that each of the items in the experiment appeared equally often in the following four critical conditions.

(1) <u>Positive Match</u>: An item seen for the first time in the test phase and which is in the SAME category as its matching old item. Such an item appears on the same background as the matching old item and differs from it only in the presence or absence of spots, a feature defined by the rule as relevant to the categorization.

(2) <u>Negative Match</u>: An item seen for the first time in the test phase and which is in the OPPOSITE category to its matching old item. As with the positive matches, a negative matching item appears on the same background as the matching old item and differs from it only in the presence or absence of spots.

(3) <u>Positive Olds</u>: Items seen in the training phase and for which the matching new item will be a positive match (e.g. upper left panel in Figure 1).

(4) Negative Olds: Items seen in the training phase and for which the matching new item will be a negative match (e.g. lower left panel in Figure 1).

The items appeared on one of four colored background scenes. Each background in the study set was home to one type of digger and one type of builder so that the backgrounds themselves were not differentially associated with builders or diggers. The association of each animal with a background as well as the counterbalancing across rules are given in the Appendix.

Procedure. Subjects were tested individually. The materials were all presented by means of a slide projector connected with an Apple II microcomputer. The projector and computer were connected through a light sensitive switch so that the computer recorded the elapsed time from the projection of a slide to the subject's response of pressing a telegraph key.

There were two conditions with 40 subjects each: The rule condition and the no rule condition. For both groups there was a training phase and a test phase. The subjects in the rule condition were informed of the rule at the start of the training phase. Those in the no rule condition were never informed of the presence of a rule and were told that the first time they saw an animal they would have to guess whether it was a builder or a digger but on subsequent trials they would be able to remember what it was.

In the training phase the subjects were given 40 trials made up of the eight old items seen on five trials each in an ordering that was random with the restriction that the same item not appear twice in a row. Each training trial consisted of three slides. The first slide showed a pair of animals simply standing in a given background (as in Figure 1). Subjects were instructed to classify the animals as quickly as possible without sacrificing accuracy. After their response they were given feedback and, if it was the first presentation of a particular animal, were shown the second two slides in the set. These second and third slides showed the particular way each animal built or dug their homes. On all subsequent presentations of each animal in the training phase, the subjects were again required to make the classification decision upon seeing the first slide but were then asked to remember how that animal built or dug before seeing the two "story" slides again. Presenting the second and third slides and requiring recall of information from them was intended to individuate the animals and prevent them from being processed solely as instances of the rules.

The test set was identical for both groups. There were 40 items, split into a phase in which the only new items were positive matches and a subsequent phase in which the only new items were negative matches. This separation into a "positive phase" and a "negative phase" was intended to allow us to evaluate a possible "generalized caution" effect when the subject discovered the presence of negative matches (although, as will be demonstrated subsequently, this separation proved unnecessary). The positive phase contained the four positive old items, the four positive matches, and four repetitions of each of the negative old items. The four repetitions of the negative old items were performing as fillers, holding the positive old and positive match items apart in the series as well as increasing the ratio of old to new items. The negative phase contained three further repetitions of each of the positive old items shown in the positive phase, now acting as fillers, and the four negative match items. There were a total of 32 old and 8 new items in the test set as a whole. The negative phase immediately followed the positive phase with no break between the two. In both halves the filler items were those old items that were least similar to the new matching items so that there would be less chance for immediate interference or enhancement from the corresponding old item.

The items in the test phase were presented as pairs of slides. The first slide showed only the background on which the upcoming test item would be displayed. The subjects were to simply look at this background and indicate when they were ready for the second slide. The second slide showed the same background as well as a pair of animals on it. As in training, subjects were instructed to classify the animals as quickly as possible without sacrificing accuracy. They were also told that they might be able to use the first slide to try to anticipate which items were most likely to appear on that background and that there would be some new items in this part. No feedback was given in the test phase.

Results

Analysis. The criterion alpha level was set to .05 for all analyses. The predicted results were elevated response times and error rates for the negative matches in comparison to the positive matches. However, a simple comparison between the positive and negative matches is not appropriate because of practice effects that occurred during the test set. Therefore the analyses were set up as two 2 X 2 analyses of variance, one for the positive phase and one for the negative phase. For each of these analyses, knowledge of the rule was a between subject factor (rule vs. no rule) and previous experience with the item (old vs. positive match or old vs. negative match) was a within subject factor. In the positive phase analysis, the comparison was between the positive old items and the positive matches; in the negative phase analysis, the comparison was between the first negative phase presentation of the positive old items and the negative matches.[1] Only those positive items (positive olds and positive matches) that had two consistent features (either one or two builder features) were included in the analyses in order to avoid a potential confounding due to the fact that for half of the positive items all three features are consistently builder or digger features whereas this is true for none of the negative items (negative olds and negative matches). This more conservative test was used, although the findings are the same whether only the items with two consistent features or all the items are included in the analysis.

The mean correct response time for each item type was calculated for each subject and the response time analyses were performed on the logarithms of these times in order to better comply with the assumption of normal distributions

1 This negative phase comparison was used in order to generate the most conservative test. Comparison of the negative matches with an additional presentation of the negative old items in the negative phase would have been inappropriate due to the use of the negative old items as fillers in the positive phase and the consequent practice effects on those items. Similarily, comparison of the negative matches with the first presentation of the negative old items in the positive phase was inappropriate because of generalized practice effects.

required for the analysis of variance. For the negative match items for the no rule group, the response times were calculated for incorrect rather than correct responses. For these subjects, who were not aware of the rule, wrong answers on the negative match items are actually the most appropriate responses. (Whether correct or incorrect answers are used for this cell is not critical to the results because whether the mean correct, 1149 ms, or incorrect, 1120 ms, response time is used the difference between the old and matching items for the no rule condition is in the opposite direction than for the rule condition).

Errors. The error data are shown in Figure 2. There were no differences between any of the items in the positive phase (all p's > .05). In the negative phase there were more errors for the no rule group than for the rule group, F(1,78) = 27.8, MSe = .05; more errors for matching items than for the old items, F(1,78) = 192.3, MSe = .06; and an interaction between rule presence and previous experience such that there was a larger difference between the old and matching items in the no rule condition than in the rule condition, F(1,78) = 34.8, MSe = .06.

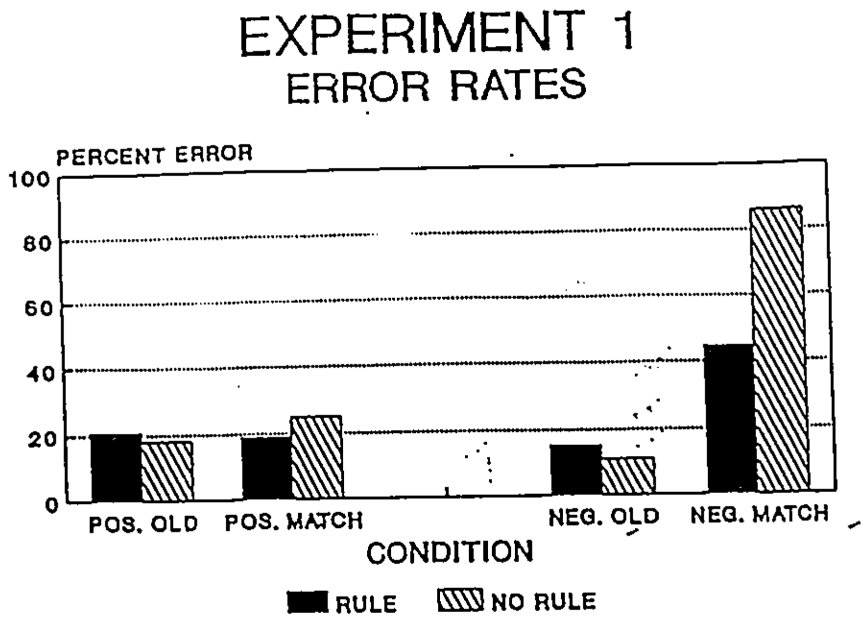
Response times. Seven subjects, six from the rule group and one from the no rule group, were dropped from the response time analysis because of empty cells.[2] The response time data for correct responses are shown in Figure 2. As with the error data there were no differences in the positive phase (all F's < 1). In the negative phase, times were longer for the rule group, F(1,71) = 4.6,

2 These were subjects who failed to correctly classify any of the items in one of the cells. In the rule group three subjects failed to correctly classify a negative match and one subject each failed to classify a positive old item, a negative old item, and a positive match. One subject from the no rule group failed to correctly classify a positive old item.

MSe = .319; longer for the matching items, $F(1,71) = 5.5$, MSe = .08; and there was an interaction between rule presence and previous experience such that the matching items took longer to classify than the old items in the rule condition whereas the opposite was true in the no rule condition, $F(1,71) = 8.0$, MSe = .08. The mean response times for the incorrect negative matches, the only items for which enough errors were made to obtain a reasonable estimate of response times, were 1172 ms for the rule group and 1120 ms for the no rule group (see previous discussion of response times for the no rule group negative matches).

Discussion

The main result of this experiment for both the rule and the no rule groups is the large increase in errors on the negative matches but not on the positive matches. This difference is consistent with the use of similarity to previously seen items as a classification strategy and is not what would be expected if subjects were simply using the rule they were given or if, in the case of the no rule subjects, they were relying on a rule that they had induced during training. On the other hand, both the accuracy and the speed results suggest that the subjects were not just ignoring the rules and simply using a similarity strategy. The no rule group made 86% errors on the negative matches (or 14% responses consistent with the rule) as opposed to only 45% errors made by the rule group. For the response times, there was a substantial slowing of the responses for the negative matches which was not evident in the no rule group, suggesting that for the rule group the rule based response and the similarity based response were in conflict. This



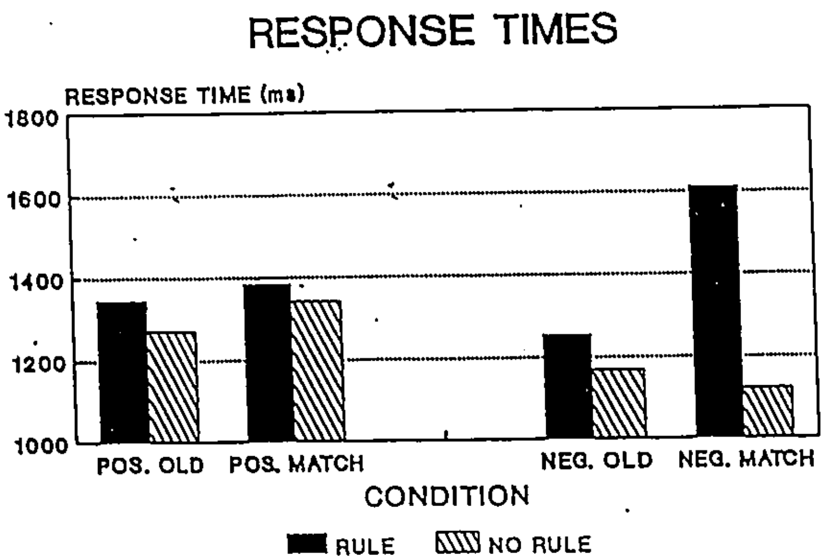


Figure 2: Results for Experiment 2. For the group that knew the rule, both errors and response times were increased for the negative matching items but were unaffected for the positive matching items relative to the Old (training) items. The rule applied equally well to both the positive and negative matching items. Although the No Rule group did not know the rule and therefore were not technically making errors, the results are plotted as percent responses inconsistent with the rule for easier comparison with the Rule group.

increase in response times would not be expected in the no rule condition, assuming that a rule was not induced that conflicted with the similarity-based response.

