INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A

SIMULTANEOUS AND DELAYED PHYSICAL IDENTITY MATCHING-TO-SAMPLE: ACQUISITION AND TRANSFER OF RESPONDING TO NOMINAL CATEGORIES OF PHOTOGRAPHIC STIMULI BY RHESUS MONKEYS (MACACA MULATTA)

BY

SHARON HIMMANEN

A dissertation submitted to the Graduate Faculty
in Psychology in partial fulfillment of the requirements
for the degree of Doctor of Philosophy,
The City University of New York

1999

UMI Number: 9946174

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy

9/13/99
Date

Chair of Examining Committee

9/13/99
Date

Executive Officer

Sheila Chase
Martin Chodorow
Robert L. Thompson
Roger K. R. Thompson

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Simultaneous and Delayed Physical Identity Matching-to-Sample: Acquisition and Transfer of Responding to Nominal Categories of Photographic Stimuli by Rhesus Monkeys (Macaca mulatta)

by

Sharon Himmanen

Advisor: Professor Karyl B. Swartz

Four monkeys were trained on an identity matching-to-sample task then transferred to novel stimuli under three delay conditions: simultaneous, single delay within a session, and multiple delays within a session to examine the effects of delay on acquisition and transfer of the identity matching. Five sets of digitized color photographs of natural objects, each divided into eight categories (amphibians, birds, fish, flowers, fruits and vegetables, insects, mammal faces and shells) were used. In Experiment 1 monkeys were trained on a simultaneous MTS task and performed significantly above chance when presented with novel stimuli. The delay between sample and comparison stimuli was gradually increased across sessions in Experiment 2, and the monkeys again were able to perform significantly above chance and no different from familiar stimuli when presented with novel stimuli. Finally, in Experiment 3, delays ranging from 8s to 80s in 8s increments were presented randomly from trial to trial. Though performance on novel stimuli was above chance, accuracy was significantly lower for novel stimuli than for familiar stimuli. Accuracy also decreased as the delay increased. Performance on trials with sample and distractor items from the same category did not differ significantly from performance on trials with the sample and distractor from different categories except during transfer to novel stimuli while the delay varied within a session. The monkeys learned an identity match when trained on photographic stimuli that transferred to novel stimuli without regard for category.

Acknowledgements

My mother, Penny Himmanen, as well as my sister Leena and my brother Eric, have been a source of unwavering support throughout my entire life, including my graduate work. Thank you for your love, encouragement and support over the years.

My father, Henry Himmanen, was an inspiration throughout my graduate career. Though he died before my dissertation project was completed, he is in my thoughts often.

My friends Sheryl Bottner, Alora Chistiakoff, Susan Garrett, Jill Gimesky, Kelly Gritten, Gretchen and Jeff Kopmanis, Annie Miranda and Karen Swanson have been tremendously supportive throughout my graduate career. I love you guys! Special thanks to Sheryl for reading the dissertation twice and providing feedback.

Phyllis Edwards is responsible for the animal care facility at Lehman College, where the monkeys used in this research project are housed. Her care and efforts contributed to this project in many ways, both small and large, and I would like to thank her as well.

I have been privileged to work with a number of students who aided me in data collection and analysis throughout the years. Thanks to Dena Sarauw, Melissa Bada, Diana Buchman, Monique Deputy, Avril Hall, Danette Jagla, and Ruthie Williams.

Sheila Chase provided moral support and encouragement from the beginning. It was through her that I was first introduced to the wonders of working with monkeys, and I shall always be grateful for that opportunity. Her contributions to this dissertation project were invaluable.

Martin Chodorow also provided many excellent suggestions, as well as being the best teacher it has ever been my privilege to learn from. My happiest memories of graduate coursework stem from Martin's classes, particularly Cognitive Psychology.

Robert Thompson was the program head of biopsychology when I began my tenure in this program, and was one of my first instructors as a graduate student. His perspectives on learning provided valuable insight into this research project.

Roger Thompson was a late addition to the dissertation committee, but I am grateful for his comments, questions, constructive criticisms, and generosity.

And finally, working with my advisor Karyl Swartz has been a tremendous privilege. She has provided guidance, support, freedom, encouragement, honesty, instruction, and challenge in every possible positive way. I have learned so much from her, and look forward to continuing to learn from her for many years to come.

TABLE OF CONTENTS

Abstract iv

Acknowledgements vi

List of Tables x

List of Illustrations xi

Introduction 1

- Open-ended concepts 2
- Abstract concepts 11
- The role of delay in MTS tasks 17

Experiment 1 19

- Method 19
  - Subjects 19
  - Apparatus 20
  - Stimuli 21
  - Procedure 21
  - Transfer test 23
- Results 23
  - Trials to criterion and percent correct 23
  - Individual category analysis 25
  - Response time 28
  - Transfer test 30
- Discussion 35

Experiment 2 39

- Method 42
  - Subjects 42
  - Apparatus 42
  - Stimuli 42
  - Procedure 43

- Results 44
  - Trials to criterion 44
  - Percent correct 44
  - Response time 47
  - Delay transition 50
  - Within and between category performance 50
  - Transfer test 53
- Discussion 53

Experiment 3 56

- Method 57
  - Subjects 57
  - Apparatus 57
  - Stimuli 57
  - Procedure 58
- Results 59
  - Acquisition 59
  - Percent correct 59
  - Within-and between-category performance 66
  - Transfer test 66
- Discussion 73

General discussion 75

Appendix A. Stimuli 82

Appendix B: Performance on individual items in Experiment 1 87

Appendix C: Performance on individual items in Experiment 2 97

Appendix D: Performance on individual items in Experiment 3 104

References 107

LIST OF TABLES

Page

Table 1. Stimulus set presentation order and experimental parameters for each subject for experiments 1, 2 and 3. Sets I, II and V were composed of four images from each category for a total of 32 images. Sets III and IV were composed of eight images from each category for a total of 64 images..22

Table 2. z-scores for each block for each category. AM=amphibians, BD=birds, FI=fish, FL=flowers, FV=fruits & vegetables, IN=insects, MA=mammal faces, SL=shells..27

Table 3. Evaluation of transfer tests on the basis of those categories that showed a significant difference between familiar and novel stimuli and above chance performance on novel items. MTS=simultaneous ((Experiment 1), DMTS=single-delay MTS (Experiment 2), and MDMTS=multiple delay within a session (Experiment 3).......................34

Table 4. Summary of post-hoc analysis on two-way ANOVA for delay and category with accuracy and response time as the dependant measures respectively..48

Table 5. Mean percent correct for each category for the delay transitions. The gaps indicate where the monkey acquired the new delay within the minimum number of sessions available. 51

Table 6. Summary of post-hoc analyses on the significant main effect of delay for each subject..67

LIST OF ILLUSTRATIONS

Page

Figure 1. Acquisition of a simultaneous MTS by four rhesus monkeys as measured by the number of completed trials (trials that ended in either a correct response or an error) to reach a criterion of 85% correct for two consecutive sessions. The dashed line indicates the minimum number of trials the monkeys could have taken to reach criterion, which was 128 24

Figure 2. Acquisition of a simultaneous MTS. Acquisition was divided into four blocks for each subject, with an equal number of sessions in each block. The number of sessions in each block differed from subject to subject as each subject required a different number of sessions to satisfy criterion. Each block represents the average of the percent correct for each session that fell within that particular block 26

Figure 3. Overall percent correct and correct median response time for each category for the final block of acquisition. The gray bars show overall accuracy on each category; the closed squares show the median response time for correct responses to samples from each category. Note the increase in correct median response time for categories on which the monkeys scored the lowest. There is a high degree of correlation between accuracy and response time 29

Figure 4. Overall percent correct for the last day of a stimulus set (familiar) and the first day of a new stimulus set (novel) for all three experiments. The dashed line indicates chance based on the number of trials in a session. There were no significant differences between performance on familiar and novel stimuli for a simultaneous MTS (Experiment 1) or a single delay MTS (Experiment 2). There was a significant difference between overall percent correct when the delay varied within a session (Experiment 3) though both scores are above chance 31

Figure 5. Percent correct for the first presentation of sample stimulus items during the transfer tests for simultaneous MTS (Experiment 1), single-delay MTS (Experiment 2), and multiple-delay MTS (Experiment 3). The delay condition of the transfer test is presented on the x-axis 32

Page

Figure 6. The results of the first transfer test broken down by category for Experiment 1. The light gray bars show accuracy on familiar stimuli and the dark gray bars show accuracy on novel stimuli. The upper line is chance based on the number of trials in a session. Chance on any given trial was 50%....................33

Figure 7. Overall completed trials to criterion for each delay for Experiment 2. Sebastian failed to reach criterion on the 20s delay. All four monkeys took a variable number of trials to reach criterion for each delay range, but only Ezra showed a systematic increase in the number of completed trials to criterion as the delay increased. All monkeys show a substantial increase at the 8s delay but this increase was not significant....................45

Figure 8. Overall percent correct for each category on each delay for Experiment 2. The error bars indicate the standard deviation of percents correct. The dotted line represents criterion. The overall scores for all the monkeys across delays was very high, and variability, as indicated by the error bars, was low....................46

Figure 9. The median response time for correct trials on each delay of Experiment 2. Response time to correct trials tended to increase as the delay increased....................49

Figure 10. Starting with Experiment 2, a proportion of the trials in a session had sample and distractor stimuli from the same category (designated as same trials). The gray bars represent overall percent correct for same trials. The dark gray bar represents accuracy on those trials in which the sample and distractor were from different categories. Data from three stimulus sets and two delay conditions are presented. There were no significant differences between same and different for the single delay condition, or the familiar 8s to 80s condition. Accuracy on same and different did differ significantly for the final set of novel stimuli presented under the variable delay condition....................52

Figure 11. The results of the transfer test for Experiment 2 broken down by category for each subject. There were no significant differences between familiar (light gray) and novel (dark gray) stimuli for any subject on any category and performance on novel stimuli was above chance for all categories except insects for Sebastian. The dashed line represents chance based on the number of trials (68%)....................54

Page

Figure 12. Overall percent correct by session for Experiment 3. The points to the left of the dashed vertical line indicate overall percent correct for each session while longer and shorter delays were systematically added to the session, until a range of 4s-40s in 4s increments was achieved. At that point, the delay range was switched to 8s-80s and the points to the right of the vertical dashed line indicate performance on that range. The same set of stimuli were used throughout this training........60

Figure 13. Performance on each category of stimuli as a function of delay for Ezra (0071). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel stimuli. The delay range presented here was 8s to 80s........61

Figure 14. Performance on each category of stimuli as a function of delay for Franklin (14ac). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel set of stimuli. The delay range presented here was 8s to 80s........62

Figure 15. Performance on each category of stimuli as a function of delay for Rutherford (301a). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel stimuli. The delay range presented here was 8s to 80s........63

Figure 16. Performance on each category of stimuli as a function of delay for Sebastian (0076). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel stimuli. The delay range presented here was 8s to 80s........64

Figure 17. Overall percent correct on each category of familiar stimuli in Experiment 3 without regard for delay........65

Figure 18. Overall percent correct as a function of delay for Experiment 3. The closed diamonds show performance on the familiar set of stimuli, and the open squares show performance on the novel set of stimuli........68

Figure 19. Overall percent correct for each session for the familiar (closed diamonds) and novel stimulus sets (open squares) for Experiment 3 on the 8s to 80s range of delays. Scores on familiar and novel stimuli were significantly different from each other for all subjects........70

Page

Figure 20. The results of the transfer test for Experiment 3 broken down by category. Light gray bars show percent correct for familiar stimuli and dark gray bars represent percent correct for novel stimuli. The upper dotted line represents chance based on 100 trials..........................72

Simultaneous and Delayed Physical Identity Matching-to-Sample: Acquisition and Transfer of Responding to Nominal Categories of Photographic Stimuli by Rhesus Monkeys (Macaca mulatta)

When an organism responds in a similar fashion to a group of objects that fall into a particular class that organism is said to have formed a concept (Keller and Schoenfeld, 1950). Conceptually mediated behavior in non-human primates can be defined further as an animal's ability to make similarity judgments between and within common categories (stimulus subsets that share a number of similar and distinguishing features) of objects (Thompson and Oden, in press). As in the human cognition concept literature, the terms "category" and "concept" have been used interchangeably when applied to concept learning in animals. In many ways these terms are synonyms since many researchers believe that an animal that can distinguish between categories of stimuli can be said to have formed a concept of that category. Presenting the organism with novel instances of the category and measuring performance then tests the generality of this concept (Estes, 1994). This is an example of what is often referred to as an "open-ended" category in that there are a potentially unlimited number of exemplars that share some constellation of features. Successful learning of this type of concept is dependent upon category learning which is often distinguished from more abstract, rule-based concepts such as same-different or identity (Herrnstein, 1990).

Brown, Brown, and Poulson (1997) proposed criteria for determining conceptual behavior in animals. First, all the transfer stimuli (stimuli introduced during a testing period to test category or concept formation) should be novel and should be sufficiently different from one another and the baseline stimuli so that discrimination confusion is unlikely. Second, transfer testing should be limited to the first presentation of each novel stimulus, so that the results will not be confounded by a history of reinforcement and subsequent learning. Third, transfer performance should be as good as baseline performance and both should be at a good performance level.

I. Open-ended concepts

The literature examining open-ended concepts in animals includes a variety of procedures and stimuli. Thompson (1995) provides a detailed overview of many aspects of this literature. The main focus of these studies involves teaching animals to discriminate between photographs that contain elements of a particular category from photographs that do not contain those elements. The negative stimuli that have been used include photographs that do not contain the item but are similar in all other respects to the positive stimuli as well as stimuli that feature other category items, such as other animals.