This reliance on similarity to specific prior episodes occurs in a situation that does not bias against the use of the rule by withholding the rule from the subjects or by supplying them with an extremely complex rule. There was no necessity in this experiment for the subjects to use anything but the rules they were given. The rules were perfectly predictive, yet the subjects made significant errors when new items were similar to previously seen items from the opposite category. However, as noted in the introduction, this experiment was designed to facilitate the availability of specific instances. Subjects had idiosyncratic information in addition to the classification of an item (the story slides). The test was conducted in a situation where the subjects were familiar with most of the items they were seeing (80% old items), and the context was established prior to the item being shown (the background slides). Such conditions appear to mimic one kind of real world situation in a number of ways. The following experiments will explore the importance of these variables by examining the limiting conditions of the similarity-based specialization of the application of explicit rules demonstrated in this experiment.

Chapter 3

LIMITING CONDITIONS

In this chapter, evidence is presented that address the conditions under which the effects of similarity will occur. Experiment 2 demonstrates that configural, memorable units are important to the use of an analogy strategy. Experiment 3 demonstrates that the effect appears even when accuracy is stressed and the subjects are paying a large premium in response times to avoid errors. Finally, a reanalysis of conditions from Experiments 1 and 3 suggests that analogy-based responding is facilitated by having a very high proportion of old items in the test set but is not critically dependent on a background of familiarity. This finding is consistent with an episodic basis of response, since overall familiarity of the test list has been demonstrated to be an important controlling factor in research on episodic memory.

Experiment 2: The Effect of Perceptual Integration

Experiment 1 demonstrated that applying a rule to a restricted set of instances can have the effect of establishing item specific bases of generalization. However, there is no reason to presume that practice with a small number of exemplars always has this effect of specializing the application of the rule. Intuitively, when an explicit rule is practiced, the effect often is simply for the person to become faster at applying the rule. One set of variables that could be expected to control these different outcomes of practice is the perceptual and mnemonic characteristics of the exemplars. For example, the list of features shown in Figure 3 is logically equivalent to the explicit dimensional structure of the animals used in Experiment 1 and would allow the application of identical classification rules. However, one could easily expect that presenting the instances in this form would not lead to the information being integrated into mnemonically separate instances. Rather, subjects might have a greater tendency to treat the displays as separable pieces of information and not learn them incidentally as special cases If this happened, then one would not expect practice on the feature lists to result in differences in performance between positive and negative matches, nor for that matter, between new and old items.

The second contribution of this experiment is to show the robustness of the basic demonstration from Experiment 1. In the current experiment, since the story slides obviously were not appropriate to the feature lists, they were not included for the drawn animals either. Further, given that the current experiment had in fact been run prior to Experiment 1 and included several types of test items that are not of current interest, there are fewer old items in the test phases. Finally, the context for individual items in the transfer set was not established by pre-exposure to background-only slides. Incidentally then, this experiment will demonstrate that the "story" slides of Experiment 1, the large proportion of old items during the test phase, and the background-only slides used during the test phase are not essential to producing the similarity effects with the animal stimuli.

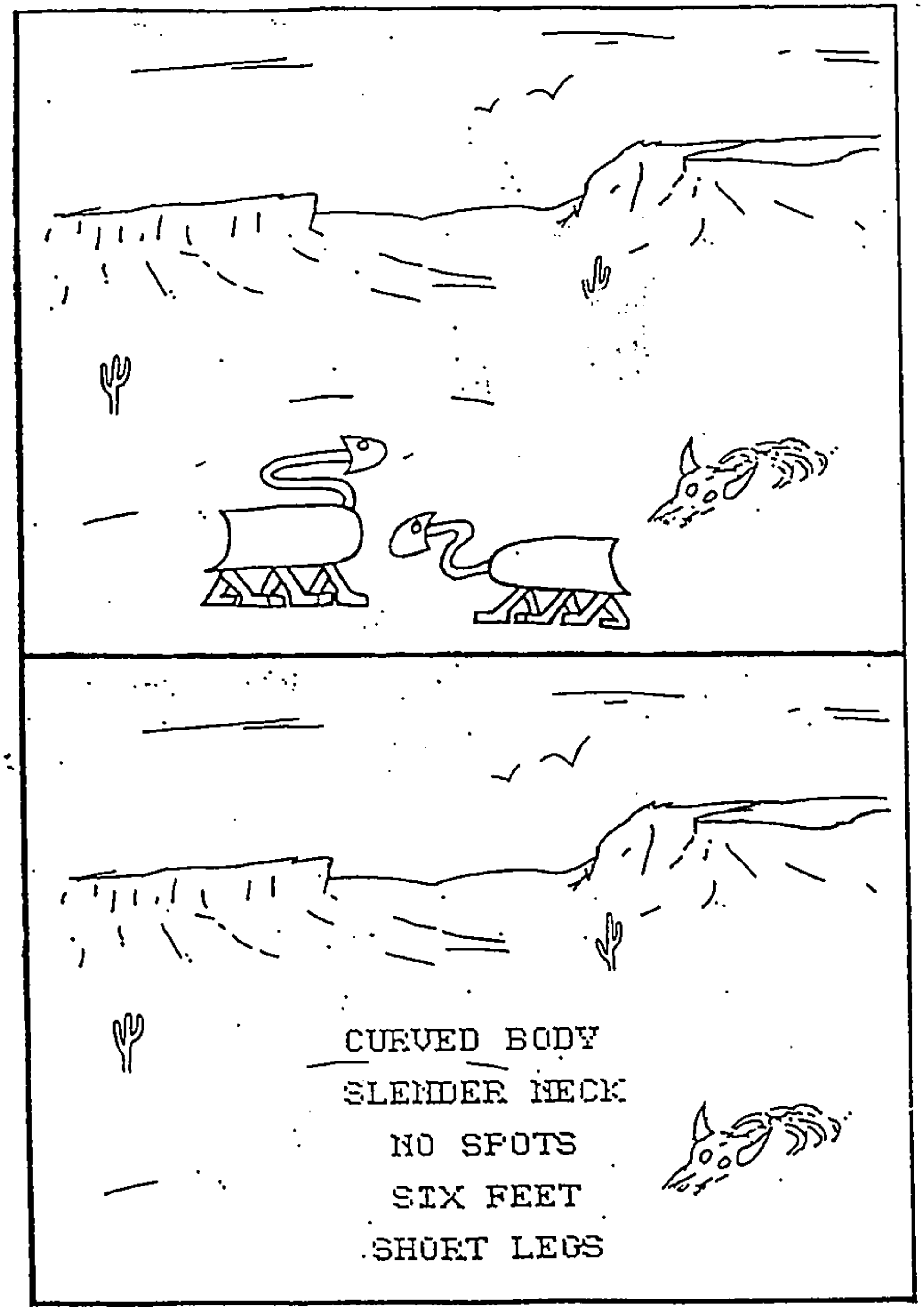


Figure 3: An example of the materials used in Experiment 2. The Drawing group saw stimulus items similar to those used in Experiment 1. The Feature List goup saw items in which the same sets of features were presented as verbal descriptions rather than pictorially.

Method

Subjects. The subjects were 72 McMaster undergraduate summer students who were paid for their participation.

Materials. For the drawing group, the stimuli were the same stimuli used in Experiment 1. For the feature list group, the stimuli were logically equivalent lists of features rather than drawings. The features were always listed in the same order. The training set consisted of 8 items as in the first experiment, however in the present experiment only one rule was used to define builders and diggers.

The test set contained more types of items in this experiment than in Experiment 1. These differences will be listed but not expanded on here since they are not important to the present discussion. In addition to varying as old items and positive and negative matches, the animals in the test phases could be presented on appropriate or inappropriate backgrounds. An appropriate background was the background on which an animal, or its matching old item, had been seen in the training set. An inappropriate background would be one of the other three backgrounds. The test set was broken into two parts. The first part of the test set contained the following items all presented in random order: the eight old items on the appropriate backgrounds, the four positive matches on appropriate backgrounds, the four positive old items on inappropriate backgrounds, and the four positive matches on inappropriate backgrounds. The second part of the test set consisted of: A second presentation of the eight old items on the appropriate backgrounds, a second presentation of the four positive matches on appropriate backgrounds, a second presentation of the four positive matches on inappropriate backgrounds, the first presentation of the four negative matches on appropriate backgrounds, and the first presentation of the four negative matches on inappropriate backgrounds. Thus, across the two test sets there were 16 old (a previously encountered animal on the previously encountered background) and 28 new items. The items used in the present analyses are the old items, the positive matches, and the negative matches, all on the appropriate backgrounds.

Procedure. The subject was first presented with a set of study items consisting of the eight old items. There were five repetitions of the eight items presented in five randomly ordered blocks for a total of 40 trials. The subject was informed of the rule and asked to classify each item by pressing one of the two response keys as quickly as possible without sacrificing accuracy. Feedback was given after each item. Then a new slide tray containing the first test set was loaded and the subject was instructed to continue doing the same task except that now there would be no feedback. When the first test set was completed another slide tray containing the second test set was loaded and again the subject was instructed to continue with the same task.

Results

Analysis. The criterion alpha level was set to .05 for all analyses. The analyses for this experiment are similar to those for Experiment 1. There were two 2 X 2 mixed ANOVA's, one for the positive items (positive olds and positive matches) and one for the negative items (negative olds and negative matches). The between subject factor was the form of the stimulus representation (animals versus feature lists), and the within subject factor was experience with the item (old versus positive or negative match). In all other respects the analysis was identical to that for Experiment 1. Two subjects, one each from the drawing and feature list conditions were dropped from the analysis of the response times because of missing data.[3]

Errors. The error data are shown in Table 1. For the positive phase there were no significant effects of the form of the stimuli or of previous experience or their interaction, all F's < 1. For the negative phase the drawing group made more errors than the feature list group, $F(1,70) = 8.42$, $MSe = 249.5$; the old items showed fewer errors than the negative matching items, $F(1,70) = 17.46$, $MSe = 223.7$; and there was an interaction between stimulus representation and previous experience, $F(1,70) = 13.12$, $MSe = 223.7$. Thus there was an increase in error rates for the negative matches only for the drawing group.

Response times. The response time data are shown in Table 1. For the positive phase the drawing group showed faster responses than the feature list Group, $F(1,68) = 27.75$, $MSe = .161$; old items showed faster responses than

3 None of the test 1 items nor the inapporopriate items from test 2 were analyzed. For the drawings group the mean percent errors for these items were 6.9, 5.6, 6.9, 5.6, 1.4, 5.6, and 25 for the negative old appropriate, positive old appropriate, positive match appropriate, positive old inappropriate, positive match inappropriate, second presentation positive match inappropriate, and negative match inappropriate items respectively. The corresponding mean correct RT's in ms were: 1303, 1355, 1382, 1599, 1543, 1421, and 1787. The corresponding means for the feature list group were, for errors; 6.3, 8.3, 6.9, 8.3, 5.6, 4.2, and 4.2; and for RTs; 1971, 1887, 1977, 1862, 1905, 1742, and 1921.

positive matching items. $\underline{F}(1,68) = 13.15$, $\underline{MSe} = .033$; and there was no interaction $\underline{F}(1,68) = 1.86$, $\underline{MSe} = .033$, n.s. For the negative phase the drawing group was again faster than the feature list group, $\underline{F}(1,68) = 5.26$, $\underline{MSe} = .164$; old items were classified faster than negative matching items, $\underline{F}(1,68) = 16.33$, $\underline{MSe} = .029$; and there was an interaction between stimulus representation and previous experience, $\underline{F}(1,68) = 16.07$, $\underline{MSe} = .029$. Thus there was an increase in response times for the negative matching items only for the drawing group. The incorrect response times for negative matches were 1224 ms for the drawing group and 2175 ms for the feature list group (although the mean for the feature list group should be viewed with some suspicion because it is based on so few observations).