An early set of experiments on open-ended concept learning in animals demonstrated that pigeons (Columbia livia) quickly learned to distinguish between photographs containing humans and similar photographs containing no humans (Herrnstein and Loveland, 1964). Based on response levels and the types of errors Herrnstein concluded that the pigeons were responding to the task on the basis of the concept "person" that had been formed prior to the experiment. Since the pigeons had already had experience with humans, Herrnstein hypothesized that they were able to transfer this concept to the laboratory setting and readily learned to discriminate pictures on the basis of whether or not they contained humans.

Herrnstein continued this work (Herrnstein, Loveland, and Cable, 1976; Herrnstein, 1979) by teaching pigeons a "tree concept." In much the same way pigeons were taught the "person" concept, they were trained to distinguish between photographs containing trees and photographs containing no trees. Herrnstein again concluded that the concept or concepts mediating behavior in this particular set of studies were not shaped by contingencies within the experiment. He argued that the subjects entered the study with the "tree" concept already formed. First, there was rapid discrimination. Additionally, there was no association between ranking of positive stimuli during initial acquisition and history of reinforcement or cumulated time of presentation of individual stimuli.

In Herrnstein's studies on the "person" concept and the "tree" concept in pigeons, it was impossible to determine what aspects of the stimulus controlled the birds' responses. The birds may have been responding to trees or people as a single class or as the union of multiple subclasses. Similarly, the negative stimuli could have been discriminated on the basis of the absence of trees or people, or on the basis of multiple subclasses of their own. The studies do not reveal which of these may be the case.

Schrier, Angarella, and Povar (1984) addressed the issue of what the basis for discrimination is by teaching stumptail monkeys (Macaca arctoides) a human concept (experiment 1) and a monkey concept (experiment 2). They also used a discrete trials procedure that eliminated any possible problems related to reinforcement schedule, and a conditional spatial discrimination task that permitted a correct response on each trial.

The general procedure for the first experiment was as follows. During a session the monkey was presented with a series of photographs. Half the photographs contained humans while the other half contained no humans. There were two response keys placed in different locations. If the photograph contained humans, the subjects were to touch one of the response keys; if the photograph had no humans in it, they were to touch the other key. This was followed by a transfer test under the same conditions, but using novel stimuli. Experiment 2 was conducted the same way as experiment 1, except that half the photographs contained monkeys and half the photographs contained other animals.

All subjects learned to discriminate between photographs containing humans, and photographs containing no humans at fairly high levels of accuracy (85%-90%). Subsequent transfer to new stimuli was higher and faster than acquisition to the original set. Similarly, all subjects learned to discriminate between photographs containing monkeys and photographs containing other animals, as well as showing similar positive transfer to that found in the first experiment. Interestingly, despite the fact that one could argue that the discrimination task in experiment 2 was much more difficult due to the presence of other animals in the non-monkey slides as opposed to slides containing just scenery, etc., acquisition here was much faster than in the previous experiment. Because one subject had not been used in the previous experiment, it was assumed that this faster acquisition had something to do with the stimulus set rather than prior training. However, this rapid acquisition did not translate into higher performance on the transfer tests.

Noting that acquisition with monkeys was much slower and the positive transfer less impressive than that reported for pigeons, Schrier and Brady (1987) looked at the effects of set size on categorization of natural stimuli by monkeys. They suggested that during original learning the monkeys were gradually learning to respond on the basis of individual photographs rather than some common feature or set of features, thereby interfering with concept formation.

In Schrier and Brady's (1987) first experiment, monkeys were presented with a two-choice discrimination task rather than the single stimulus conditional spatial task used by Schrier, Angarella, and Povar (1984). In the two-choice discrimination task the monkeys were only rewarded for responding to images containing humans when presented with two photographs, one containing humans and the other containing no humans. In addition, a significantly larger set of human photographs was used. These photographs were classified by the experimenters into good, intermediate to good, intermediate to poor and poor photographs on the basis of the percentage of the photograph taken up by humans. Acquisition was rapid, and the percents correct were monotonically related to subcategory, with the highest percentages associated with photographs containing a high percentage of humans (good) and the lowest percentages associated with photographs containing the lowest percentage of humans (poor). Response times were also related to the subcategories, with the fastest response times associated with the photographs containing the highest percentage of humans.

The second experiment examined the properties of the images that might lead to categorization. A series of human and non-human probe trials were introduced in which the images were silhouetted, scrambled, contained a rightside-up human in an upside down scene, contained upside-down humans, and contained monkeys and apes. When paired with photographs of humans, unless the probe contained the image of a rightside-up human, the monkeys seldom chose it. These results suggest that with the exception of the rightside-up human in an upside down scene, none of the probe photographs were similar enough to the human photographs to cause the monkey to respond to them. Additionally, the monkeys maintained their high levels of performance found in experiment 1, even with the introduction of probe trials. Additionally, the monkeys in these experiments acquired the discrimination faster and responded with higher levels of accuracy than did the monkeys in Schrier, Angarella, and Povar (1984) experiments. The better performance by the monkeys in the later study was attributed to a larger stimulus set size that eliminated the possibility of learning the characteristics of individual slides. They rule out a number of other factors, including the species of animals used in the two studies (Macaca mulatta rather than Macaca arctoides) as well as differences in the procedures used for the two studies (two-choice discrimination rather than go/no-go). Finally, however, as was the case with Herrnstein's results, it is difficult to say what aspects of the stimuli were controlling responding. Responding may have been based on stimulus generalization to a well-defined physical dimension, based on the concept "human," or based on some factors between these two extremes.

D'Amato and Van Sant (1988) used Schrier and Brady's (1987) stimuli to examine the effects of the number of exemplars on the person concept in capuchin monkeys (Cebus apella). They attempted to limit the number of exemplars used in the training phase while still obtaining high levels of transfer to new instances of the target concept. In experiment 1 20 exemplars (10 with humans and 10 without) were used to train a person concept in monkeys using a go/no-go procedure. Acquisition occurred within 600 trials. Early transfer tests produced no significant transfer. However, after additional training with 30 person and 30 non-person photographs the monkeys showed significant transfer.

However, it is questionable whether the monkeys' behavior could be labeled conceptually mediated. Like Schrier and Brady (1987), they too found better performance on photographs that could be considered "good" or "fair." By examining errors to individual stimuli they discovered that the monkeys seemed to have some degree of initial difficulty with close-up, portrait photographs of humans. Additionally, a number of the person photographs contained patches of red, and the researchers theorized that this feature led to a number of errors made on one non-person photograph that also contained a patch of red. A more detailed analysis confirmed that more errors were made to non-person stimuli containing red patches than to those containing no red patches. D'Amato and Van Sant concluded that if a person concept was mediating responding in these monkeys, it was neither the sole nor the dominant source of control, as evidenced by the responding to red patches in certain stimuli regardless of the presence or absence of humans. The authors further suggested that rather than a higher-order "person concept," the monkeys may have identified a limited set of features (eyes, limbs, etc.) and used those as the basis for responding to the stimuli.

Roberts and Mazmanian (1988) examined learning in relation to the level of abstraction as defined by the breadth of the category in humans, monkeys and pigeons. The most concrete level of abstraction was a category of a single bird species , the common kingfisher (Alcedo atthis), with other birds as negative items. The next level of abstraction was a single class of animals, all birds, with other animals as the negative items. The highest level of abstraction was a category composed of all animals (including all negative images used in the previous two levels of abstraction), with photographs containing no animals as the negative items. Pigeons, monkeys and humans performed this task using a two-choice simultaneous discrimination learning task

Pigeons acquired the kingfisher/bird task significantly faster than the other two sets of stimuli. For monkeys, both the kingfisher/bird task and the animal/no animal task were learned at the same rate, and both were faster than the bird/animal rate of acquisition. Humans learned all three much faster than pigeons and monkeys, and their relative rate of learning was opposite that of pigeons. Probe tests showed that all three species acquired a kingfisher concept insofar as they were able to pick out novel pictures of kingfishers from other birds. While this agrees well with the previous pigeon data, the monkeys were equally efficient at acquiring the animal/non-animal concept, yet did not show the same level of transfer as with kingfisher/bird. These findings suggest that pigeons and monkeys may be able to form concepts at only the most concrete levels, and suggest that they were able to learn the kingfisher concept due to a relatively small number of salient features. Further experiments indicated that with further training monkeys and pigeons could acquire the animal/no animal concept, but that it was not specific to a particular species (Schrier, et al., 1994, show similar acquisition of a monkey/animal discrimination). This makes sense as there was no reason for the subjects to learn anything about the particular species used as negative instances in this task, other than to learn not to respond to them. These findings, in general suggest that monkeys and pigeons can learn very concrete and very broad concepts, but have difficulty with the intermediate concepts (birds from other animals, for example).

Schrier, et al. noted that learning and memorization of individual items was a significant factor in their experiment, particularly with respect to the transfer test. This may also have been a factor in Roberts and Mazmanian's results as well. Not only were the animals trained on the respective discriminations with the same pairs of stimuli presented in the same order every day, some of the same stimuli were used across all three discriminations.

As noted previously, it is difficult to determine what elements of the stimulus are controlling behavior in experiments on open-ended concepts. Roberts and Mazmanian suggested that the monkeys in their study learned a kingfisher concept in that they were able to discriminate kingfishers from other birds. However, it is possible that the monkeys were responding to certain physical properties common to all kingfishers that the non-reinforced slides of other birds did not contain. The real test of whether or not the monkeys learned a kingfisher concept would be to see if they could tell the difference between one kingfisher and another.

Finally, and most importantly, there is the issue of level of abstraction as in this particular set of experiments. The difficulty of determining what aspect of the stimulus controls behavior may be a factor in what constitutes concrete and abstract concepts from the animal's point of view. What may seem to be a discrimination done at a fairly abstract level from a human observer's perspective could be simple stimulus generalization. Additionally, it could also be the case that the most abstract concept to learn in Roberts and Mazmanian's study was the bird/animal discrimination because all positive and negative items in the discrimination belonged to the same class of stimuli, namely the class of animals. This may explain why that particular discrimination was learned more slowly and with less success than the other two discriminations.

Pigeons can also be taught to discriminate and classify multiple categories of stimuli (Bhatt, Wasserman, and Reynolds, 1988; Astley and Wasserman, 1992). Pigeons were trained on a go/no-go procedure using colored photographs of people, flowers, cars and chairs. They found that the birds grouped similar stimuli together, even when such grouping was unrelated to the prevailing reinforcement contingency. They suggest that stimuli from the same category may be represented along a common perceptual dimension, which they refer to as categorical cohesion. Unfortunately, the authors do not elaborate on the term, so it is difficult to determine precisely what they mean.

Wasserman, Kiedinger, and Bhatt (1988) also examined how pigeons learned to classify both familiar and novel pictures. In a forced-choice procedure that included training on discriminating items within a category, they found the birds were more likely to confuse items within a given category than between different categories. Further, in a second experiment, pigeons learned to classify items more quickly according to experimenter imposed or semantic categories, as opposed to arbitrary categories.

In summary, photographically represented open-ended discriminations can be learned and transferred along nominal category distinctions. The basis for learning may be learning of individual photographs, specific features, or some perceptual similarity among individual exemplars that fall into a nominal category. More precise distinctions such as kingfisher/bird may be easier to learn than more broadly defined distinctions, such as bird/animal possibly because the latter has a greater number of overlapping features. Though the exact mechanisms underlying acquisition and transfer of nominal categories by non-humans is not yet known, a growing body of evidence suggests that animals do learn them, and that it is based in some part on perceptual similarities. Whether or not an animal's ability to make these discriminations constitutes conceptually mediated behavior continues to be a subject of considerable debate. Most researchers agree that transfer to novel stimuli is an important test to determine whether or not an animal has formed a concept of a particular category. Recently, however (Thompson and Oden, in press), the need for within-category discriminations in such studies are necessary to rule out simple stimulus generalization as an explanation for behavior.

II. Abstract concepts

While it may be the case that animals make decisions in an open-ended concept task by using a rule (i.e. when learning a "person" concept respond to a picture if certain features common to humans appear in the stimulus) there are other, more abstract, rule-based concepts. These concepts also rely on features of the stimulus, but unlike open-ended concepts, decisions are not dependant upon the presence or absence of specific features, but on the degree to which stimuli share features. An example of this type of abstract concept is the identity concept, which is often measured by the same-different and matching-to-sample paradigms.

In the same-different concept task, animals learn to make the appropriate response when two or more stimuli are identical or similar to one another, and another response when two or more stimuli are different. While Premack (1983) pointed out that the true same-different concept refers to the ability to judge relations between relations (i.e. being able to say that pictures of two identical flowers are the same as pictures of two identical butterflies) same-different discriminations comprises a significant portion of the animal concept literature, and continues to be an area of active research.

Wright, Santiago, and Sands (1984) trained rhesus monkeys on a same/different task using six sets of 70 pairs of stimuli to a high degree of accuracy. Initial transfer was poor, possibly because the monkeys learned particular trial configurations rather than learning the same/different concept. Additional training with daily item changes in the training stimuli improved transfer performance greatly. They also found that once the monkeys learned the same/different concept they took less time to learn new sets of stimuli. Bhatt and Wright (1992) obtained similar acquisition and transfer of the same/different concept with monkeys using digitized color images of natural categories presented on a monitor fitted with a touch-sensitive screen.

In an earlier study, Sands, Lincoln, and Wright (1982) examined errors made by monkeys on a same-different task using natural category stimuli. Rhesus monkeys were able to learn to make same/different distinctions across a wide range of stimuli. Sands, Lincoln, and Wright also looked at confusion matrices to determine whether or not there was any categorization of similar items. They found that faces tended to cluster together in the same region of multi-dimensional space. Fruits and vegetables also tended to cluster together separately from the other items. However, there was a certain degree of uniformity in terms of stimulus composition and backgrounds. It was difficult to determine which aspects of the stimuli were controlling behavior without further manipulation.

The class of stimuli used for initial training of the same/different concept appears to be a limitation of the same/different concept. D'Amato and Colombo (1989) used very experienced monkeys and trained them to differentially respond to pairs of static stimuli that were either identical or different. All monkeys readily learned to perform this task. Normally, a transfer test consists of stimuli similar to, though discriminable from the stimuli used in training. In this case, however, the monkeys were presented with dynamic stimuli with which they had no experience. Flashing and steady green lights were used. and none of the monkeys were unable to transfer the same/different concept to dynamic stimuli. Control subjects were able to learn a simple discrimination between the steady and flashing light, so failure to discriminate the steady and flashing lights on the part of experimental subjects was not an explanation.

The same/different concept task has also been used to train non-identity matching. Neiworth and Wright (1994) used the same-different classification to train monkeys to judge objects with which they were familiar. Objects used in the study were further classified as objects the monkeys had had observed from a distance (a broom, a trashcan) and objects with which the monkeys had actively interacted (an apple, the transport box).

They found that the monkeys were able to successfully designate different photographs of the same object as “same.” Additionally, transfer was better to those objects with which the monkeys had had direct, physical experience.

The identity concept, usually measured using a matching-to-sample (MTS) procedure, is the ability to determine that two stimuli are physically identical to one another, and has been demonstrated with both auditory and visual stimuli in a variety of non-human species, including pigeons (Cummings and Berryman , 1961; Carter and Warner, 1978; D’Amato, Salmon, Loukas, and Tomie, 1986), rats (Nakagawa, 1993), monkeys (D’Amato, Salmon, and Colombo, 1985; Fujita, 1982), dolphins (Herman, Gory, Hovancik, and Bradshaw, 1989; Roitblat, Penner, and Nachtigal, 1990), sea lions (Kastak and Schusterman, 1994; Pack, Herman, and Roitblat, 1991), and chimpanzees (Oden, Thompson, and Premack, 1988; Tomonaga, 1993). It is believed that subjects in identity matching paradigms learn a matching rule. Depending on the species, this matching rule can be very specific to the stimulus configuration, or it can be a larger, conceptual rule to match two items that are identical.

In an early experiment Kojima (1979) trained monkeys to make identity matches to four geometric patterns. When presented with similar novel stimuli, formerly high performance rates dropped significantly. While it was possible to eliminate specific sample-match-distractor configurations as the source of learning, the authors concluded that familiarity with the sample stimuli played a significant role in the monkey’s ability to perform the task. The lack of successful transfer to novel stimuli strongly suggests that the monkeys did not learn a general matching rule.

However, Malone, Tolan, and Rogers (1980) were able to train rhesus monkeys to match photographs to real-world objects they could feel but not see. Haptic stimuli were widely varied, and included such items as door knobs, hinges, etc. Full-color photographs are of these items were used as photographic stimuli. Two monkeys learned to match three-dimensional objects to their photographs with accuracies greater than 80%. Accuracy decreased by only 5% for novel stimuli.

Attempting to determine the limits of the matching concept in monkeys, D'Amato, Salmon and Colombo (1985) used an extremely limited set of stimuli to train monkeys. In experiment 1, experimentally naïve capuchin monkeys (Cebus apella) were trained on a simultaneous MTS using three colored disks and eight black and white form stimuli. Each monkey was trained on a two-sample set of stimuli composed of two of the stimuli listed above. After approximately 24 sessions the monkeys were given a transfer test session in which half of the stimuli were familiar and half the stimuli were novel. Half of the monkeys failed to meet the transfer criterion of 70% correct or better on novel stimuli and were given additional training with the new stimuli until they met a criterion of 90% correct or better, at which time they were given another transfer test.

Monkeys who were trained with at least one color stimulus, particularly red, in the training set took fewer trials to reach criterion than those that were trained with black and white form stimuli. Of the monkeys who failed the first transfer test, three learned the set of novel stimuli faster than they had learned the original set of stimuli, and two were then able to pass a second transfer test. After further training on the third set of stimuli, the third monkey did extremely well on a third transfer test. Despite the differences in acquisition and the number of tests needed to show successful transfer, all the monkeys in this study were able to transfer the matching rule to novel stimuli after training with an extremely limited number of stimuli (between two and six stimulus items).

In a second experiment, the monkeys who had successfully transferred the matching rule to similar stimuli in the previous experiment failed to show transfer to steady versus flashing green light. D'Amato, Salmon, and Colombo (1985) note that their results suggest that monkeys acquire a stronger representation of the matching concept than pigeons do, but, like the same/different concept, it seems to be limited to the particular class of stimuli on which the monkeys were initially trained. Several attempts by the authors, as well as other researchers, to train auditory matching in monkeys who had had prior visual identity matching training also support these results.

A number of experiments suggest that monkeys are capable of learning more abstract concepts that they can apply to a variety of stimuli, such as matching and same-different. Animals are able to compare two stimuli and determine whether or not they are physically identical to one another. However, matching in these cases is limited to the class of stimuli used for initial training, in that monkeys may be incapable of transferring the matching rule to a seemingly higher level of abstraction by crossing stimulus classes. In addition, there are a number of distinctions between same-different procedures and MTS, both in terms of the materials used, and how its theoretical application. Photographic stimuli have been used to a much greater extent in same/different procedures than for matching. The same-different concept has also been used to investigate how animals might categorize and represent groups of similar items. The use of the MTS procedure in animal research has moved into the area of memory by employing a delay between the sample and the choice.

III. The role of delay in MTS tasks

In a delayed MTS (DMTS) procedure subjects are presented with a sample stimulus followed by a retention interval. During the retention interval no stimuli are present and no responses relevant to the task are possible. The subject must simply remember the sample stimulus during this period of time and then make the appropriate response when presented with a choice of stimuli. This procedure is useful for studying a number of cognitive phenomena, especially the parameters of memory, as well as various procedural factors that contribute to retention of memory across delay intervals. Of particular interest to the research here are experiments looking at the effects of stimulus similarity in DMTS.

Rhesus monkeys were tested using a DMTS procedure and ten two-dimensional stimuli, half of which were colored disks and half of which were black and white shapes (Devine, Jones, Neville, and Sakai, 1977). All possible combinations of color and form stimuli were used as samples and distractors and combined with all possible delays (which ranged from 0.03s to 32s) as well as several sample stimulus durations, and all were presented randomly across trials. The results showed no significant effect of delay or sample duration, but, more importantly, the combination of stimulus types did have a significant effect. The monkeys performed poorest when the sample, match and distractor were all shapes than when the sample, match and distractor were all colors or were a combination of colors and shapes. These were replicated in a second experiment that held either the sample duration or the delay constant across trials.

Similarly, in their first experiment, Devine, Burke, and Rohack (1979) used a DMTS task to examine the effects of stimulus similarity on visual short term memory in rhesus monkeys for a small number of two-dimensional stimuli (again, white shapes on a black background, and colors). Delays ranged from 0.03s to 32s. Again, the monkeys made more errors when the sample and distractor stimuli were similar to one another (either both shapes or both colors). They made fewer errors when the sample and distractor were different (a color and a shape). Experimental experience may have also played a role with respect to the effects of delay since one subject's performance was affected by delay while a more experienced subject did not show this effect to the same degree.

In summary, pigeons and monkeys can learn to match stimuli without the sample stimulus present. In addition, they can do this after some period of time between the sample and the choice has passed.. Within this procedure, however, evidence suggests that stimulus discriminability plays a significant role, and that monkeys are likely to make more errors when the distractor is from the same category of stimuli as the sample..

As noted above, most research done using the MTS procedure has used colors and/or forms as stimuli. The thinking behind the use of such stimuli is obvious: complex stimuli, such as photographs, often difficult to describe in terms of a single dimension, unlike the stimuli used in the described experiments. Colors and forms often vary along a single dimension, or along two easily manipulated dimensions when combined. While this is certainly not true for photographs, photographs do relate more closely to real-world objects than colors and forms, and as such may be easier to discriminate and remember than colors and forms (Fetterman, 1996).

The present research combines specific photographic stimuli that can be classified into nominal categories with memory requirements to investigate the effects of delay on the identity concept using multi-dimensional stimuli. Three delay conditions, simultaneous MTS, single delay MTS (DMTS), and multiple delay MTS (MDMTS) were used to assess how well matching can be sustained when the demands placed on memory are increased.

Experiment 1

In this experiment, monkeys were trained on a simultaneous MTS task to determine how quickly they could learn this task with a larger, more complex stimulus set than had been used previously. This experiment also assessed transfer to a similar, but novel set of stimuli.

Method

Subjects. Four adult male rhesus monkeys (Macaca mulatta) were subjects. Two of the monkeys, Franklin and Rutherford, were previously trained on a series of serial learning experiments (Swartz, Chen, and Terrace, 1991; Chen, Swartz, and Terrace, 1997). Ezra and Sebastian previously learned a same-different task involving several sets of natural category stimuli.

All the monkeys had ad lib access to water and were maintained on a diet of standard monkey chow (Purina Monkey Chow), whole wheat bread and fruit in sufficient quantities to maintain a healthy body weight. All were laboratory born and were housed in the Lehman College animal care facility in accordance with NIH and USDA guidelines.

Apparatus. A training and testing apparatus similar to that used by Swartz, Chen and Terrace (1991) was used to train the subjects on this procedure. The monkey was transferred from the home cage to a Plexiglas and steel chamber (53.82 cm high x 48.90 cm wide x 53.34 cm deep) for experimental sessions. A 27.94 cm by 25.40 cm square was cut out of one side of the chamber for the placement of a Sony GVM1310 high-resolution RGB monitor measuring 33.02 cm along the diagonal. Fitted to the surface of the monitor screen was a MicroTouch touch sensitive screen on which the subjects' responses were detected. A transparent Plexiglas template with nine square 4.40 cm by 3.20 cm holes arranged in a three by three matrix was placed between the front wall of the chamber and the touch screen. The template allowed the monkey access to the items on the display while preventing "swiping errors."

Both the monitor and the Plexiglas chamber were located inside an 86.36 cm high by 87.00 cm wide by 73.66 cm deep soundproof chamber (Model BRS/LVE). Lights and speakers for a white noise generator and feedback tones were located inside the soundproof chamber. Two Gerbrands 5120 pellet dispensers delivered Noyes 190mg banana-flavored pellets to food cups located at floor level on either side of the monitor and touch screen. A closed-circuit video camera was attached to the door of the soundproof chamber to allow viewing and videotaping via an outside monitor. A 486 33mH computer controlled and recorded data for all experimental sessions.

Stimuli. Four stimulus sets composed of color images of natural categories (16-bit, 5.33 cm wide by 4.24 cm high) digitized from periodicals, personal photographs, and videotapes were used. The categories were birds, fish, flowers, frogs, fruits and vegetables, insects, mammal faces (non-primate), and shells. In addition, the background for each image was digitally replaced with a neutral gray color (red=148, green=140, blue=148), leaving only the image of the specific category item(s). The stimuli for all sets used in all three experiments are presented in Appendix A.

Procedure. A simultaneous MTS procedure was used, with trials consisting of three parts: a start stimulus (a purple triangle on a black background), a sample stimulus, and two choice stimuli (choice phase). Stimulus sets and experimental parameters for all three experiments are outlined in Table 1.

Touches to the start stimulus, the sample stimulus, and the correct comparison stimulus (the match) produced a feedback tone; touches to the incorrect comparison stimulus (the distractor) produced a different "error" tone. When first touched, a red border appeared around the sample and the comparison stimulus for 100ms. Repeated touches to the sample stimulus produced no additional feedback. Failure to touch a stimulus within the allotted time during any phase of the trial resulted in a timeout (the screen was cleared and the house lights extinguished), as did touching the incorrect stimulus during the choice phase of the trial. A new trial began after the timeout. Correct responses during the choice phase of the trial were rewarded with a 190mg Noyes banana-flavored pellet delivered randomly to one of two food cups located on either side of the monitor. The screen was then cleared, and the house lights remained on until the

	Sets			Durations				
Subject	Exp.	Training	Transfer	ITI	Start	Sample	Choice	Timeout
Ezra (0071)	1	II	IV	5s	5s	5s	20s	20s
	2	IV	III	5s	5s	5s	20s	20s
	3	III	V	5s	5s	5s	20s	20s
Franklin (14ac)	1	I	III	9s	10s	10s	20s	10s
	2	III	IV	9s	10s	10s	20s	10s->15s
	3	IV	V	9s	10s	10s	20s	15s
Rutherford (301a)	1	I	IV	5s	5s	5s	20s	20s
	2	IV	III	5s	5s	5s	20s	20s
	3	III	V	5s	5s	5s	20s	20s
Sebastian (0076)	1	II	IV	5s	5s	5s	20s	20s
	2	IV	III	5s	5s	5s	20s	20s
	3	III	V	5s	5s	5s	20s	20s

Table 1. Stimulus set presentation order and experimental parameters for each subject for experiments 1, 2 and 3. Sets I, II and V were composed of four images from each category for a total of 32 images. Sets III and IV were composed of eight images from each category for a total of 64 images.

next trial. A flash of the house lights signaled a new trial two seconds prior to the appearance of the start stimulus.

There were 64 trials per session. Stimulus presentation for all phases of training was constructed so that eight presentations of a particular category appeared as the sample randomly throughout the session. Samples and distractors were all drawn from the same pool of stimuli such that on any given trial any image in the set could appear as either a sample-match, or a distractor, with the restriction that a distractor was never presented with a sample from the same category.

Transfer test. The criterion for completion of training on the simultaneous MTS was 85% correct or better for two consecutive sessions. The monkeys were then transferred to one of the two sets of 64 novel stimuli. The transfer test consisted of sessions that were identical to all previous sessions, except that novel stimuli were presented.