Table 1

Mean responses in Experiment 2

	Percent Errors			
	Positive Phase		Negative Phase	
Group	Old	Match	Old	Match
Drawings	4.2	4.8	4.9	24.3
Feature Lists	4.2	4.2	6.3	7.6
	Correct Response Times (ms)			
Drawings	1192	1413	1368	1691
Feature Lists	1750	1881	1752	1751

An additional 2 X 2 within subjects ANOVA on response times was performed on only the old items from the feature list group where the factors were

Test (first or second) and status of the matching item (positive or negative). This analysis compares the response times for the first presentation of the old items during test, 1887 ms for the positive olds and 1971 ms for the negative olds, with the times for their second presentation, 1751 ms for the positive olds and 1753 ms for the negative olds. The only significant effect was that of Test where the first test showed slower response times than the second, $\underline{F}(1,34) = 12.33$, $\underline{MSe} = .02$, suggesting that the subjects were getting faster at answering in the feature list condition even though they weren't showing signs of nonanalytic specialization (no difference in performance on positive matches and negative matches).

Discussion

Subjects who saw drawings of the animals in both training and test showed evidence of instance based specialization of the application of the rule, namely increased errors and response times for the negative matches. Subjects who had an equivalent number of exposures to feature lists in both training and test did not show this evidence of instance based specialization. For these subjects the effect of practice was simply to speed up the rule application. These results suggest that evidence of instance based classification is more likely with stimuli that can easily be integrated into mnemonically distinct instances. The subjects who saw drawings were able to respond very much more rapidly with an equivalent number of training trials, but apparently at the cost of greater vulnerability to negative matches when they occurred.

In addition, the similarity between the results of the drawing condition in this experiment and of the rule condition in Experiment 1 suggests a few things that are not in themselves critical to the occurrence of instance based specialization of the rule. These are: Information in training other than category membership (the "story" slides in the training set), establishing the context before the appearance of the animal (the background-only slides in the test set), and the very high proportion of old items in Experiment 1 (32/40 in Experiment 1 as compared to 16/44 in this experiment). The next experiment and the subsequent reanalysis will demonstrate that the similarity effects are robust across other potentially limiting features of the initial demonstration.

Experiment 3: Caution and Awareness of Error Source

In both of the previous experiments, subjects were encouraged to respond as rapidly as was consistent with accuracy. It is possible that the intrusion of specific familiarity on rule-governed classification is limited to conditions with this emphasis on speed. The purpose of the current experiment is to change the emphasis of the subjects increasingly toward accuracy to see where the effects of negative matches disappear. The first of three instructional groups, the speed group, had identical instructions to those used in Experiments 1 and 2. With the exception of a list context manipulation that will be described in the method section, this group is a replication of the Rules condition of Experiment 1. The accuracy group was told that their main concern was to be with accurate classification and were given feedback on any errors in the test phase with a re-emphasis of the need for accuracy. Finally, the alert (to negative matches) group, in addition to having accuracy instruction and feedback, was told of the presence of negative matches and had the first negative match pointed out to them immediately after they classified it. In the previous two experiments subjects almost invariably immediately commented on any errors that they made, even though they were not given specific feedback during the test phase. However, without the current manipulations it is not clear whether this realization that a mistake had been made would be translated by the subjects into an emphasis on accuracy or that they realized the source of their mistakes on negative matches.

Method

Subjects. The subjects were 120 students from the introductory cognition course at McMaster University who participated for course credit.

Materials. The materials were exactly those used in Experiment 1 except for the items used as fillers in the test set. The makeup of the training set was the same for all three conditions and identical to the training set for the Rule group in experiment 1. The test set was also identical to the test set in Experiment 1 with the exception that the filler items were all seen on inappropriate backgrounds. In Experiment 1 all items were always seen on the appropriate background, that is the background on which that item or its matching old item was seen in training. For reasons that will be discussed in the subsequent section, I wanted to decrease the proportion of precisely old items in the test. To accomplish this, all the filler items in the present experiment were seen on inappropriate backgrounds so that the only items seen on the appropriate backgrounds were the critical items. Thus the positive phase of the test set was made up of the four positive old items, the four positive matching items, and four repetitions of each of the negative old items seen on a different inappropriate background for each repetition (except for the fourth presentation on which the first inappropriate background had to be repeated since there were only three inappropriate backgrounds). The negative phase of the test set was made up of the four negative matching items along with three repetitions of each of the positive old items seen on a different inappropriate background each time. It is worth repeating that the old items are not themselves positive or negative with respect to the rule, but only that their matching new item would, according to the rule, be in the same or opposite category.

Procedure. There were three conditions in the experiment with 40 subjects participating in each group. The groups differed in the instructions they received prior to being given the training set and again prior to the test set. The speed group was given the same instructions as the rule group from Experiment 1, that is, to classify the items as quickly as possible without sacrificing accuracy. The accuracy group was given instructions that emphasized accuracy. They were told that response times were also being measured but that this was of secondary importance and that their main concern should be with accuracy. These subjects were also given feedback regarding the accuracy of their responses in the test phase, again emphasizing the importance of accuracy. The instructions to the subjects in the alert condition were identical to those in the accuracy condition except that they were alerted to the presence of negative matches. They were told at the beginning of the test sequence that there would be some new items in the test set that closely resembled items they had seen before but that would in fact be in the opposite category and that they should watch out for these items. In addition, when the first negative match was presented it was pointed out to the subject by the experimenter that it was one of the "tricky" items. The experimenter told the subject which old item it was similar to, pointed out that it was in the opposite category and again warned the subject to watch out for this type of item.

Results

The criterion alpha level was set to .05 for all analyses. The analysis of the results of this experiment, like the materials and design, was very similar to that of Experiment 1. The data were subjected to two 3 X 2 mixed Anovas, one for the positive phase and one for the negative phase, where the between subjects factor was the instructional condition (speed vs. accuracy vs. alert) and the within subject factor was item type (old vs. match). Two subjects from each of the speed and accuracy groups and one from the alert group had to be dropped from the response time analyses because of empty cells. All other aspects of data analysis were as in Experiment 1.

Errors. The error data are shown in Table 2. In the positive phase there was a marginal interaction between the instructions and item type. $F(2,117) = 3.02$, $MSe = .05$, $.05 < p < .10$. Neither main effect approached significance. In the negative phase there was no main effect of the instructions nor was there an instruction by item-type interaction. There was, however, a significant effect of item type with the subjects showing more errors on the negative matching items than on the negative old items, F(1,117) = 35.14, MSe = .04.

Table 2

Mean responses in Experiment 3

	Percent Errors			
	Positive Phase		Negative Phase	
Group	Old	Match	Old	Match
Speed	13	13	6	28
Accuracy	5	19	8	19
Alert	9	8	4	18
	Correct Response Times (ms)			
Speed	1404	1609	1345	1348
Accuracy	1469	1478	1352	1519
Alert	2012	2157	2003	1936

Response times. The response time data are displayed in Table 2. In the positive phase there was a significant main effect of instructions, F(2,112) = 11.88, MSe = .25, as well as a main effect of item type F(1,112) = 11, MSe = .07. There was no interaction between the two. In the negative phase there was a main effect of instructions F(2,112) = 17.49, MSe = .19, but the item type and interaction effects were not significant. The incorrect response times for the negative matching items were 1254 ms, 1146 ms, and 1535 ms for the speed, accuracy, and alert conditions respectively.

Discussion

The increase in errors on the negative matches in both the accuracy and alert groups suggests that the instance based effects in the speed group (and the comparable groups in Experiments 1 and 2) were not entirely due to an emphasis on speed on the part of the subjects. When subjects were asked to concentrate on accuracy and were given feedback in order to point out their errors, there was little change from the results under the "speed without sacrificing accuracy" instructions given to the speed group and to subjects in the previous studies. When subjects were alerted to the nature of the difficult items they would encounter they decreased their speed considerably but gained little in their accuracy and continued to show increased errors on the negative matching items. This pattern of results suggests that an effect of specific training items intrudes despite at least some effort on the part of the subject to prevent it. This is especially true in the alert condition where the subjects should have been alerted to the problem but still seemed unable to protect themselves.

The high error rate for positive matches in the accuracy condition is problematic. This was unexpected and doesn't fit with either a rule based or a similarity based strategy. This is the only time that such a difference between old and positive matching items has been demonstrated. The increase in error rates for these items is accompanied by a lesser increase in response times than in the speed or alert conditions and may represent a speed-accuracy tradeoff favoring speed for these items. This is not a particularly satisfying explanation though because the tradeoff is in the opposite direction to what the instructions required. However, even though there is an approximately equal increase in errors on both positive and negative matches in the accuracy condition, it seems clear that overall there is a difference in the two types of items; the increase in errors on positive matches disappears in the alert condition but the increase on the negative matches does not.

Additional evidence from the accuracy group suggests that the similarity information intrudes without the subject even being particularly aware of it. In a post experimental interview, 24 subjects from the accuracy condition were questioned regarding which items gave them problems and why they thought they might be faster on some items than on others. They were shown pairs of items consisting of one positive match and one negative match and told they answered faster on the positive match. They were asked why they thought this might be the case. Only 10 of the 24 subjects ever mentioned similarity to an old item in the opposite category as a possible reason for the relatively poor performance on the negative matches. Furthermore, those subjects who were counted as mentioning similarity were those who made any mention of similarity at any point in the discussion, often after considerable prompting from the experimenter. Thus, even using the most generous of criteria, over half of the subjects failed to place the blame for their poor performance on those items on the fact that they were negative matches. They generally picked features of the animal that seemed to be plausible causes. For example "this one had a long neck so I must have thought it had long legs and called it a builder" and "this one had long curvy legs and so I guess I got them mixed up with its body which is angular and called it a builder by mistake". Occasionally the subject couldn't find a plausible reason in the slide and would say "I guess I was trying to answer too fast" which of course is not a reason for relative decrements in performance. The lack of awareness of the error source as well as the subjects' inability to protect themselves from similarity to previous items when told about it suggests that similarity to old items may be an intrusive process that doesn't require a deliberate decision to occur.