Results

Trials to criterion and percent correct. Figure 1 presents the completed trials to criterion for each subject. Three subjects, Ezra (0071), Franklin (14ac) and Sebastian (0076) all acquired the simultaneous MTS within 3,000 trials. Rutherford (301a) needed more training to acquire the simultaneous matching to sample than the others.

Acquisition was divided into four blocks for each subject, with each block comprised of an equal number of sessions. The number of sessions in each block varied from individual to individual since each monkey took a different number of sessions to reach criterion. Each subject's acquisition and performance was considered individually

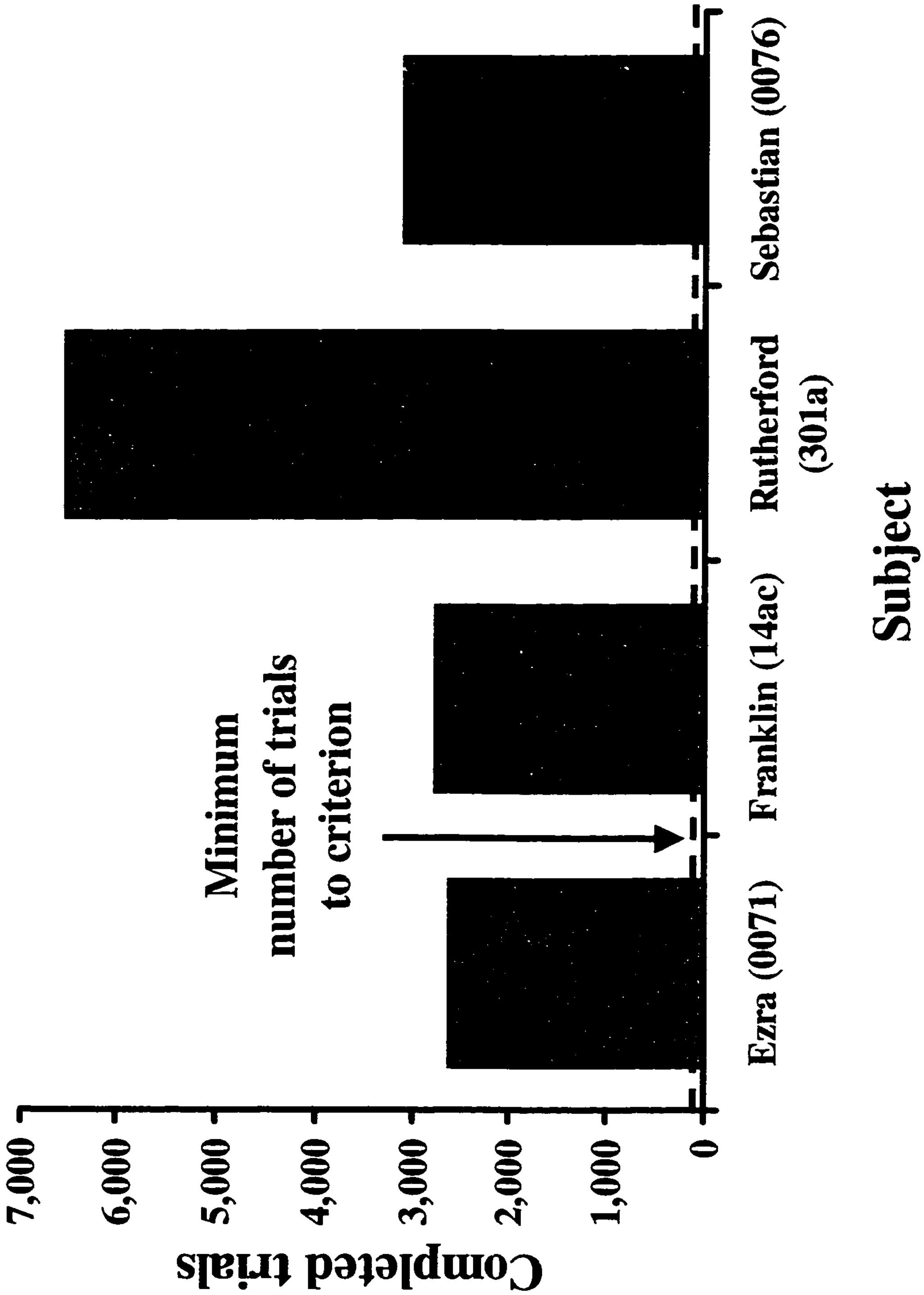


Figure 1. Acquisition of a simultaneous MTS by four rhesus monkeys as measured by the number of completed trials (trials that ended in either a correct response or an error) to reach a criterion of 85% correct for two consecutive sessions. The dashed line indicates the minimum number of trials the monkeys could have taken to reach criterion, which was 128.

in terms of statistical analysis, unless noted otherwise, and data transformations were performed to stabilize variances in the case of proportion correct and for positively skewed data in the case of response time (Winer, 1962). As can be seen in Figure 2, which shows the overall percent correct for each block, all four subjects initially performed at chance (50%), with accuracy gradually increasing until the criterion of 85% correct for two consecutive sessions was met. A one-way ANOVA on the arcsine-transformed proportion correct for each block showed a highly significant effect of training block for all four monkeys [Ezra - $\underline{F}(3,42)=29.21$, $\underline{p}<0.0001$; Franklin – $(3,46)=13.40$, $\underline{p}<0.0001$; Rutherford – $\underline{F}(3,101)=12.72$, $\underline{p}<0.0001$; Sebastian – $\underline{F}(3,48)=44.107$, $\underline{p}<0.0001$]. Tukey's HSD post hoc analyses showed that accuracy on the fourth block was significantly different from accuracy on the first and second blocks for all four monkeys.

Individual category analyses. During the first block of training, each subject performed significantly above chance on at least one category. The z-scores for each category for each block were calculated to determine if the monkeys were responding at above chance levels to the stimuli within a particular category across acquisition. These data are presented in Table 2. By the fourth block performance on all categories was significantly above chance for all subjects except Franklin's performance on birds, which remained at chance.

Within the final block some significant differences between performance on various categories for each subject became apparent. A one-way ANOVA on the arcsine-transformed proportion correct for each category for the final block of training revealed a

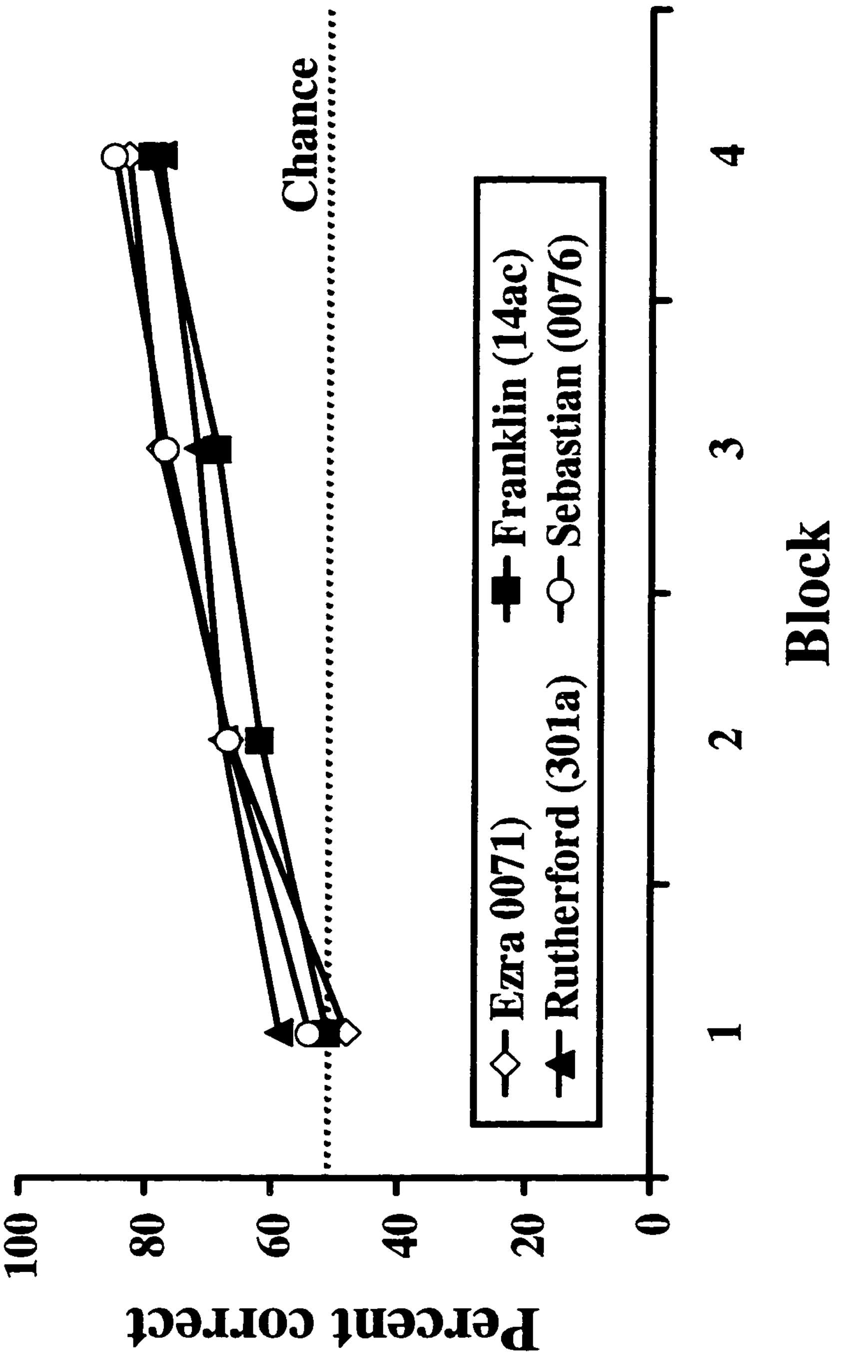


Figure 2. Acquisition of a simultaneous MTS. Acquisition was divided into four blocks for each subject, with an equal number of sessions in each block. The number of sessions in each block differed from subject to subject as each subject required a different number of sessions to satisfy criterion. Each block represents the average of the percent correct for each session that fell within that particular block.

		Block			
Subject	Category	1	2	3	4
Ezra (0071)	AM	-1.01	2.91	4.36	5.48
	BD	0.35	2.01	4.00	5.13
	FI	1.59	2.01	6.33	6.81
	FL	2.59	4.70	4.23	3.62
	FV	-1.56	5.59	6.46	6.86
	IN	0.45	-0.67	3.67	5.84
	MA	0.11	2.68	6.97	6.40
	SL	0.23	5.59	7.42	8.71
Franklin (14ac)	AM	1.22	2.13	2.56	5.31
	BD	-2.86	-2.56	-2.13	-0.20
	FI	1.43	4.69	3.20	5.72
	FL	-1.84	4.69	3.20	5.72
	FV	-3.47	3.84	0.85	5.31
	IN	1.63	3.50	3.62	4.90
	MA	-1.84	1.43	-2.23	4.21
	SL	-0.20	4.05	4.48	5.72
Rutherford (301a)	AM	4.71	5.23	6.84	9.85
	BD	-0.55	2.26	3.82	4.30
	FI	2.08	3.82	6.59	9.43
	FL	2.63	7.50	8.34	10.12
	FV	4.85	7.35	7.35	9.29
	IN	-0.14	3.25	5.17	7.90
	MA	3.47	4.38	2.55	3.47
	SL	2.63	6.65	6.98	8.04
Sebastian (0076)	AM	0.78	2.86	6.05	7.53
	BD	-2.35	-1.02	0.20	2.89
	FI	0.89	3.27	6.53	9.29
	FL	3.68	3.27	6.53	7.55
	FV	-0.10	5.31	7.14	8.49
	IN	-0.20	0.41	0.30	2.38
	MA	2.86	3.88	7.35	8.11
	SL	3.14	8.98	9.39	8.91

Table 2. z-scores for each block for each category. AM=amphibians, BD=birds, FI=fish, FL=flowers, FV=fruits and vegetables, IN=insects, MA=mammal faces, SL=shells.

significant main effect of category for each subject [Ezra – $F(7.87)=2.87$, $p<0.01$; Franklin – $F(7,95)=2.70$, $p<0.014$; Rutherford – $F(7,207)=6.55$, $p<0.0001$; Sebastian – $F(7,94)=15.01$, $p<0.0001$]. Tukey's post hoc analysis showed that each monkey performed significantly better or worse on at least one category compared to the others, but as Figure 3 shows, there was nothing consistent across subjects. Scores on birds was low for three subjects (Franklin, Rutherford and Sebastian). In addition, Rutherford and Sebastian performed significantly different on mammal faces and insects respectively in the final block of training. Ezra performed significantly different on shells in the final block.

Response time. A one-way repeated measures ANOVA on correct natural log-transformed median response times for sessions within the last block of training for each category showed a significant main effect of category at the for two of the monkeys [Rutherford – $F(7,207)=2.83$, $p<0.008$; Sebastian – $F(7,95)=3.24$, $p<0.004$]. As was the case with percent correct, these two monkeys performed differently on at least one category compared to one or two of the others. Response time data for each category are also presented in Figure 3. Rutherford's response time to mammal faces was significantly different from his response time to fish and flowers, and Sebastian's response time to insects was significantly different from his response time to fruits and vegetables (FV) and sea life (SL). As Figure 3 indicates, all monkeys showed a tendency to take longer to respond to those categories on which their scores were lower than to other categories, though this did not always result in significant differences across

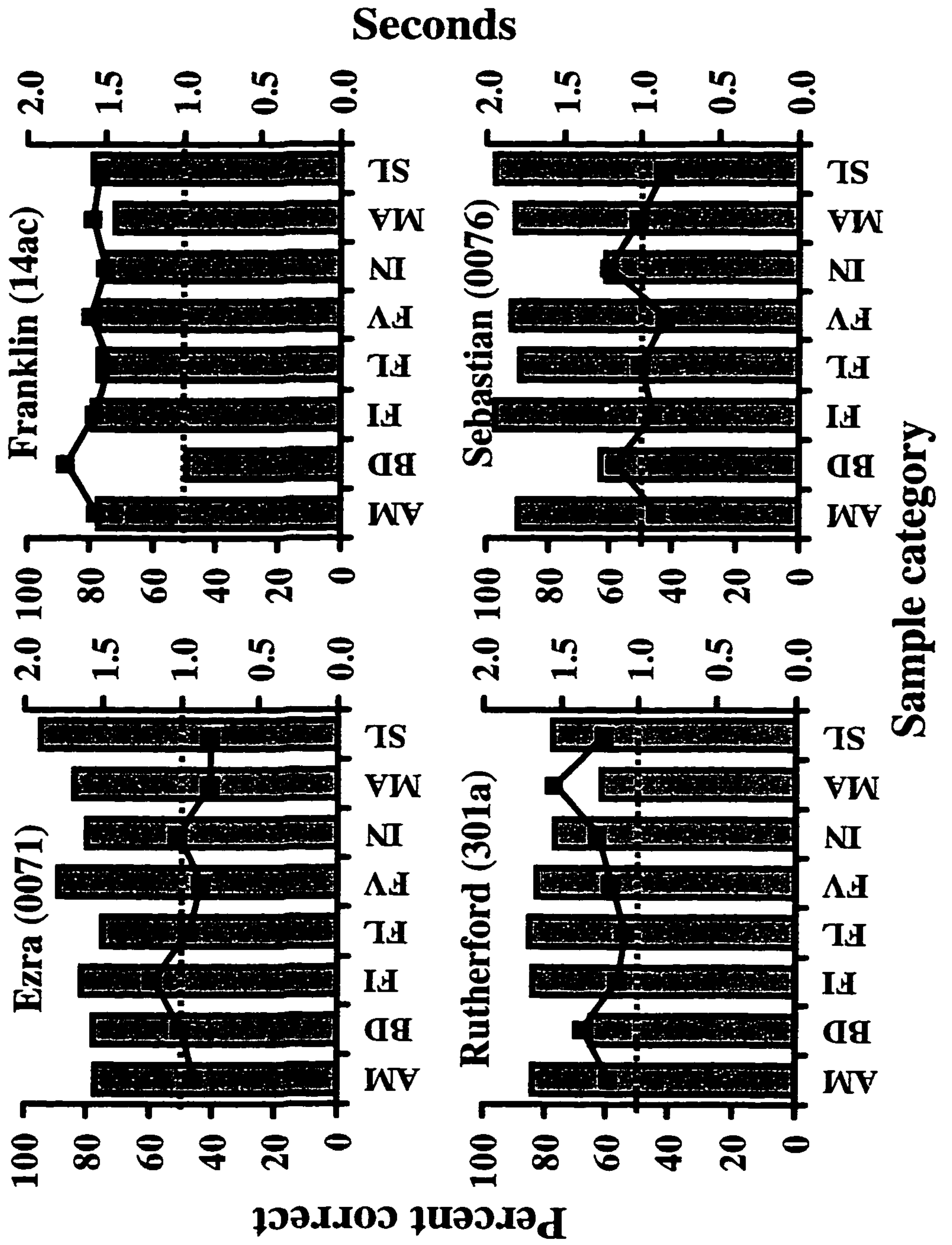


Figure 3. Overall percent correct and correct median response time for each category in the final block of acquisition. The gray bars show overall accuracy on each category; the closed squares show the median response time for correct responses to samples from each category. The dashed line is chance (50%). Note the increase in correct median response time for categories on which the monkeys scored the lowest. There is a high degree of correlation between accuracy and response time.

categories. A Spearman rank-order correlation on category percent correct and correct response times also yielded high negative correlations for each subject [Ezra – r_s=-0.49; Franklin – r_s=-0.43; Rutherford – r_s=-0.91; Sebastian – r_s=-0.62].

Transfer test. The overall percent correct for the last session of familiar stimuli and the first session of novel stimuli are presented in Figure 4 (left-most set of columns on each graph). Though all scores on the novel stimuli were above chance for 64 trials (68%) a paired t-test showed that percent correct scores for novel stimuli were not significantly different from familiar, $t(3)=2.88$. The left-most column on each graph of Figure 5 shows accuracy on the sample stimuli for those trials on which they appeared for the first time on the transfer test. All four monkeys are well above chance on these trials for this transfer test.

Transfer test data for each category are presented in Figure 6. Two-way 2 X 8 ANOVAs on the arcsine-transformed proportion correct for stimulus set (training vs. novel) and category for each subject revealed a significant main effect of sample category for three monkeys [Ezra – $F(7,64)=3.31$, $p<0.005$; Rutherford – $F(7,64)=3.97$, $p<0.001$; Sebastian - $F(7.64)=2.31$, $p<0.001$]. Additionally, Sebastian had a significant main effect of set (familiar vs. novel), $F(1,64)=55.75$, $p<0.0001$). Finally, Franklin had a significant interaction between set type and category, $F(7.64)=2.27$, $p<0.037$. A simple effects analysis indicated that his scores on familiar and novel amphibians and fish were significantly different from one another.

The left most columns beneath each subheading in Table 3 summarize the results of the transfer test for this experiment. By evaluating above and below chance performance

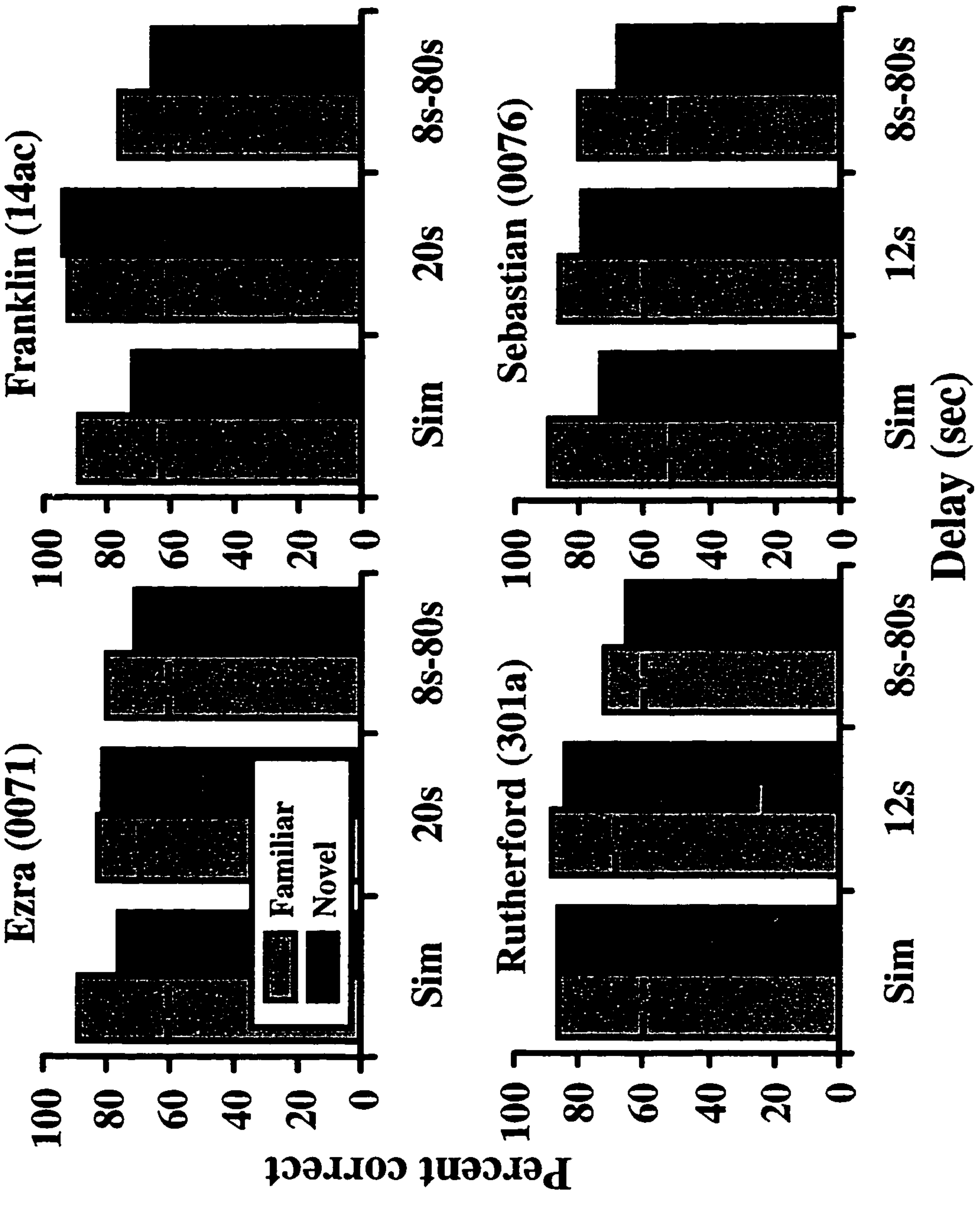


Figure 4. Overall percent correct for the last day of a stimulus set (familiar) and the first day of a new stimulus set (novel) for all three experiment. The dashed line indicates chance based on the number of trials in a session. There were no significant differences between performance on familiar and novel stimuli for a simultaneous MTS (Experiment 1) or a single delay MTS (Experiment 2). There was a significant difference between overall percent correct when the delay varied within a session (Experiment 3) though both scores are above chance.

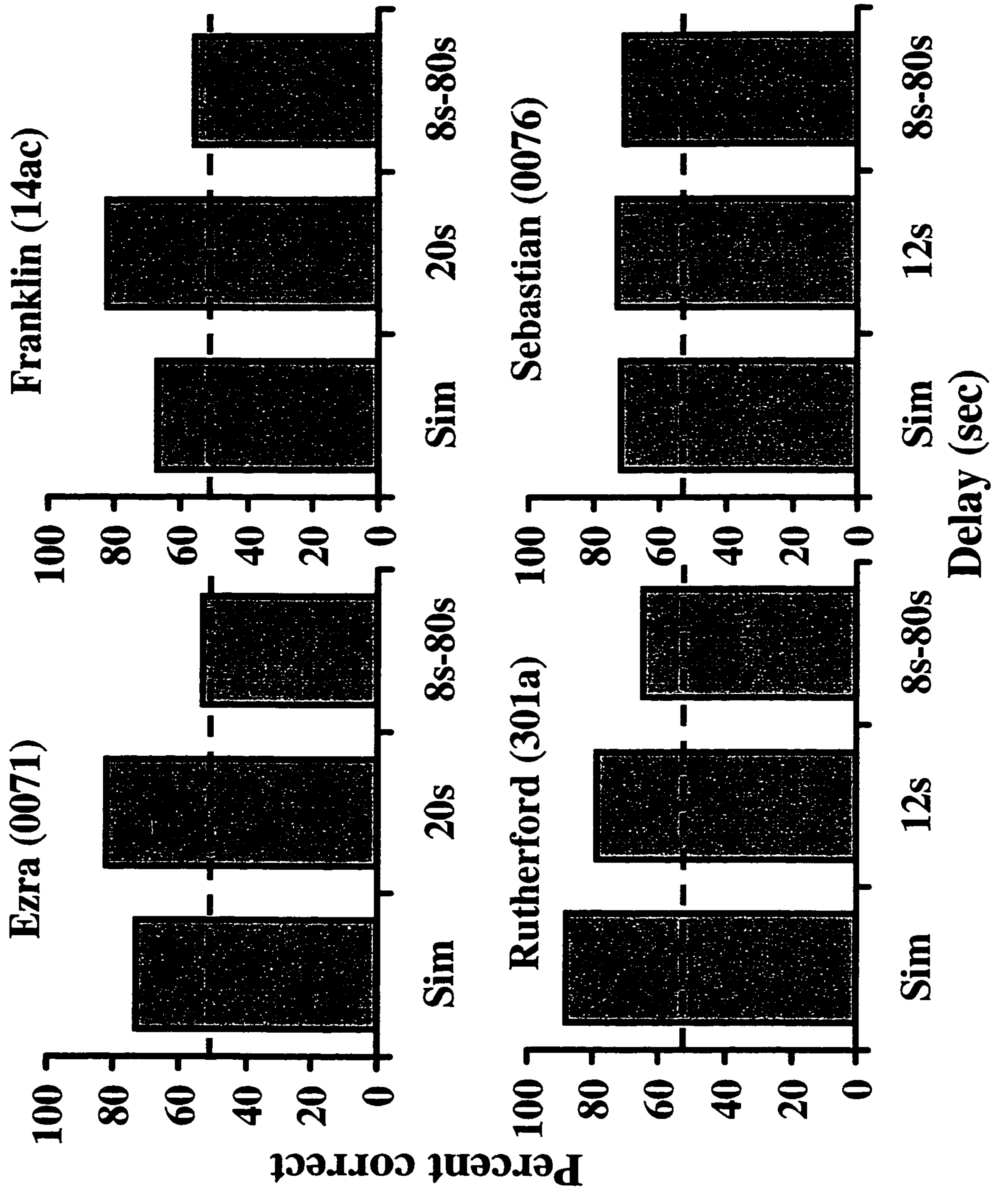


Figure 5. Percent correct for the first presentation of sample stimulus items during the transfer tests for simultaneous MTS (Experiment 1), single-delay MTS (Experiment 2), and multiple-delay MTS (Experiment 3). The delay condition of the transfer test is presented on the x-axis.