Reanalysis: Familiarity of List Context

One of the ecological constraints described in the introduction was a test context that contained a predominance of familiar items. In this section two groups from previous experiments will be compared to evaluate the effect of this variable. The rule group from Experiment 1 and the speed group from Experiment 3 differed only in the relative familiarity of the filler items. The rule group from Experiment 1, referred to below as friends list context, used old items as fillers. This resulted in 32 of the 40 item test list being items seen in exactly the same form as in training. The filler test items for the speed group from Experiment 3, referred to below as strangers list context, consisted of old animals appearing on a variety of different backgrounds (a different background for each trial in which a given filler animal appeared). This re-pairing of animals and backgrounds was designed to provide a list context that was less familiar (4 old items, 8 new matching items, and 28 re-paired items) than with the friends list context. This changed list context could make the subject rely on the rule more, or possibly even make the training items less available -- analogous to the mnemonic deficit resulting from changing list context in memory experiments (a comparable effect of familiar list context was found for the perceptual identification of words by Jacoby, 1983b). Although the list contexts (filler items) differed between the two test lists, the analyses below (and in the previous experiments) are based only on those 12 critical items that were identical in the two lists.

Analysis. The data are presented in Table 3. The analyses consisted of two 2 X 2 mixed Anovas with list context (friends vs. strangers) as the between subjects factor and Item Type (old vs. match) as the within subject factor. The criterion alpha level was set to .05 for all analyses.

Errors. In the positive phase there was a marginal effect of the list context with more errors in the friends condition, $F(1,78) = 3.31$, $MSe = .07$, $.05 < p < .10$, and no effect of item type or the list context by item type interaction, both F's < 1. So that, although there was a trend toward an advantage for the strangers group, there was no advantage for old items or interaction between list context and item type. In the negative phase there was a significant effect of list context, $F(1,78) = 11.15$, $MSe = .06$, and a significant effect of item type, $F(1,78) = 50.59$, $MSe = .05$, but no interaction between the two, $F(1,78) = 1.24$, $MSe = .05$, n.s.. Thus, in the negative phase, there was an advantage for the strangers group and an advantage for old items.

Response times. In the positive phase there was no overall effect of either the list context, $F(1,70) = 1.56$, $MSe = .35$, n.s.. or item type $F(1,70) = 1.48$, $MSe = .09$, n.s., while their interaction was marginally significant, $F(1,70) = 3.23$, $MSe = .09$, $.05<p<.10$. In the negative phase there was again no overall effect of list context, $F < 1$, but there was a main effect of item type, $F(1,70) = 6.64$, $MSe = .08$, and an interaction, $\underline{F}(1,70) = 6.01$, $\underline{MSe} = .08$. Thus, in the negative phase, subjects responded more slowly on the matching items and this was due to longer response times for the negative matches in the friends condition but not in the strangers condition.

Table 3

<u>Mean responses in Reanalysis</u>

	Percent Errors			
	Positive Phase		Negative Phase	
Group	Old	Match	Old	Match
Friends	12	19	15	45
Strangers	13	13	6	28
	Correct Response Times (ms)			
Friends	1345	1385	1253	1612
Strangers	1404	1609	1345	1348

<u>Discussion</u>

Clearly the list context during test did influence instance-based performance. The negative matches in the friends list context showed many more errors than in the strangers list context. However, the basis of this effect is not as clear. The simplest result to support the hypothesis that list context affects the <u>availability</u> of prior exem; .ars would be to have olds and positive matches both faster and more accurate and the negative matches more disrupting in the friends condition than in the strangers condition. This increased disruption did happen for the negative matches, but the results for the olds and the positive matches showed signs of a speed-accuracy tradeoff. Negative olds, positive olds, and positive matches were all in the direction of more errors and faster times in the friends than in the strangers condition. A post-hoc analysis of only the old items showed a significant effect of list context with subjects making more errors in the friends list context, $F(1,78) = 5.12$, MSe $= .06$, and a marginal effect of test phase with subjects performing better in the negative phase than in the positive phase, $F(1,78) = 3.17$, MSe $= .05$, $.05<p<.10$. The interaction did not approach significance and a similar analysis on the response times showed no significant effects. Even though the planned comparisons did not reach significance, the direction of the differences was consistently in the opposite direction to that predicted by a simple effect of increased exemplar availability, a reversal that should not be ignored. However, even if there were a lower criterion for accuracy, this would not exclude an effect based on instance availability. It would not be unreasonable for a more familiar list context to both result in relaxed criteria (leading to somewhat faster and less accurate responses) and to provide more available prior instances. Regardless of the basis, what is clear is that the familiarity of the context is important in determining the relative contributions of rule-based and of instance-based generalization.

Chapter 4

CONFLICTS BETWEEN RULE INFORMATION AND ITEM HISTORY

The rule has been held constant in the preceding experiments. But in some cases in the world, the rule is either an approximation to be superceded or is changed in the light of new understandings. Such circumstances raise the question of the role of prior episodes in the resulting new mix. In this chapter, changes in the rule are used to clarify the relations between rules and episodic knowledge.

Experiment 4: Reversing the Rule to Produce Conflict with Episodes

In previous experiments, rule governed responding was set at odds with item history by presenting negative matches, items that belong in one category according to the rule but look like a previously seen member of the opposite category. In this experiment, a partial reversal shift procedure (e.g. Robbins, et al, 1978) is used to set rule based responding at odds with item history. In the partial reversal shift paradigm, a classification rule is taught, and then the rule is reversed with some items being given practice under the reversed rule (items extinguished for the previous rule) and others not (unextinguished items). This sets up a conflict between the previous experience obtained under the old rule and the current requirement to respond according to the new rule. This conflict should be especially strong for the unextinguished items, those that have not been practiced under the new rule. In this way performance differences between extinguished and unextinguished items become diagnostic of the influence of prior instances on the application of the current rule. This design allows us to examine a situation in which similarity is not benign, a situation in which familiarity and similarity are suspect for a large number of the items.

One could entertain two possible results of the comparison of extinguished and unextinguished items. If classification continues to be partly based on item history despite the accuracy advantage that could result from tight strategic control by the subject, one would expect to see better performance for the extinguished items, where the subjects have previous experience with the currently appropriate response. They should have special trouble with the unextinguished items, where all the previous experience is opposite to the current correct answer. This should be true for both the old and positive match items, although the prediction for the negative match items is not clear.[4] A second possibility is that the subject, under these confusing conditions, might resort to a strictly rule based strategy, possibly paying a cost in time. In this case there should be no difference between extinguished and unextinguished items as well as no difference between positive and negative matches. Performance should be determined only by the amount of practice on the rule overall, not on whether a given item had been practiced with the new rule.

---

4 The unextinguished items could show better performance for the negative matches because the switch in category from the old item to the matching item might counteract the switch due the rule change and the subject may output the correct response before making either of the adjustments. This is the "two wrongs make a right" argument. Alternatively, the two additional sources of opposing information may simply lead to a great deal of confusion causing especially poor performance on the negative match items. Regardless, the predictions, or lack thereof, for the negative match items are not critical to the experiment since the predictions for the positive matches are clear.

Method

Subjects. The subjects were 48 students from the introductory cognition course at McMaster University who participated for course credit. The subjects were run in groups of from 1 to 3 per session, each with his or her own set of response keys.

Materials. The stimuli were the same as were used in all the previous experiments. The same four rules as used in previous experiments were varied across subjects in order to avoid confounding the particular items with the type of item they represented.

Procedure. As in the accuracy and alert conditions of Experiment 3, the instructions stressed accuracy. Subjects were told the rule prior to the start of the training set and were then given 24 trials of the story training used in Experiments 1 and 3. On each story trial the subjects were shown a pair of leptons standing in their environment and asked to classify them. As soon as the subjects had done so, they were shown slides of the leptons engaged in building or digging. This story training phase consisted of a random sequence of the sets of 3 story trials for each of the 8 training items. The training with the initial rule was completed by giving the subjects practice on a speeded classification task in which the training items were presented twice each followed immediately by two extra presentations of each of half of the training items.

At this point the training was halted, and the experimenter announced that the new metric classification rule had taken effect, which meant that the characteristics that had been indicative of a builder now indicated a digger. All items that used to be builders were now diggers and vice versa. This was followed by two presentations of each of the practice items that had not had extra practice before the rule reversal. Thus at the end of the reversal training phase, the subjects had an equal amount of practice classifying all of the old items, but for the extinguished items two trials of this practice had been in classifying the items according to the new rule, whereas for the other half of the items. the unextinguished items, all the practice had been with the old rule.

This design resulted in 8 types of test items, defined by a factorial combination of old versus new items, extinguished versus unextinguished items, and positive versus negative matching items. Each subject saw two of each type of item twice each making a test set of 32 presentations. Since no "generalized caution" effect had emerged in the previous experiments, the test slides were integrated into a single random test sequence rather than separating them into a positive phase and a negative phase. The task at test was speeded classification with feedback. There were no background slides prior to the test slides in this experiment.

Results

Analysis. The mean number of errors and mean response times for this experiment are displayed in Table 4. The results for each dependent measure were analyzed in a 2 X 3 within subject ANOVA where the factors were extinction (extinguished or unextinguished) and Item Type (old, positive match, negative match). Only the data from the first presentation of each item were analyzed.

Because only the data from items with 1 or 2 builder features were included in the analysis, and the extinction manipulation further split the items among the cells, the two logical types of old items were combined to give a more stable estimate thus the means are based on 3 observations per subject for the old items (2 negative olds and 1 positive old), 1 observation per subject for the positive match items, and 2 observations per subject for the negative match items. This reduced number of observations per cell led to 12 subjects being excluded from the analysis of the response times due to missing data.

Errors. Analysis of the errors revealed no effect of Extinction or of the Extinction by Item Type interaction, both F's < 1. There was a significant main effect of Item Type, F(2,94) = 10.12, MSe = .062. A separate 2 X 2 ANOVA including only the analogy items showed that more errors were made on the negative matches than on the positive matches, F(1,94) = 9.34, MSe = .074.

Table 4

Mean responses in Experiment 4

		Percent Errors	
		Positive	Negative
Group	Old	Match	Match
Extinguished	8	10	22
Unextinguished	8	13	25
	Correct Response Times (ms)		
Extinguished	2316	2362	2423
Unextinguished	2594	2647	2315

Response times. For the response times there was a marginal effect of Extinction, $F(1,35) = 3.63$, $MSe = .104$, $.05<p<.1$, with unextinguished items in the direction of taking longer to classify. There was no main effect of Item Type, but it did interact marginally with Extinction, $F(2,70) = 2.59$, $MSe = .07$, $.05<p<.1$, suggesting that negative matches took more time to classify than olds or positive matches for extinguished items but less time for unextinguished items.

Discussion

The expected results for this experiment were either more difficulties for the unextinguished items than the extinguished items or strict rule control resulting in no differences among any of the item types (positive match, negative match, extinguished, or unextinguished). Neither pattern materialized; instead these results look remarkably like those of the other experiments. In retrospect it seems probable that two presentations of the extinguished items following the shift in the rule were not enough to significantly affect the original response. This experiment is being reported nonetheless because of some interesting observations relevant to the main theme of this thesis: specialization of rules around specific prior instances.

The subjects apparently developed a special policy for dealing with the intrusion of conflicting item history. The old and positive match items in this study are similar to the negative match items in the alert condition in Experiment 3: In both cases the rule and item history are at odds with each other, and in both cases the subjects know about the problem. The difference between the two experiments is that in this experiment conflict occurs for almost all the items while in the alert condition it occurs only for a few of the items. Possibly because of this, the subjects in the present experiment seemed to have developed a better way of protecting themselves against the conflict than in the previous experiments as shown by the lower error rates on the conflicting items. Apparently, when the exceptions outnumber the non-exceptions, they generate a policy to deal with the exceptions. This policy seems to have exacted a cost in time since the response times in this experiment are much longer than in any other of the experiments.