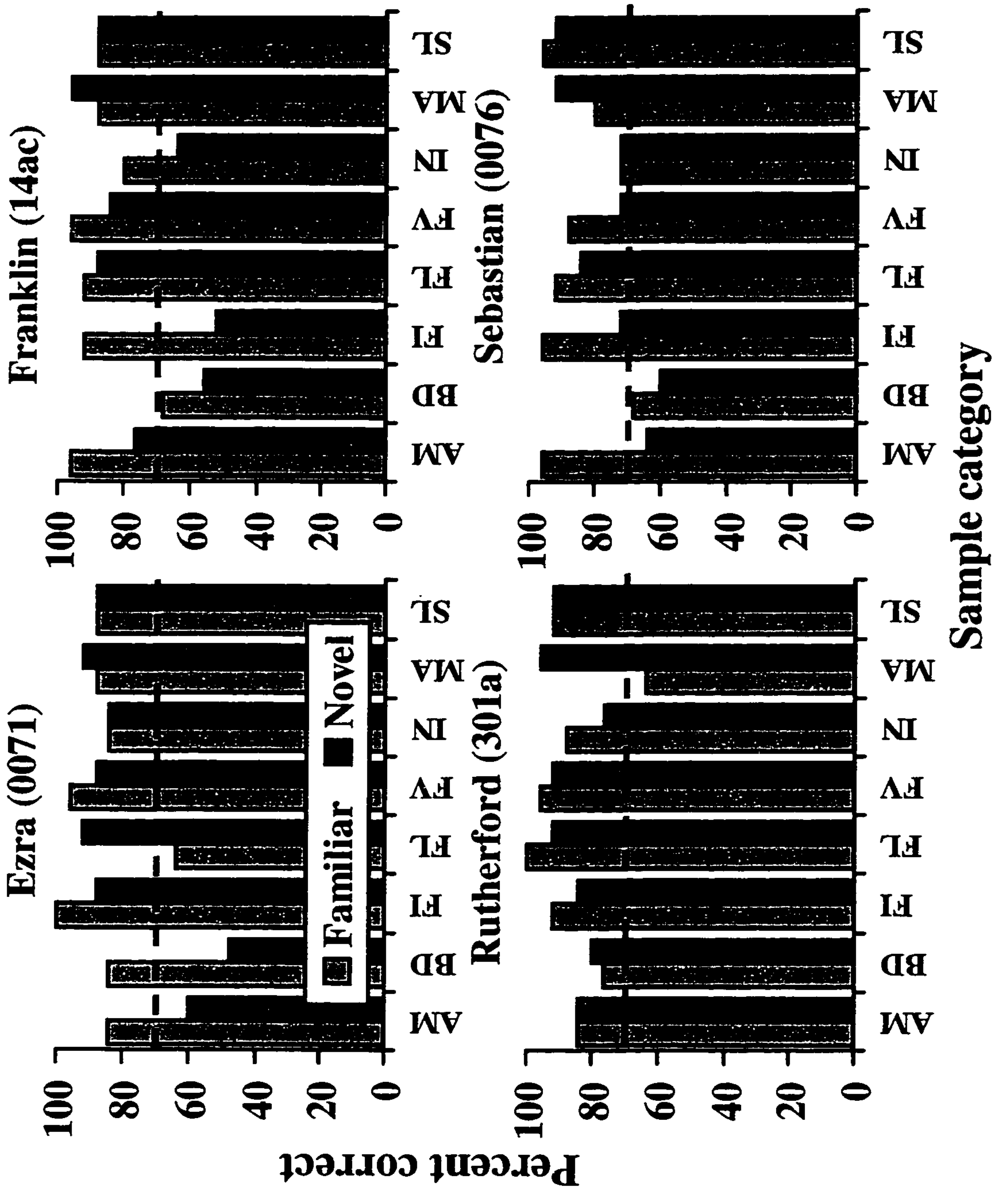


Figure 6. The results of the first transfer test broken down by category for Experiment 1. The light gray bars show accuracy on familiar stimuli and the dark gray bars show accuracy on novel stimuli. The upper line is chance based on the number of trials in a session, which was 68%. Chance on any given trial was 50%.

	Above chance			Novel diff. Fam.		
Subject	**MTS**	**DMTS**	**MDMTS**	**MTS**	**DMTS**	**MDMTS**
Ezra	FI FL FV IN MA SL	AM BD FI FL FV MA SL	AM FI FV IN MA SL	None	None	None
Franklin	AM FL FV MA SL	All	AM FI FV IN MA SL	AM FI	None	AM BD FI
Rutherford	AM BD FI FL FV IN SL	BD FL FV IN MA SL	AM FI FV IN MA SL	None	None	AM FL
Sebastian	FI FL MA SL	AM BD FL MA SL	AM FV IN MA	None	None	BD FL SL

* Familiar significantly lower than novel.

Table 3. Evaluation of transfer tests on the basis of those categories that showed a significant difference between familiar and novel stimuli and above chance performance on novel items. MTS=simultaneous (Experiment 1), DMTS=single-delay MTS (Experiment 2), and MDMTS=multiple delay within a session (Experiment 3).

as well as whether novel items were significantly different from familiar items on each of the categories for this experiment, a category was judged to have successfully transferred to novel stimuli when performance on novel items was not significantly different from familiar items, and performance was above chance. In that respect, each monkeys each showed transfer on several categories. Ezra successfully transferred to all categories except amphibians and birds, though his performance on flowers improved significantly between familiar and novel stimuli. Rutherford showed successful transfer to all categories, but performance on mammal faces was significantly worse for familiar stimuli than novel. Franklin successfully transferred to flowers, fruits and vegetables, mammal faces and shells. Sebastian successfully transferred to flowers, mammal faces and shells.

Discussion

The monkeys were able to learn to perform an identity match on stimuli containing more information than colored disks or monochromatic geometric shapes. Three of the monkeys took approximately 2,500 trials to learn this task, with the fourth monkey taking almost twice that number. The subjects in this experiment were within the ranges previously reported for monkeys (D'Amato, Salmon, and Colombo, 1985) on colors and geometric shapes.

A number of researchers feel that conceptual behavior in animals should be demonstrated with trial-unique stimuli, to avoid having the animal learn to respond to specific stimuli, or to particular configurations of stimuli. In this experiment, as well as the subsequent ones, a limited set of stimuli was used, and a stimulus could appear as either a sample or a distractor on any given trial. In MTS experiments using trial-unique stimuli responding can simply be on the basis of novelty rather than identity. In the present research, the monkeys were required to pay attention to the particular features of the stimuli since they could appear as either the sample or the distractor on any given trial. Additionally, after the first presentation, all the stimuli were familiar, thus ruling out responding on the basis of novelty.

During the first block of training most of the subjects began approximately at chance levels of accuracy, as indicated by the results of the z-test. Franklin was an exception; he had a very high number of categories on which he performed significantly below chance. An examination of his history as an experimental subject might explain his performance. Prior to learning this task, Franklin participated in a number of serial learning experiments. The procedure used in those studies required him to touch a series of items in a particular order. The same item was never presented twice in a trial, and there were no consequences to touching the same item more than once. Franklin initially may have attempted to solve the MTS problem in the same way he would a serial learning problem; after touching the "first" item, the sample in the present case, he moved on to the next item of the "list," the distractor here. On the serial learning task he had some probability of being correct, depending on how many additional items were on the screen. In this case, this strategy resulted in a very high number of errors, and significantly below-chance performance during the initial stages of training.

An analysis of performance on the various categories indicated that three of the subjects had difficulty performing the identity match on one or more categories. Poor performance on those categories was without regard for the category of the distractor stimulus. Performance on individual items (see figures B1, B2, B3 and B4 in Appendix B) suggests that accuracy on individual items rather than on a group of items comprising a category could account for some of the low scores found in the last quarter of training on this task (see figure B5 in Appendix B for performance on individual items for the last block of training). There was a tremendous amount of variability across items within certain categories, though there were a few instances of comparable acquisition of all the items in a particular category.

Transfer tests are crucial to assessing the identity concept. All the monkeys performed successfully above chance in the first session on which they were presented with novel stimuli, and performance was not significantly different from performance on the final session with familiar stimuli. The monkeys were able to successfully transfer the identity match to novel stimuli, since their overall percent correct for the first day of the novel set were all significantly above chance.

In terms of the categories, successful transfer was also determined both by performance in comparison to chance and by whether or not there was a significant change in performance on a particular category when presented with novel stimuli. Each monkey showed a significant change in performance for at least one category on the transfer test. Performance on nearly all the novel items of certain categories was below chance (see Figures B6, B7, B8 and B9 in Appendix B). Conversely, performance on several categories remained very accurate across all items, even novel items. All the monkeys performed significantly above chance on all, or all but one, of the stimuli composing fruits and vegetables and mammal faces. Ezra and Sebastian continued to do well on the shell category, even when presented with novel stimuli, and Rutherford and Ezra performed significantly above chance on all pictures of novel flowers.

Since two different sets were used, and since the monkeys all showed a great deal of similarity in terms of performance on a few particular categories, there was some suggestion that the features of certain categories may lend themselves to transfer of the identity matching more readily than others. Two of the categories, fruits and vegetables and mammal faces may be explained by experience. All the monkeys were given a variety of fruit, and undoubtedly encountered some or all of the items pictured in the stimulus set. There is some suggestion (Neiworth and Wright, 1994) that experience with real-world objects enhances conceptual behavior. The same may also be true of mammal faces. Though the monkeys experience with faces was limited to humans and conspecifics, all mammal faces have a specific and consistent configuration. Most importantly, all the mammal faces used here had very prominent eyes, which may have drawn the monkey's attention. Their experience with even a limited set of faces may have had some effect on their ability to transfer the identity match to novel photographs of mammal faces.

Rutherford's performance on the transfer test for faces, however, questions this explanation, since his performance improved substantially when presented with novel stimuli. This increase in accuracy from familiar to novel stimuli suggests that there were potential physical factors other than features of the category. It also should be noted, however, that within most categories, there was a great deal of variability in performance across novel items, again suggesting that individual stimuli as opposed to categories were more important in controlling responding.

In summary, the results of this experiment suggest that monkeys can successfully learn the identity matching task with a large set of photographic stimuli, and they can sustain overall performance at above-chance levels when presented with novel stimuli. The discriminability of individual items appeared to play a more significant role in performance, without regard to the category to which they belong. All the monkeys did, however, demonstrate a high level of transfer to certain categories, such as fruits and vegetables and mammal faces.

Experiment 2

Once the monkeys acquired and demonstrated successful transfer of the simultaneous MTS to novel stimuli, a delay was introduced to examine memory for individual stimuli and transfer of identity matching. A large number of researchers have found that delay durations have an effect on performance. Specifically, as the delay increases, accuracy in a DMTS task tends to decrease.

Decreased accuracy due to longer delays compared to no or short delays has been demonstrated in pigeons for fairly moderate delays, and with substantial effects on performance. Berryman, Cumming, and Nevin (1963) trained pigeons to match colored response keys in a DMTS task with delays of up to 24 seconds. The pigeons could not learn the MTS task when first trained using delays that varied across sessions. Only after several sessions of simultaneous MTS training were the pigeons able to learn the task. Then, when placed back on the variable DMTS procedure, performance for the shorter delays was high, but quickly dropped off as the delay increased. Specifically, they found that the pigeons ability to respond accurately on this task dropped from close to 95% in the simultaneous MTS condition to 50% with a delay of 24s, and that the amount of errors was closely related to the length of the delay (i.e. accuracy decreased more or less monotonically with respect to the length of the delay).

Perkins, Lydersen, and Beaman (1973) found similar results in a series of experiments that specifically looked at the effects of delay on acquisition of the identity match in pigeons. In one manipulation, different delays were associated with specific stimuli. In the other manipulation delays were presented randomly across trials and not associated with specific stimuli. When specific delays were associated with specific stimuli, all the pigeons showed acquisition. When delays were mixed only one pigeon out of three acquired the matching task. Since the pigeons were able to acquire the matching task when delays and stimuli were consistently paired they were obviously able to remember the stimuli for the delay durations used in these experiments (2s to 6s).

White and Cooney (1996) trained pigeons on a combination of short and long (4s) delays within a session with differential reinforcement outcomes. They varied reinforcement at short intervals and kept reinforcement the same at long intervals and vice versa. They found that performance at one retention interval was independent of factors biasing performance at another. Additionally, results showed that accuracy decreased as a function of delay.

Clearly, pigeons are not able to remember stimuli for any length of time in a DMTS task, since their performance decreases to chance levels at approximately 24s and delays of 4s to 6s are considered to be "long" by most researchers using pigeons. However, there is evidence that pigeons can learn to perform adequately when specific delays are explicitly paired with specific stimuli, supporting the notion that pigeons have a tendency to perform matching tasks by learning to make stimulus-specific responses based on reinforcement history. Many researchers have had better results with monkeys in matching tasks, and as a consequence, the effects of delay on performance is of interest.

In a series of experiments addressing short-term memory for haptic cues in monkeys, Bauer and Steele (1985) found a decrease in percent correct as a function of the retention interval that was independent of size and shape of the sample and comparison. They used retention intervals of up to 64s, which were considerably longer than those used with pigeons. At each retention interval accuracy was higher when comparisons differed in both size and shape than in only one dimension, when the size differences between comparison stimuli increased, and when the samples were spheres instead of cubes. Dimensions of the haptic stimuli influenced discrimination but had little effect on short-term memory.

A delayed non-MTS, a variation of the DMTS procedure was used by Moss, Rosene and Peters (1988) and Rapp and Amaral (1991) was used to examine age differences in monkeys. Using delays of up to 10 minutes, a monkey's ability to select a novel item over a familiar item after a variable delay decreased as a function of the delay. There were significant differences between young and old monkeys, but all showed a decrease in percent correct as the retention interval increased.

Roberts, Mazmanian, and Kraemer (1987) used delays ranging from 0.5s to 10s in a session to test squirrel monkey memory for picture fragments. Pictures were masked in the center, in the periphery, or with a random dispersal, and covered 25%, 50% and 75% of the picture. Subjects were shown the whole picture as the sample, and had to respond to the masked image, and vice versa. Recognition accuracy improved as the percentage of the picture exposed increased and accuracy was lowest with the center mask. They also found that accuracy decreased as a function of the length of the delay.