The nature of the subjects' policy for dealing with the conflict between item history and current rule is suggested by their reports that they remembered how that item, or the analogous old item, was classified according to the old rule, and then classified the item as belonging to the opposite category. This strategy would generate the observed pattern of results that the old and positive match items, for which the rule and item history are at odds, show better performance than the negative match items, where the rule and item history are consistent. If it were the case that the subjects had continued trying to treat the item history as conflict rather than the primary source of the response, one would expect that the olds and positive matches would show worse performance than the negative match items. Thus, the results suggest that a reasonable number of the subjects adopted the policy of reversing the decision as a way of dealing with intruding item history. This experiment then, in common with the others, shows evidence of classification performance being based both on rules and on similar prior items, even though there is no logical need for and indeed some cost from these specific analogies.

Experiment 5: Dealing with a Suspect Rule

In all the experiments so far the rule has had absolute authority. It has been perfectly predictive and presented to the subjects as such. But in many natural situations the rule is substantially less well established. A rule may be presented as only a rule of thumb, or may be a rule that a learner is in the process of developing. When a provisional rule such as these starts producing errors, memory for prior instances takes on a special status. Under most natural circumstances, when errors on new items suggest the need to modify a rule, it is unlikely that the responses to known items should be changed. In addition, if a new item looks sufficiently similar to a known item, it is worth considering classifying the new item on the basis of its similarity to the old item rather than on the basis of the suspect rule. That is, if the two items match on many characteristics and there is uncertainty about which characteristics are relevant, then it is a reasonable bet that the few non-matches are not sufficiently important to change the categorization. With a suspect rule, the use of a literal similarity strategy may be advantageous for both generating better classification performance during a transition period and as evidence to be used in attempting to frame a better rule.

In the present experiment the rule was originally presented to the subjects as a rule of thumb; it would be correct the majority of the time, but there would be a few exceptions. In no case did the classification of a previously encountered item change. The exception items were produced by making the "correct" response for the negative match items the response associated with the similar prior item (the negative olds). Thus the exceptions would be classified incorrectly if the rule were used but correctly if an appeal to the most similar old item were used. The non-exceptions should be classified correctly in either case. This feedback rule essentially reifies the ecological argument regarding the use of similarity strategy given in the preceding paragraph.

Classification on the basis of similarity to old items should work better if the old items are more memorable and overall similarity is more salient. Without mnemonic salience of instances, the learner might be more likely to respond to exceptions by changing the features specified in the rule. As in Experiment 2, the drawings version of the materials is set against the feature lists version with the prediction that there should be more evidence of similarity based decisions for the drawings version. With the feature lists, people might be more likely to alter their classification policy in such a way as to inadvertently reclassify some of the old items.

Method

Subjects. The subjects were 24 students from the introductory cognition course at McMaster University who participated for course credit. All subjects were tested individually.

Material. The stimuli used were the same as in all the previous experiments. Both the drawing and feature list versions of the materials were used. Once again the set of four different rules was used across subjects to avoid confounding the particular items with the logical type of item they represented.

Procedure. The subjects were divided into a feature list group and a drawing group. The feature list group saw the feature list descriptions of the animals described in Experiment 2 throughout the experiment. The drawing group saw the drawing stimuli throughout the experiment. Throughout the experiment, both groups were instructed to respond as rapidly as is consistent with accuracy.

The first phase of the experiment was designed to familiarize the subjects with the rule and some of the animals. The subjects were told the rule and then given a set of 8 animals to classify. For the feature list group the task was to classify each slide as a builder or digger with feedback being given following each slide. For the drawing group the procedure was the same as the procedure in the training phase of Experiment 1. At the end of the familiarization phase, all subjects had seen 8 items 5 times each and in addition the drawing group had idiosyncratic information about the behavior of each animal.

The second phase was the "drift" training phase (the correct rule "drifted" away from the original rule -- rather than dramatically changing). At the start of this phase the subjects were informed that the rule they had been given was a rule of thumb rather than a perfectly predictive rule; that some of the items they would see in this part would be exceptions to the rule and their job was to try to improve on the rule. They were then presented with a set of 16 slides of the following three types: (a) one presentation of each of the eight familiarization slides (olds), (b) two repetitions of each of two new non-exceptions, and (c) two repetitions of each of two new exceptions. With this distribution of items, as in previous experiments, there was a predominance of old items, and half of the new items were not exceptions. The new items were generated, as in previous experiments, by changing one relevant feature (the spots). The exception items were those items whose category membership changed when the relevant feature was changed (the negative matching items in previous experiments). Thus the exceptions would be classified incorrectly if the rule were used but correctly if an appeal to the most similar old item were used. The non-exceptions should be classified correctly in either case. The subjects classified the items and were given feedback following each slide.

The third phase was the test phase. Again there were 16 items: (a) one presentation of each of the eight old items, (b) two presentations of each of two new non-exceptions, (c) two presentations of each of two new exceptions. The new items were each presented once on the background on which the corresponding old items were seen in the familiarization set (appropriate background) and once on a background that had been seen with different animals (inappropriate background). The order of presentation for background appropriateness was counterbalanced across items with the restriction that half the new test items preceded their matching old item (with maximum average lag) and half followed them; there was no evident difference resulting from this balancing, so it was dropped from subsequent analyses. There was no feedback given in this section.

<u>Results</u>

The criterion alpha level was set to .05 for all analyses. The data are summarized in Table 5. Because of the small number of new items, nonparametric analyses were used throughout. For old items, the drawing group showed more correct (rule-based) answers in both the drift training and test phase, Mann-Whitney $U(12) = 28.5$ and $U(12) = 35.5$ respectively. In the drift training phase, both groups showed a higher level of rule-based responding for the first presentation of an exception item than for its second presentation: Using a Wilcoxon matched pairs signed ranks test, for the drawing group $T(12) = 1.5$, for the feature list group $T(12) = 10.5$. However, the level of rule-based responding on the second presentation of the exception items was higher for the feature list group than the drawing group, $U(12) = 32.5$. In the test phase, the feature list group showed a higher level of rule-based responding than the drawing group on exception items seen on the appropriate background, $U(12) = 37.5$, and there was a marginally higher level of rule-based responding within the drawing group on the inappropriate background items than on the appropriate background items, $T(12) = 19.5$, $.05<p<.10$.

<u>Discussion</u>

In the drift training phase, both the feature list group and the drawing group showed better performance (less rule-based) on the second presentation of the exception items. The subjects were evidently remembering the correct response from the first presentation of the exception items and using this to produce better performance on the second presentation. Apparently the drawings were easier to use for this purpose since the drawing group performed better than the feature list group on the second presentation (but not the first).

Table 5

Mean percent rule-based responses in Experiment 5

		Drift training Phase			
		New Non-Exceptions		New Exceptions	
		Presentation		Presentation	
Group	Old	1st	2nd	1st	2nd
Drawing + Story	99	100	96	88	33
Feature List	89	100	100	96	71
		Test Phase			
		New Non-Exceptions		New Exceptions	
		Background		Background	
Group	Old	App.	Inapp.	App.	Inapp.
Drawing + Story	97	92	100	54	75
Feature List	87	88	96	88	83

In the test phase there was better responding (less rule-based, more consistent with the classification of matching old items) for the drawing group than the feature list group for the exception items that appeared on the appropriate background; when the exception items appeared on an inappropriate background the amount of this difference was reduced. This finding indicates that the drawings group made more use of similarity based responding, where the subjects will override the rule response if a new item is similar enough to a previously seen item. The marginally better performance (less rule-based responses) for the drawing group on the appropriate background exceptions than on the inappropriate background exceptions, suggests, as did the previous "friends-strangers" analysis, that this similarity strategy is dependent on nominally irrelevant aspects of context.

The difference in performance on old items is also consistent with this interpretation. In both the drift training phase and the test phase the drawing group performed better than the feature list group on the old items (more rule-based responses -- correct for these items). This would follow if the drawing group could recognize the old items better and would thus be less likely to change their operative rule for these items. Because they have a way of dealing with exceptions (using similarity to old items) and are more likely to recognize old items, there is less pressure on this group to adjust a rule that is correct most of the time. The feature list group, on the other hand, faced with errors that cannot easily be corrected must change their operative rule and, in doing so, risk errors on previously seen but unrecognized items that were correctly classified in the past.

In summary, I have shown that the drawing stimuli, that were earlier demonstrated as showing more evidence of similarity based responding, led to superior performance under conditions where the rule is no longer perfectly predictive. Admittedly, the method used here for classifying the exceptions (all the negative matches were to be categorized according to similarity) was designed to generate better performance from a subject using a similarity strategy. However, if we accept the argument, made for example by Rosch (1978), that many natural categories have a clustered structure that is not particularly amenable to simple rules, then the type of feedback rule used here should not be ecologically uncommon. Looked at in this fashion, the present experiment demonstrated the subjects did not have any particular trouble with such a category structure provided that the stimuli were perceptually and mnemonically distinct. When the stimuli were the more separable feature lists, there is no evidence that similarity was used at all.

Chapter 5

DEVELOPMENTAL CHANGES

There is a substantial base of research in the developmental literature that suggests that a more "holistic" or global mode of processing precedes children's learning of analytic strategies for classification. e.g. Kemler 1983a,b; Kemler-Nelson, 1984; Kossan, 1981; Markman 1989). In Kemler-Nelson's (1984) terms, categorization based on overall similarity is a more "primitive" form than that based on an analytic use of criterial features and thus is "more prominent in the cognitive activities of young children"(p. 750). In her paper she presents evidence that younger children perform as well as older children when categories are defined by a strong family resemblance structure but more poorly when they are defined by a single criterial attribute thus supporting the developmental priority of nonanalytic, similarity based categorization.

Like most of the investigators studying categorization with adults, Kemler-Nelson, and others working in the area chose an experimental paradigm where subjects were not told a classification rule beforehand but were required to induce the rule for themselves. Although, as Kemler-Nelson has argued, this form of experiment mirrors many informal learning situations where rules and feedback are not immediately available, it may be a poorer match to some formal learning situations than the experiments reported in this document where the relevant classification rule is made available to the subjects from the beginning. The research of Kemler-Nelson and others (Kemler, 1983; Kossan, 1981) demonstrates that younger children are likely to perform better when trying to induce categories with a strong family resemblance structure, but they do not bear directly on the present question of the coordination of information from analytic and nonanalytic sources in learning to apply classification rules.

One of the implications of the work on category induction with children is that children are more comfortable dealing with overall similarity information than rule information. In a situation like the present experimental setup, both rule-based and episodic information are sufficient to allow correct classification in the training phase. If the children's preference for more wholistic similarity based classification translates readily from category induction to rule learning one might expect that younger children, who will have more difficulty with the rules, will rely more on the episodic information to make a classification decision. That is, we might expect that young children, although at a level where they could understand and apply the rule, might find it easier to solve the task by a more episodic based system and thus show us results that look more like the no-rule group of adults. If so this would suggest that the children, as suggested by Kemler-Nelson (1983a&b, 1984) and Kossan (1981), are more inclined to use a similarity based strategy for classification than are adults when faced with a situation where either strategy is possible. That is the materials are amenable to either type of strategy -- they form configural recognizable wholes and so allow for a similarity based strategy, and the features are simple and obvious and thus allow for an analytic strategy.