In this experiment, a progressively increasing delay was added across sessions to determine whether or not the monkeys were able to learn and transfer the identity concept with a delay. A larger stimulus set composed of items from the same categories as those used in Experiment 1 was presented to the monkeys. Additionally, trials with distractors from the same category as the sample were added

Method

Subjects. The same four subjects who served in Experiment 1 served as subjects in Experiment 2.

Apparatus. The same apparatus used in Experiment 1 was used here.

Stimuli. The first stimulus set each monkey was trained on in this experiment was the novel set presented during the transfer test from the previous experiment. Once a monkey reached criterion on this training set, he was given a transfer test with the second set of 64 stimuli that was novel to that particular monkey (see Table 1). See Appendix A for stimuli.

Procedure. Initially, training continued exactly as described in Experiment 1. Once an animal achieved a criterion of two consecutive sessions at 85% correct or better, a delay was introduced between the sample and the choice phase of the trial. The same delay was used across trials within a session.

The number of trials and delay length were adjusted for each individual based on performance. The progression of delays was as follows: 0s, 1s, 2s, 4s, 8s, 12s, 16s, 20s, etc. Criterion for increasing the delay was initially two consecutive sessions at 85% correct or better. Once the monkeys reached a 20s delay, the number of trials was gradually increased from 64 to 100.

Rutherford progressed to a 24s delay, and Franklin progressed to a 32s delay. Both monkeys remained at these delays for a substantial number of trials, and began to exhibit a number of motivational problems, such as excessive balking, low scores, and auto-aggressive behavior. Consequently, their delay was reduced to 20s. Rutherford continued to exhibit a number of problems, and his delay was further reduced to 16s, then 12s. At this point, criterion for all subjects was reduced to 80% correct or better for two consecutive sessions.

Sebastian's delay was also reduced first to 16s, then to 12s, due to low scores and a large number of trials on the 20s delay. At the same time, the number of trials in a session for both Rutherford and Sebastian was reduced to 70.

After Franklin and Ezra reached criterion on a 20s delay, and Rutherford and Sebastian reached criterion on a 12s delay all monkeys were then transferred to a novel set of 64 stimuli (the second set of 64 stimuli). In transfer the same delay that was presented during previous sessions was presented on 86-90% percent of the trials (depending on how many trials were in a session) with 5-7% of the trials presented with longer and shorter delays. In Franklin and Ezra's cases, the delays ranged from 16s to 24s in 4s increments. For Rutherford and Sebastian, the delays ranged from 8s to 16s in 4s increments.

Results

Trials to criterion. The completed trials to criterion for each delay are presented in Figure 7. The number of trials Ezra needed to reach criterion increased as the delay increased. The other monkeys were more variable, needing comparatively fewer trials to reach criterion. Acquisition of the 8s delay appeared to be an exception for all monkeys except Ezra. However, a one-way ANOVA on completed trials to criterion for each delay indicated no significant differences between the number of trials needed to acquire a particular delay, F(8,31)=1.39.

Percent correct. Figure 8 shows overall percent correct for each delay for each subject. The error bars indicate minimum and maximum percent correct scores for that particular delay. Scores were above chance (68%). A line of best fit for session percents correct indicated y-intercepts that were significantly different from a theoretical y-intercept of 50% (see Table C1 in Appendix C). A similar analysis of the slopes showed that all except the slopes for the acquisition of the 1s delay were not significantly different from a theoretical slope of 0 (see Table C1 in Appendix C).

Two-way 9 X 8 ANOVAs on delay vs. category were performed on the arcsine-transformed proportion correct for each subject. Criterion sessions were used because

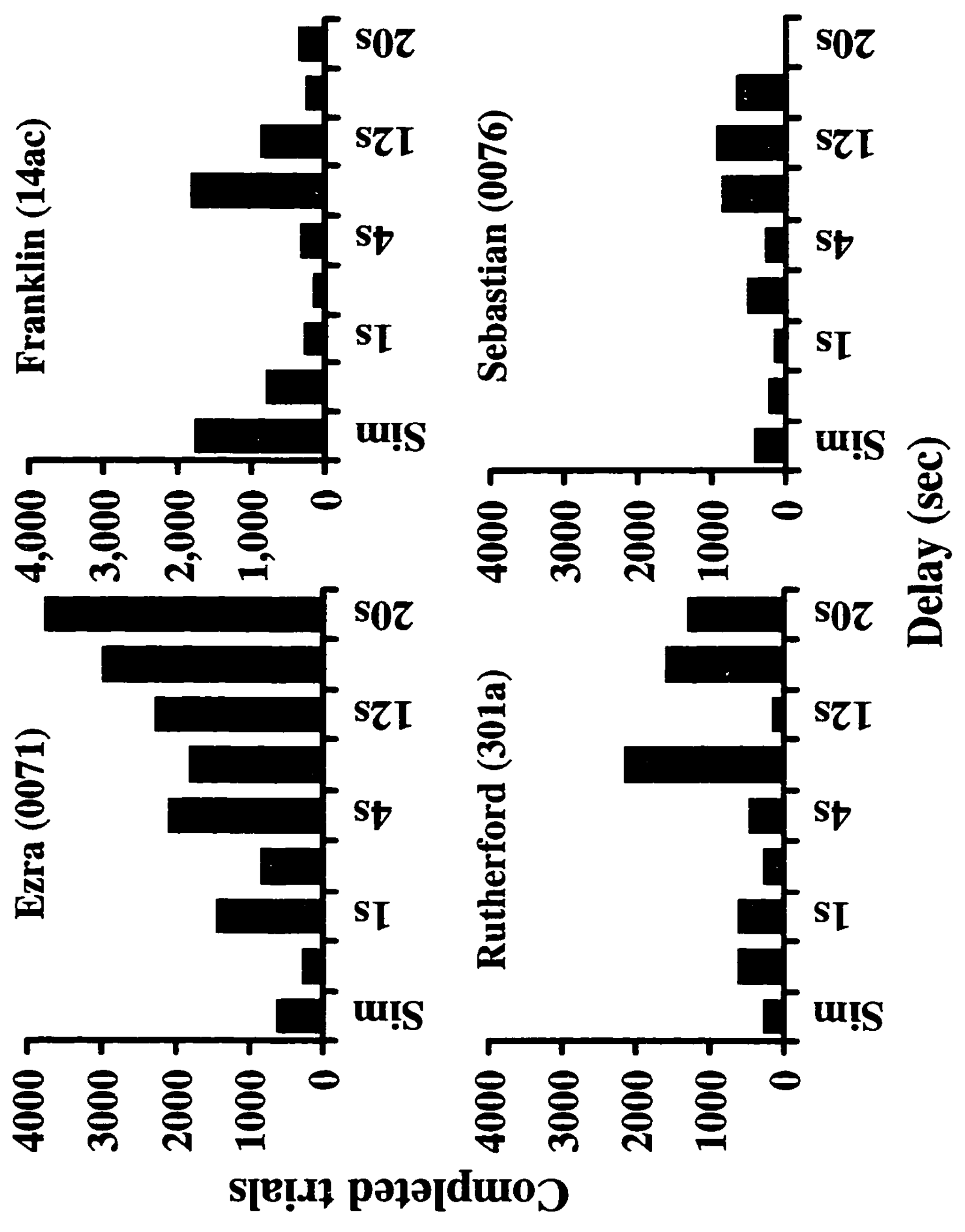


Figure 7. Overall completed trials to criterion for each delay for Experiment 2. Sebastian failed to reach criterion on the 20s delay. All four monkeys took a variable number of trials to reach criterion for each delay range, but only Ezra showed a systematic increase in the number of completed trials to criterion as the delay increased. All monkeys show a substantial increase at the 8s delay but this increase was not significant.

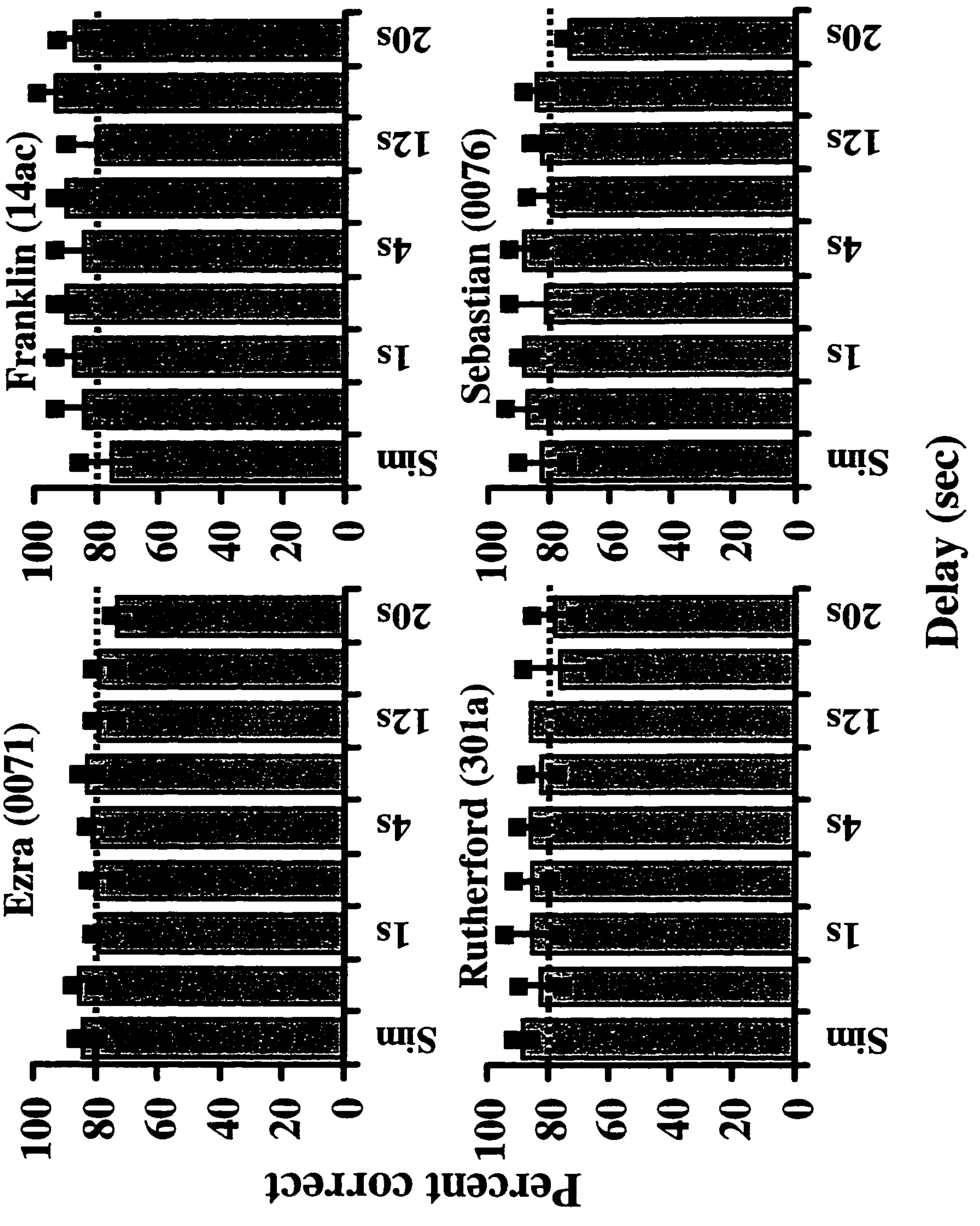


Figure 8. Overall percent correct for each category on each delay for Experiment 2. The error bars indicate the standard deviation of percents correct. The dotted line represents criterion. The overall scores for all the monkeys across delays was very high, and variability, as indicated by the error bars, was low.

there were several cases in which a subject completed a particular delay within two sessions. The total number of trials for each category was divided into four blocks, and a percent correct calculated for each block. All subjects showed a significant main effect of category [Ezra - *F*(7,72)=3.81, *p*<0.001; Franklin – *F*(7,72)=4.07, *p*<0.01; Rutherford – *F*(7,72)=2.68, *p*<0.016; Sebastian – *F*(7,64)=4.15, *p*<0.001]. These data are presented in Table 4, along with the results of the post-hoc analysis.

Response time. The natural log-transformed correct median choice response times for each block of the criterion sessions were also calculated for a 9 X 8 two-way ANOVA on delay vs. category. Only correct median response times were used because there were too few errors to allow a comparison between correct and incorrect response times. Three subjects showed a significant main effect of delay [Ezra – *F*(8,72)=14.67, *p*<0.001; Rutherford – *F*(8,72)=27.35, *p*<0.001; Sebastian – *F*(7,64)=31.34, *p*<0.001]. Tukey's post-hoc analysis indicated that correct response times on trials with extreme delays (i.e. 0s vs. 20s) tended to differ significantly from response times on trials with intermediate delays. As Figure 9 shows, the correct median response time tended to increase as the delay increased.

Rutherford also had a significant main effect for category, *F*(7,72)=5.40, *p*<0.001. Tukey's post-hoc test showed that Rutherford's response time to flowers (FL) was significantly different from several other categories. Finally, Franklin's data revealed a significant interaction between category and delay, *F*(56,72)=2.15, *p*<0.01. A simple main effects analysis indicated an effect of category at the simultaneous, 2s, 4s, 12s and 20s delay, and an effect of delay on all categories excepts birds.