Experiment 6: Investigating the developmental priority of similarity

Method

Subjects. The subjects were 72 grade 3 students from elementary schools in the Saint Catharines area. Four subjects from the rule condition were replaced due to their failure to reach the learning criterion. One subject from each of the rule, no rule, and feature list conditions was replaced due to equipment malfunction.

Materials. The materials were exactly those used in Experiments 1 and 2. In the rule and no rule conditions the drawing stimuli described in Experiment 1 were used while the feature list condition employed the feature list stimuli described in Experiment 2. In all three conditions the training set consisted of a set of eight items seen anywhere from five to eight times. As in previous experiments four different rules were used to counterbalance items across logical item types. The test series was identical to the test series in Experiment 1 except that no background-only slides were used.

Procedure. The children were tested individually. The materials were all presented by means of a slide projector connected with an Apple microcomputer. The projector and computer were connected through a light sensitive switch so that the computer recorded the elapsed time from the projection of a slide to the subject's response of pressing a telegraph key.

There were three conditions with 24 subjects each: The rule condition, the no rule condition, and the feature list condition. For all groups there was a training phase followed by a test phase. The children in the rule condition and the feature list condition were informed of the rule at the start of the training phase. Those in the no rule condition were never informed of the presence of a rule and were told that the first time they saw an animal they would have to guess whether it was a builder or a digger but on subsequent trials they would be able to remember what it was.

For children in the rule and no rule conditions (those conditions employing the drawing materials) each training trial consisted of three slides as in Experiment 1. The first slide showed a pair of animals standing in a given background (as in Figure 1). Subjects were instructed to classify the animals as quickly as possible without sacrificing accuracy. After their response, they were given feedback and, if it was the first presentation of a particular animal, were shown the second two slides in the set and told a little story about how those animals built or dug (e.g. These ones use their unusual ability to breathe fire to build themselves little ice huts. They kick up a big pile of snow and then use little puffs of fire to melt it down and hollow it out until they have a little hut made out of ice. Then they lay down and rest inside the ice huts that they've built using their fire breath.). The second and third slides showed the particular way each animal built or dug their homes. On all subsequent presentations of each animal in the training phase, the subjects were again required to make the classification decision upon seeing the first slide but were then asked to remember how that animal built or dug before seeing the two "story" slides again. Presenting the second and third slides and requiring recall of information from them was intended to individuate the animals and prevent them from being processed solely as instances of the rules. Children in the feature list condition saw only one slide per training trial and were not given any story information.

A criterion of once through the full set of 8 training items with no classification errors was set. Each child continued repeating the training set, each time in a different order, for a minimum of 5 times through the training set and until criterion was reached. Children who had not reached criterion by the eighth pass through the training set were replaced. Four such subjects were replaced in the rule condition, none were replaced in either of the other two conditions.

The test set was identical for all three groups except that for the feature list group, the feature list materials were used. There were 40 items, split into a phase in which the only new items were positive matches and a subsequent phase in which the only new items were negative matches as in Experiment 1. The positive phase contained the four positive old items, the four positive matches, and four repetitions of each of the negative old items in random order. The four repetitions of the negative old items were performing as fillers, holding the positive old and positive match items apart in the series as well as increasing the ratio of old to new items. The negative phase contained three further repetitions of each of the positive old items shown in the positive phase, now acting as fillers, and the four negative match items. There were a total of 32 old and 8 new items in the test set as a whole. The negative phase immediately followed the positive phase with no break between the two. No feedback was given in the test phase.

Results

Analysis. The criterion alpha level was set to .05 for all analyses. The main focus is on the pattern of results for each individual condition rather than comparisons between conditions. Therefore a 2 X 2 within subjects ANOVA with test phase (positive phase vs. negative phase) and item type (old vs. match) as the factors was carried out for each condition. Because of problems with heterogeneity of variance all response times were subject to a log transformation prior to analysis. The data are displayed in Table 6.

Learning phase. The number of trials to criterion, that is the number of passes through the eight item training set, was 3.5 for the rule condition, 4.5 for the no rule condition, and 3.1 for the feature list condition. These differ according to a one way ANOVA F (2, 69) = 4.36, MSe = 2.97 and t-tests show that it is due to the no rule group taking longer than the rule group, $t(46) = 2.31$, and longer than the feature list group, $t(46) = 2.71$, but no difference between the rule and feature list groups $t(46) = 0.73$, n.s.. The trial of last error showed a similar pattern of results.

Feature list condition. A 2 X 2 ANOVA showed no effect of test phase ($F < 1$), no effect of item type ($F < 1$), and no interaction ($F(1,23) = 1.56$, MSe = 0.03, n.s.). Similar to the adults, there were relatively few errors and an equal number of errors across all types of items suggesting, that the children were using predominantly rule information to make their decisions. Thus these children could handle this type of rule quite well. Looking at their response times, we see that they took substantially longer than the adults, but that although they were reliably faster in the negative phase than in the positive phase $\underline{F}$ (1,23) = 6.70, $\underline{MSe}$ = 0.05 (probably a practice effect), they were not differentially slow on the match items ($\underline{F}$ < 1) and this was true for both the positive and negative phases (interaction $\underline{F}$ < 1).

No rule condition. Now we turn to the question of whether the children have a strong tendency to spontaneously use a similarity strategy, as suggested by previously discussed work (e.g. Kemler 1983a, b; Kemler-Nelson 1984; Kossan 1981). Once again the results from the children are comparable to those of the adults.For the error rates there is a large effect of test phase, $\underline{F}$ (1,23) = 37.8, $\underline{MSe}$ = 0.06; a large effect of item type, $\underline{F}$ (1,23) = 54.3, $\underline{MSe}$ = 0.06; and a substantial interaction between the two, $\underline{F}$ (1,23) = 25.7, $\underline{MSe}$ = 0.09. The children are making more errors on the match items than on the old items and this is especially true in the negative phase.

For the response times there was no effect of the test phase, $\underline{F}$ < 1: no effect of item type, $\underline{F}$ (1,11) = 2.37, $\underline{MSe}$ = 0.33; and a marginal interaction. $\underline{F}$ (1,11) = 3.76, $\underline{MSe}$ =0.32, .05 < $\underline{p}$ < .10. Thus, although the apparent increase in response times is not reliable overall, there appears to be a marginally larger increase for response times for the match items in the negative phase than in the positive phase (probably due to the old items getting faster, NOT the match items getting slower). So that, although the children appear to be slowing down more for the new items than the adults do, the results are for the most part, very similar

Table 6

Mean responses in Experiment 6

	Percent Errors			
	Positive Phase		Negative Phase	
Group	Old	Match	Old	Match
Feature Lists				
Children	10.4	4.2	4.2	6.3
Adults	4.2	4.2	6.3	7.6
No Rule				
Children	12.5	18.8	12.5	80.2
Adults	18	25	11	86
Rule				
Children	10.4	25.0	14.6	39.6
Adults	21	19	15	45

	Correct Response Times (ms)			
	Positive Phase		Negative Phase	
Group	Old	Match	Old	Match
Feature Lists				
Children	5056	4948	4344	4230
Adults	1750	1881	1752	1751
No Rule				
Children	2601	3704	2127	3837
Adults	1272	1342	1168	1120
Rule				
Children	2760	3270	1950	3770
Adults	1345	1385	1253	1612

to the results obtained from the adults. These children appear to be using a similarity strategy to make their decisions in the absence of a rule.

Rule condition. When the children are in a position to use both types of information, it appears that they balance the two in a manner similar to adults. For the error rates there was no effect of test phase, F (1,23) = 2.06, MSe =0.10 n.s.; a reliable effect of item type, F (1,23) = 13.3, MSe = 0.07; and no interaction, F (1,23) = 1.09, MSe = 0.06 n.s. Thus, although the children show more errors on the match items than on the old items, they do not make reliably more errors on the negative matches than on the positive matches. This diverges from the adult data primarily in that the children seem to make more errors on the positive matches than did the adults or indeed than was expected. The response times show a pattern more similar to the adult data with no effect of test phase, $F < 1$; a reliable effect of item type, F (1,18) = 14.3, MSe = 0.19; and a reliable interaction, F (1,18) = 7.14, MSe = 0.12. The children show a larger increase in response times for negative matches than for positive matches. For this group of children, although they are not reliably less correct on negative matches than on positive matches, it does take them longer to make a correct decision on a negative match item.

Discussion

As with the adults, the rules do not drop out very quickly as demonstrated by the few differences from adults evident in the children's data. The children are slower than the adults, they seem to have more trouble with positive matches, and they show less self correcting -- virtually no self correcting for children in contrast to virtually total self correcting for adults.

The children appear to be able to apply the rules with little trouble as shown by the feature list group. They also appear comfortable with a similarity basis for classification as shown by the no rule group. Additionally, somewhat surprisingly, they do not appear to solve the problem of balancing the two processes differently than do the adults, as shown by the rule group. They don't make reliably more errors on the negative matches as do the adults but they do take longer to make their decisions. So that, although they may be trading accuracy with response time slightly differently, the balance between rule and similarity based responding appears to be very similar. It is clear that the children's rule group does not look like the adult's no rule group as would be expected if the children had dropped the rule in favour of the item information in order to make their decisions.

There appears to be little difference between the performance of the adults and the performance of these children. However, it is possible that we have not moved to a young enough group of children in order to show a marked preference for similarity based information in this sort of rule-learning paradigm. Accordingly, an identical version of the experiment using only the rule condition was carried out with kindergarten children. Many of the kindergarten children had trouble reaching the criterion of one errorless run through the set of 8 training items but regardless of whether one examines the data from those who met or were close to the learning criterion or those who did not do as well, the results show the same intermediate level of performance on the negative match items suggesting that the kindergarten children were being affected by the negative matches but that they had not totally abandoned the rules. Overall, for these children the correct response rates were .69, .77, .73, and .52 for the positive old, positive match, negative old, and negative match items respectively ($\underline{F}(1,23)$ 3.57, $\underline{MSe}$ = 0.07, $\underline{p}$ =.07, for the match (good vs. bad) effect; $\underline{F}(1,23)$ = 1.1, $\underline{MSe}$ = 0.09, n.s., for the experience (old vs. match) effect; and $\underline{F}(1,23)$ = 5.10, $\underline{MSe}$ = 0.10, for the match X experience interaction). With these kindergarten children it is becoming apparent that we have pushed down to an age level where the children are having substantial trouble with the task. Even at this level, there is no evidence of complete reliance on similarity based strategies, either in the children who successfully completed training or those who failed to meet the criterion.

So even pushing down to an age level where the children were having substantial difficulty with the task did not push them into a more "wholistic" similarity based mode of responding. Thus, while it may be true that in a rule <u>induction</u> paradigm younger children will prefer a more overall similarity based style of responding, this does not appear to be the case in a rule learning paradigm where both episodic and rule based information are easily available.

One possible way for the rule and episodic information to combine in a task such as this would be in the form of a race between the two processes (see Logan 1988 for a race model of automatization). When a person is first learning the rule it would take longer to apply the rule than to recall a similar prior episode therefore the episodic basis for response would take precedence. As the person became more practiced with the rule their ability to apply the rule would increase in speed so the chances of a rule-based response would increase with practice. A race model would be consistent with the self-correcting seen with the adults. A fast episode-based response would be corrected when the slower, more reliable (for negative matches) rule-based process was completed.

There are two results that would argue against a race model. The results of Experiment 3, where instructions alerting the subjects to the negative matches caused substantial increases in response times, did not show an appreciable decrease in errors on negative match items. Similarly, with the children, who one might expect to take longer to apply the explicit rule, there was no change from the balance of episodic and rule based responding that was seen in the adults. If there was a race between the two processes happening one would expect that the children, especially the kindergarten children, would show more episodic responding. These results, although they do not rule out a race model, make it unlikely as an explanation of the results.

A more likely explanation is that the children's responses are becoming specialized to the concrete feature level -- they are looking for particular instantiations of the abstract features so that rather than looking for long legs in general the subject is looking for a particular set of long legs that are suggested by the rest of the available information (surrounding features, background, etc.). This type of rule specialization would be consistent with Kemler-Nelson's stress on the concrete, although it wouldn't be in the spirit of the term "wholistic" since it would still be true that the children would be attending to the individual features.

The striking difference between the adults and the children in the amount of self correcting suggests some interesting possibilities but, unfortunately, at this writing it is not possible to separate these from some less interesting ones. It is apparent that with the adults, once they have made their decision they continue checking it with the rule so that, in virtually every case, they correct their own errors. This self correcting was almost totally absent in the children. It may be that the children did not continue rule checking once they had made a response and thus did not notice their errors so that there may be a difference in children and adults in their use of rule information in a self checking fashion. Unfortunately there are other possible reasons for the lack of self correcting behaviour in these children. It may be that the children don't view the task as being embarrasingly easy as the adults do and thus are not embarrassed enough by an error to orally correct it even though they may be just as aware that they have made an error. Similarly, the children, like the adults, may be well aware that they have made an error but may not assume, like that adults, that the experimenter has noticed the error so they don't give it away by orally correcting themselves. Clearly more experimentation needs to be done to discern whether the children are as aware of their errors as are the adults but simply do not vocalize the fact or whether they in fact are less aware of their errors.

Chapter 6

GENERAL DISCUSSION

The experiments in this thesis have shown that practice with a sufficient and simple rule can lead to application that is specialized around particular exemplars. New items that were matched to old items in the same category were responded to approximately as rapidly as the old items themselves. New items that were negative matches, that is similar to old items that were in another category than that designated by the rule, produced slower reaction times and a much higher rate of error. By comparison to a group that did not have a rule, it is clear that knowledge of the rule helped to maintain accuracy, but it did not prevent facilitation or interference from instance and context-specific knowledge.

The phenomenon seems robust in the face of manipulations that might have forced subjects away from being influenced by past items. Removing additional individuating information in the form of stories (Experiment 2), changing the instructions to emphasize the importance of accuracy (Experiment 3), and decreasing the proportion of familiar items in the test series (Reanalysis), all failed to erase evidence of similarity based responding -- although emphasizing accuracy and decreasing the proportion of old items at test both decreased the size of the effects. The failure to avoid the detrimental effects of similarity on the negative matching items in the alert condition of Experiment 3 was despite paying a large premium in terms of response time.

It is important to stress that these results occurred under conditions that were selected to be of ecological interest, but not under conditions that were maximally biased to produce effects of prior episodes. To the extent that the conditions of the present experiments do not match those of interesting, real world situations, they tend to bias in the direction of more, rather than less rule use. First, the rule was the simplest possible additive rule; a fixed classification rule with any less than three binary features doesn't allow even an approximation to a graded structure. In all the experiments except Experiment 5, the rule was perfectly predictive, an unusual natural situation for a rule this simple. Further, the dimensions were binary with clearly discriminable values, characteristics which might have led to easier application of the rule than, for example, in many medical settings. Thus there was no necessity in these experiments for the subjects to use anything but a straightforward rule strategy in order to perform flawlessly.

Conditions for Producing a Specialization Effect

The impression is that the intrusion of negative matches is not a fragile phenomenon as long as familiarity remains predominantly useful within the general domain. The combination of old and positive matching items, items for which a retrieved answer would be useful, were in a strong majority in all experiments except Experiment 4 (neglecting the variations in background of Experiment 2). Given that the same answer was usually given by specific retrieval as by the rule, it apparently was hard not to relax into quickly accepting the answer that came from appearances. The distribution of familiar and unfamiliar items is probably also important when considering the potential effect of more extensive practice. The subjects in these experiments obviously had limited practice with the whole task. To make a convincing argument about automaticity, these experiments would have to have been continued for many more trials than they were. However, if additional practice had continued to favor a few old items, as is true with much of our commerce with natural objects, then one would expect, subject to testing, that specialization effects would continue to occur. Having said this, however, it is worth repeating that with the limited amount of practice in the current experiments, the rule did continue to be used, as evidenced by spontaneous error correction in virtually all adult subjects.

In addition to predominant familiarity, having a classification rule at least as complicated as an additive rule is probably important for producing specialization. A simple attributive rule based on a single, perceptually clear feature likely would not have given a specialization effect. However, simple additive rules of thumb seem to be prevalent in instructing adults in new visual categorizations. Medicine and biology contain a sufficient number of examples of simple additive rules for this not to be an overly restrictive condition.

Providing perceptually coherent, memorable visual units also seems to be important. In Experiments 2 and 5, similarity effects were seen when the items were presented to the subjects as drawings but were not seen when the same information was presented as written lists of features. This is consistent with the present emphasis on specific retrieval, because an effect of similar prior episodes requires that the material be processed as distinctive events in memory. However, under different processing circumstances verbal stimuli can be processed as distinct, memorable units. For example, when the subjects in Whittlesea's Experiment 6 (Whittlesea, 1987; as well as in Whittlesea & Cantwell, 1987) were encouraged to treat pseudowords as whole units, then a later test of the visual perception of new pseudowords showed striking similarity-to-old effects. When the initial experience with the pseudowords involved comparing them one letter at a time with a pseudoword that was prototypical for the training set, then the perception of new pseudowords was facilitated in proportion to their similarity to the prototype rather than similarity to the training instances. In the current experiments, the feature list stimuli were also presented in such a way as to break up the lists into separate features and minimize the amount of item-distinctive processing: The relevant features were always in the same place in the list, no unique irrelevant features were included, and no memory for unique information was required as had been done in Experiment 1. If the desire was to produce similarity effects with the feature list material, one would presumably have to change at least some of these processing characteristics.

Although it is at least useful to provide perceptually coherent whole items, it is probably important that there is not a predictive overall similarity among all of the members of the category; that is, the relevant features not be perceptually integrable independently of the irrelevant features, a stricture very much in the spirit of Garner's (1976) work on the sorting of integral and separable dimensions. One is left with the impression that since the relevant features of the current material do not seem to translate into something resembling an overall shape, one has to keep checking the features individually. This seems to be characteristic of many medical categories; the relevant, verbally specified features look different in combination with different irrelevant, or at least unspecified variables.

Taken together, the conditions that seem conducive to similarity effects are that testing take place in a context of largely familiar items, the rule be as complex as is often used for initial instruction, the items be easily integrated into mnemonic units, the relevant features not form a coherent perceptual whole independently of the irrelevant features, and possibly that extended experience bias toward repetition of old items. These joint conditions leave a domain of considerable interest.

Prior Processing Episodes

The effects of previous instances have been described as occurring because the items that were presented to the subjects act as retrieval cues for prior processing episodes involving those instances, as well as giving information for applying a rule. Using the term "prior processing episodes," in contrast with the term "instance" is meant to stress the processing dependent nature of both similarity and retrieval. In the current experiments there is a high degree of literal similarity between items that are described as similar. However, that is simply an operationally convenient way to define similarity and should not be taken to imply that a literal level of similarity in a positivist sense is necessary for the operation of the effects seen here. What makes a prior instance similar and available in the broader context is not necessarily its "objective" similarity, but the extent to which the items have been processed the same way.

Malt (1989) has a clear example of this where two different groups of subjects were given the same order of stimuli with slightly different tasks. In both cases there were old and new items and an old item was preceded by a similar new item. In the case where the processing required for the similar new item was the same as that required for the old item (is this category A or B?) there was a priming effect of the similar new item on the time taken to categorize the old item. When the categorization decision required was different for the new item (is this a large or small animal?) than for the following old item (is this category A or B?) there was no priming effect of the same physical stimulus. Thus it appears that it is the similarity of the item as processed that is important to categorization performance.

Rather than showing no effect of a literally similar stimulus if the processing didn't match, Vokey and Brooks (1990) have recently demonstrated the opposite. They demonstrated strong similarity to old effects in a grammaticality judgment task using artificial grammars despite changing all the individual features between training and test leaving only the relations between the individual features the same. That is, replacing all the individual letters in a letter string so that there was literally no similarity at the individual letter level did not interfere with the subject's increased willingness to call a string grammatical if it was logically similar to a training item. That result suggests that, although literal similarity was used in the present experiments it is probably not the only type of similarity that will lead to such episodic effects.

This approach is similar to research on episodic memory tasks which is characterized by very close attention to specifically what was done with a stimulus on a prior trial, the specific conditions of retrieval, and the way in which these two fit with one another; an approach referred to as encoding or processing specificity (Tulving & Thomson, 1973; Kolers, 1979). Such work demonstrates the importance of small variations in the setting and processing context of an item.

A processing episode approach has proven to be useful in a variety of areas that do not directly require memory for particular episodes. For example several of the experiments in this thesis were modelled after experiments by Jacoby and his colleagues on such topics as word identification (e.g. Jacoby 1983a, 1983b), judgments of fame (Jacoby, Kelley, Brown, & Jasechko, 1989; Jacoby, Woloshyn, & Kelley, 1989), and judgments of the loudness of background noise (Jacoby, Allan, Collins, & Larwill, 1988). Whittlesea has shown processing dependent episodic effects in perception of pseudowords and the word superiority effect (Whittlesea & Cantwell, 1987; Whittlesea & Brooks, 1988; Whittlesea, Brooks, & Westcott, 1990). Perception of line drawings (Jacoby, Baker, & Brooks, 1989) has also shown processing dependent retrieval effects. A general review is in Jacoby and Brooks (1984) and a more specific application to categorization tasks is in Brooks (1987). Overall, this work concurs with the research in the current paper in suggesting that episodic effects are a prominent part of many tasks in which they are not definitionally required.

The distinction between "prior processing episodes" and "instances" is not strongly tested in this thesis. The familiarity of the overall test list has been demonstrated to be important in controlling the size of episodic effects. But to have provided a strong test, one would have had to vary the processing of particular prior items and show that the same prior item had more of an effect when those conditions of processing were matched at the moment of test, similar to the work by Malt (1989) reported earlier. However, the "processing episode" terminology has been used because the approach has proven useful in related work and as a way of stressing that I am not assuming literal, "positivist" similarity when I refer to similarity effects.

Relation to Other Hybrid Models

This work is certainly not unique in suggesting that the operative form of knowledge is some form of hybrid between rule or prototype and instance or episodic knowledge (e.g. Medin, Dewey, & Murphy, 1983; Medin & Ross, 1989). This work can be viewed as part of an ongoing effort to embed this suggestion into ecological and process rationales. Investigating the hybrid question with the standard concept learning materials and procedures often leads to neither satisfying parallels to everyday problems nor to our experience when dealing with these problems. The episodic rationale just outlined and the ecological rationale of previous sections are one attempt to generate more satisfactory preparations. However, there are several other approaches in the current literature that have made interesting progress on the same general problem.

One approach to instantiating the hybrid suggestion is Ross's work on remindings in problem solving (Ross, 1984, 1987, 1989a, 1989b; Ross & Kennedy, 1990; Ross, Perkins, & Tenpenny, 1990). His work has shown that performance while learning a word processing program or learning to solve simple probability problems is influenced by similarity to previously encountered problems in the same domain. Ross (1987) demonstrated that while superficial similarity has an effect (e.g. whether the story line for the current and similar problem is the same), there was independent analytic competence (e.g. similarity of story lines had no effect if the correct formula was provided with the test problem), as well as a more analytic effect of similarity not tested in the current studies (the similar problem influenced the identification of the objects in the story with variables in the formulae). The parallel with the current results is interesting because of the very different level of challenge being provided for the learner. In Ross's work, understanding the problem and the algorithms is the issue. For those problems there is doubt whether the person will retrieve the correct formula and will identify the correct semantic objects with the variables in the formula. The purpose of his research is to explicate the processes by which the learner accomplishes these tasks, and similarity to specific prior problems is clearly implicated. In the current task, the analytic competence of the learner is hardly the issue. There is only one rule, it is an easy rule, and the identification of the item features with the elements of the rule is not in doubt. Rather, the issue here is whether the effect of practice with a well understood rule is to make its application faster or to also specialize it around previously experienced exemplars. It would seem that within-domain similarity is important for both levels of problem difficulty. I see this work as complementary to that of Ross. Both demonstrate a role for prior instances, but they have different rationales and apply to different circumstances.

In another approach, Logan (1988) published a series of experiments which address instance effects in automaticity. He uses tasks like basic arithmetic in which he sees two ways to get an answer: 1) a purposeful use of an algorithm, and 2) automatic instance based memory for one's previous answer to the same question. Performance is based on a race between the two processes. As practice continues, the base of instances increases, thereby increasing the probability that the fastest access of an instance is fast enough to beat the algorithm. Automatic performance is the extreme case of this race process where there are so many episodes in the knowledge base that the instance based process virtually always wins the race. Thus he equates automatic performance with instance based processes and purposeful performance with algorithm based processes. The results of at least some of the present experiments, that is the lack of any substantial difference in proportion of episodic based responses with substantially slowed response time in the Alert condition of Experiment 3 and with the children in Experiment 6, suggest that a straightforward race explanation is not likely in these situations (see the discussion of Experiment 6). However, one could still argue that the simultaneous effects of rules and prior episodes in the present studies could be looked at as a very early version of the process that Logan describes.

Further experiments with longer practice sets would be required to answer that argument.

Although Logan's views seem fundamentally compatible with the present experiments there are some complementary differences between this research and his. Conceptuallv, his work describes a pure instance model that does not have any of the coding specificity or retrieval variations discussed in the previous section. Methodologically, he treats instances as statistical entities: he does not attempt to keep track of the influence of individual items. While this does not allow the specificity of effects provided by the negative matches in these studies, he is also not tied to a small number of specific items. Although he has not yet directly addressed the problem of transfer to new instances, the manner in which he could extend his explanation seems clear, and could provide interesting convergence with the results of the present methodological approach.

Finally, I believe the studies in this paper are very much in the spirit of a recent chapter by Medin and Ross (1989). They argued for the specific nature of much abstract thought taking the view that induction is not necessarily autonomous but is a function of how examples are processed and in practice induction tends to be far more conservative than implied by abstractionist models. Although they are concentrating, to a large extent on induction, or the learning and application of difficult complex rules, the implied end point of the process may be the same as when the subject is coming from the direction of the present research. Medin and Ross argue that the process may never become as abstract as previous abstractionist model would imply. In the present studies the induction is already done for the subject but they seem to be moving in a direction of more specific processing than is necessary from the abstract information that they already have -- the process is becoming more specific than is necessitated by the demands of the task. It appears that we may be converging on a level of coexistence of abstracted and specific information that is the general state of people's knowledge, regardless of whether it was gained by induction on the part of the person or presented to the person as an abstract generalization.

Bibliography

Allen, S. W., Norman, G. R., & Brooks, L. R. (1990). Experimental studies of learning dermatological diagnosis: The impact of examples. Manuscript submitted for publication.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch and B. Lloyd (Eds.), Cognition and Categorization. Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.

Brooks, L.R., Allen, S.W. & Norman, G.R. (1990). Rule and Instance Based Inference in a Medical Diagnostic Task. Manuscript submitted for publication.

Bruner, J.S., Goodnow, J.J., & Austin, G.A., (1956). A Study of Thinking. New York: John Wiley.

Estes, W. K. (1986a). Array models for category learning. Cognitive Psychology, 18, 500-549.

Estes, W. K. (1986b). Memory storage and retrieval processes in category learning. Journal of Experimental Psychology: General, 115, 155-174.

Garner, W.R. (1976). Interaction of stimulus dimensions in concept and choice processes. Cognitive Psychology, 8, 98-123.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. Behavior Research Methods, Instruments, and Computers, 16, 96-101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple trace memory model. Psychological Review, 95, 528-551.

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

Jacoby, L. L. (1983a). Remembering the data: Analyzing interactive processes in reading. Journal of Verbal Learning and Verbal Behavior, 22, 485-508.

Jacoby, L. L. (1983b). Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 21-38.

Jacoby, L.L. (1990). Separating consciously controlled from automatic influences of memory: Effects of dividing attention at test. Manuscript submitted for publication.

Jacoby, L.L., Allan, L.G., Collins, J.C., & Larwill, L.K. (1988). Memory influences subjective experience: Noise judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 240-247.

Jacoby, L. L., Baker, J. G., & Brooks, L.R. (1989). The Priority of the Specific: Episodic Effects in Picture Identification. Journal of Experimental Psychology: Learning, Memory & Cognition, 15, 275-281.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception and concept learning. In G. Bower (Ed.), The Psychology of Learning and Motivation: Advances in Research and Theory, Volume 18. New York: Academic Press.

Jacoby, L.L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. Journal of Personality and Social Psychology, 56, 326-338.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. Journal of Experimental Psychology: General, 118, 115-125.

Kemler, D. (1983a). Holistic and analytic modes in perceptual and cognitive development. In TJ Tighe and B.E. Shepp, eds., Perception, cognition, and development: Interactional analyses. Hillsdale NJ: Erlbaum.

Kemler, D. (1983b). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. Journal of Experimental Child Psychology, 36, 365-379.

Kemler-Nelson, D. (1984). The effect of intention on what concepts are acquired. Journal of Verbal Learning and Verbal Behavior, 23, 734-759.

Kolers, P. A. (1979). A pattern-analyzing basis for recognition memory. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of Processing and Human Memory. Hillsdale, N.J.: Erlbaum.

Kolodner, J. (Ed.), (1988). Case-based reasoning: Proceedings of a workshop on case-based reasoning. San Mateo, CA: Morgan Kaufman Publishers.

Kossan, N.E. (1981). Developmental differences in concept acquisition strategies. Child Development, 52, 290-298.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Malt, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 539-555.

Markman, E. M., (1989). Categorization and naming in children: Problems of induction. Cambridge, MS:MIT Press.

Medin, D. L., Dewey, G.I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. Journal of Experimental Psychology: Learning, Memory & Cognition, 9, 607-625.

Medin, D. L. & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R. Sternberg (Ed.), Advances in the Psychology of Human Intelligence, 5, 189-223. New York: Academic Press.

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

Medin, D. L. & Smith, E. E. (1984). Concepts and concept formation. Annual Review of Psychology, 35, 113-138.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115, 39-57.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 282-304.

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

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

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

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

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

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

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

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

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. Cognitive Psychology, 16, 371-416.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 629-639.

Ross, B. H. (1989a). Distinguishing types of superficial similarities; Different effects on the access and use of earlier problems. Journal of Experimental Psychology: Learning, Memory & Cognition, 15, 456-468.

Ross, B. H. (1989b). Remindings in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp.438-469). Cambridge: Cambridge University Press.

Ross, B. H. & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 42-55.

Ross, B. H., Perkins, S. J., & Tenpenny, P. L. (in press). Reminding based category learning. Cognitive Psychology.

Shank, R. C. (1982). Dynamic memory: A theory of reminding and learning in computers and people. Cambridge: Cambridge University Press.

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

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

Vokey, J. R., & Brooks, L. R. (1990). The salience of item knowledge in learning artificial grammars. Journal of Experimental Psychology: Learning, Memory & Cognition, Manuscript submitted for publication.

Whittlesea, B.W.A. (1987). Preservation of specific experiences in the representation of general knowledge. Journal of Experimental Psychology: Learning, Memory & Cognition, 13, 3-17.

Whittlesea, B. W. A. & Brooks, L. R. (1988). Critical influence of particular experiences in the perception of letters, words, and phrases. Memory & Cognition, 19, 387-399.

Whittlesea, B. W. A., Brooks, L. R. & Westcott, C. (1990). Utilization of Analytic and Nonanalytic Processing: Factors Affecting Reliance on General and Particular Knowledge. Manuscript submitted for publication.

Whittlesea, B. W. A., & Cantwell, A. L. (1987). Enduring influence of the purpose of experiences: Encoding-retrieval interactions in word and pseudoword perception. Memory & Cognition, 15, 465-472.

Appendix

Logical Descriptions of Experimental Stimuli

Item Number	Body Shape	Spots	Leg Length	Neck Length	Number of Legs	Back-ground
1	1	1	1	0	0	4
2	0	1	1	0	1	1
3	1	0	1	1	1	2
4	1	1	0	1	0	3
5	1	0	0	0	1	2
6	0	1	0	1	1	1
7	0	0	1	1	0	4
8	0	0	0	0	0	3
9	1	1	1	1	1	2
10	1	0	1	0	0	4
11	1	1	0	0	1	2
12	0	1	1	1	0	4
13	0	1	0	0	0	3
14	0	0	1	0	1	1
15	0	0	0	1	1	1
16	1	0	0	1	0	3

The 1's and 0's represent the two different values on the dimension. Body shape, spots, and leg length are the dimensions that are relevant with respect to the rule while neck length and number of legs are irrelevant. Notice that each item numbered from 1 through 8 has a matching item among those numbered from 9 through 16 that is identical on all dimensions except spots. Any other item differs by at least two dimensions. The assignment of items to conditions for experiments 1, 3, 4, 5, & 6 was as follows (in experiment 2 only rule 1 was used):

Rule 1:

Positive Old - 1, 3, 6, 8 Positive Match - 10, 9, 15, 13

Negative Old - 2, 4, 5, 7 Negative Match - 14, 16, 11, 12

Rule 2:

Positive Old - 11, 12, 14, 16 Positive Match - 5, 7, 2, 4

Negative Old - 9, 10, 13, 15 Negative Match - 3, 1, 8, 6

Rule 3:

Positive Old - 9, 10, 13, 15 Positive Match - 3, 1, 8, 6

Negative Old - 11, 12, 14, 16 Negative Match - 5, 7, 2, 4

Rule 4:

Positive Old - 2, 4, 5, 7 Positive Match - 14, 16, 11, 12

Negative Old - 1, 3, 6, 8 Negative Match - 10, 9, 15, 13