	Percent correct		Response time	
Subject	Category	Different from	Delay	Different from
Ezra	AM	FV	1s	8s, 12s, 16s, 20s
			2s	Sims, 0s, 4s, 8s, 12s, 16s, 20s
			8s	0s, 1s, 2s, 4s, 12s
Rutherford	FI	MA	Sim	16s, 20s
			0s	Sim, 4s, 8s, 16s, 20s
			1s	4s, 8s, 16s, 20s
			2s	8s, 16s, 20s
			4s	16s, 20s
			8s	20s
Sebastian	AM	FL, MA	Sim	1s, 4s, 8s, 16s
			0s	1s, 2s, 4s, 8s, 12s, 16s
			2s	1s, 8s, 12s, 16s
			4s	16s

Table 4. Summary of post-hoc analysis on two-way ANOVA for delay and category with accuracy and response time as the dependant measures respectively.

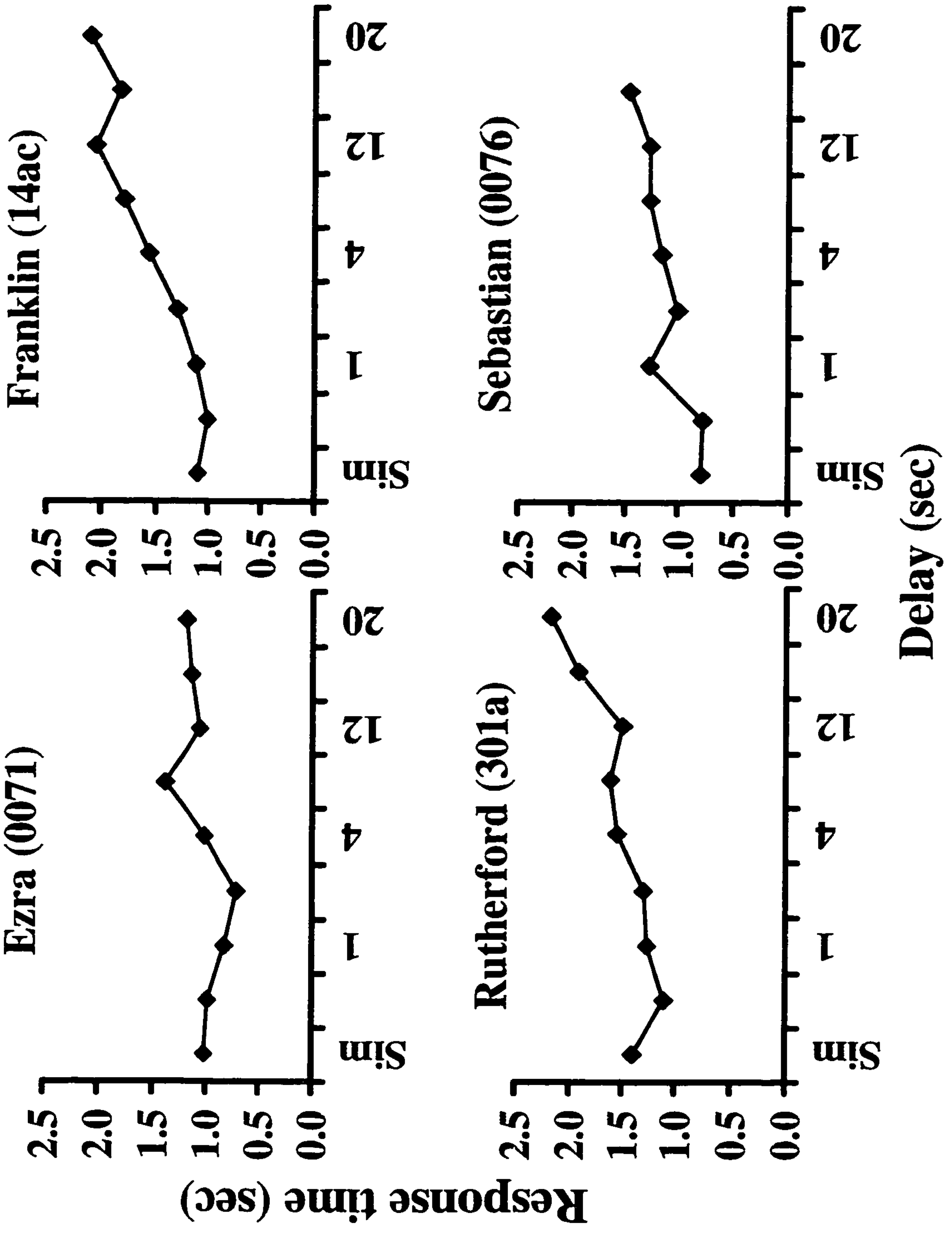


Figure 9. The median response time for correct trials on each delay of Experiment 2. Response time to correct trials tended to increase as the delay increased.

Delay transition. Where possible, performance on the transition from one delay to the next was also examined in a manner similar to that used for the general transfer test. The first and last 25 trials of two consecutive delays were compared to determine whether category, delay, or a combination of these factors had any influence on performance. A series of two-way ANOVAs on category vs. delay were performed on the arcsine-transformed proportion correct data. In cases in which the new delay was acquired within the minimum number of sessions necessary (two) an analysis of performance transition was not done for two reasons. First, there was a limited amount of data in those cases. Secondly, an analysis in those cases seemed moot since the new delay clearly had no effect on performance as it remained at the criterion level. Table 5 shows the percent correct scores for each category for each transition. The statistical analyses revealed no significant main effects, and no significant interaction between delay and category.

Within and between category performance. The two left-most columns in each panel of Figure 10 show the overall percent correct for trials on which the distractor was sample (same) and trials on which the distractor was from a different category (different) for this experiment. A paired t-test on accuracy for all subjects showed no significant difference between overall scores on same and different, $t(3)=-0.33$. Chi-square analyses for accuracy on sample-distractor pairs for each delay showed that the distractor category had no effect on accuracy to the sample.

		Sim to 0s		0s to 1s		1s to 2s		2s to 4s		4s to 8s		8s to 12s		12s to 16s		16s to 20s	
	Dela	**Last**	**First**	**Last**	**First**	**Last**	**First**	**Last**	**First**	**Last**	**First**	**Last**	**First**	**Last**	**First**	**Last**	**First**
Ezra	**AM**	76	68	76	68	68	60	80	60	88	76	88	76	80	56	80	72
	BD	84	76	76	60	76	52	76	72	80	80	88	60	72	84	92	56
	FI	88	68	72	72	88	68	80	64	76	84	92	80	96	76	80	84
	FL	96	88	88	76	92	76	80	80	88	88	88	76	80	92	84	84
	FV	100	96	92	100	84	96	100	92	88	80	84	100	84	80	84	76
	IN	68	88	88	84	88	80	84	88	84	84	84	80	84	72	92	72
	MA	92	92	96	80	92	84	76	88	92	88	84	92	92	96	64	88
	SL	92	92	92	96	96	92	96	76	96	84	96	88	88	76	84	88
Franklin	**AM**	72	72	80	76	80	80			88	96	96	68	84	88		
	BD	60	72	92	68	72	75			76	80	76	72	92	83		
	FI	68	84	80	84	84	93			84	76	76	52	88	93		
	FL	84	92	88	92	96	90			100	88	96	80	92	100		
	FV	92	88	96	92	92	95			96	92	92	88	100	100		
	IN	52	76	76	88	88	87			84	92	92	72	96	100		
	MA	72	76	96	88	88	88			84	88	88	96	88	95		
	SL	68	80	96	96	96	100			84	96	100	72	88	82		
Rutherford	**AM**	88	64	88	60	84	68	80	88	80	72	72	75			84	80
	BD	84	56	80	80	84	80	84	76	92	72	92	93			88	72
	FI	88	76	72	56	84	80	68	84	72	80	92	83			84	72
	FL	88	96	92	84	96	96	92	96	100	88	92	82			96	84
	FV	96	92	100	72	100	88	88	88	92	100	84	93			80	76
	IN	76	76	80	92	88	84	96	72	80	92	72	75			76	84
	MA	96	76	100	68	100	92	96	84	96	88	92	100			88	84
	SL	88	80	84	100	92	80	84	96	84	88	92	87			88	76
Sebastian	**AM**	80	88					76	76	68	52	84	84	88	84		
	BD	84	88					84	88	84	56	96	68	92	88		
	FI	84	83					88	84	84	88	80	76	84	80		
	FL	84	96					100	84	88	96	96	92	88	96		
	FV	80	84					92	92	84	88	92	92	80	100		
	IN	84	80					84	84	88	72	76	68	84	76		
	MA	96	96					88	100	84	76	88	80	92	72		
	SL	96	87					92	92	88	88	92	84	92	84		

Table 5. Mean percent correct for each category for the delay transitions. The gaps indicate where the monkey acquired the new delay within the minimum number of sessions available.

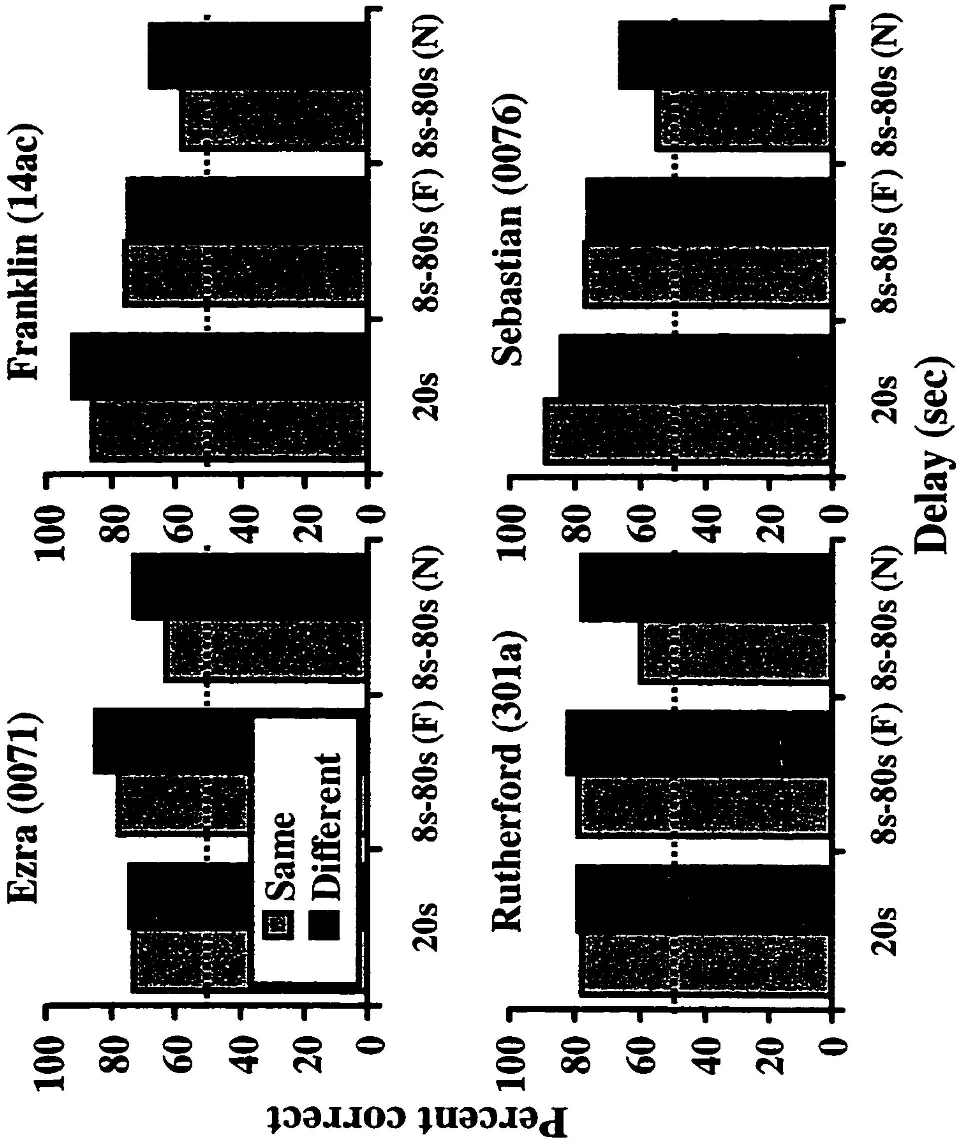


Figure 10. Starting with Experiment 2, a proportion of the trials in a session had sample and distractor stimuli from the same category (designated as same trials). The gray bars represent overall percent correct for same trials. The dark gray bar represents accuracy on those trials in which the sample and distractor were from different categories. Data from three stimulus sets and two delay conditions are presented. There were no significant differences between same and different for the single delay condition, or the familiar 8s to 80s condition. Accuracy on same and different did differ significantly for the final set of novel stimuli presented under the variable delay condition.

Transfer test. The middle set of columns for each graph of Figure 4 show the overall percent correct for the last session of familiar stimuli and the first session of novel stimuli for each subject. Overall accuracy on novel stimuli was above chance, and a paired t-test revealed no significant difference between performance on familiar and novel stimuli, t(3)=1.53. The middle column in Figure 5 shows accuracy on the items for their first presentation. As was the case for the transfer test in Experiment 1, all monkeys performed well above chance on the first presentation of novel stimuli.

The first and last 25 trials for each set were grouped into five-trial blocks in order to determine whether the identity match could successfully transfer to novel stimuli under delay conditions. The transfer test data are presented in Figure 11. Two-way 2 X 8 ANOVAs (stimulus set vs. category) were performed on the arcsine-transformed proportion correct for each block for each subject. The analysis revealed no significant differences between the stimulus set (familiar vs. novel) or categories for any of the subjects. Ezra performance on insects was below chance for both familiar and novel, and there are several categories on which his performance was below chance on familiar stimuli. This was also true for Rutherford. Sebastian's performance on novel fish, fruits and vegetables and insects were also below chance.

Discussion

In this experiment the monkeys were able to maintain high levels of accuracy even with delays as long as 20s with a large stimulus set of color photographic images. The lack of significant differences throughout this experiment suggests that once the monkeys learned the identity match, the delay had little effect on performance. Certainly,

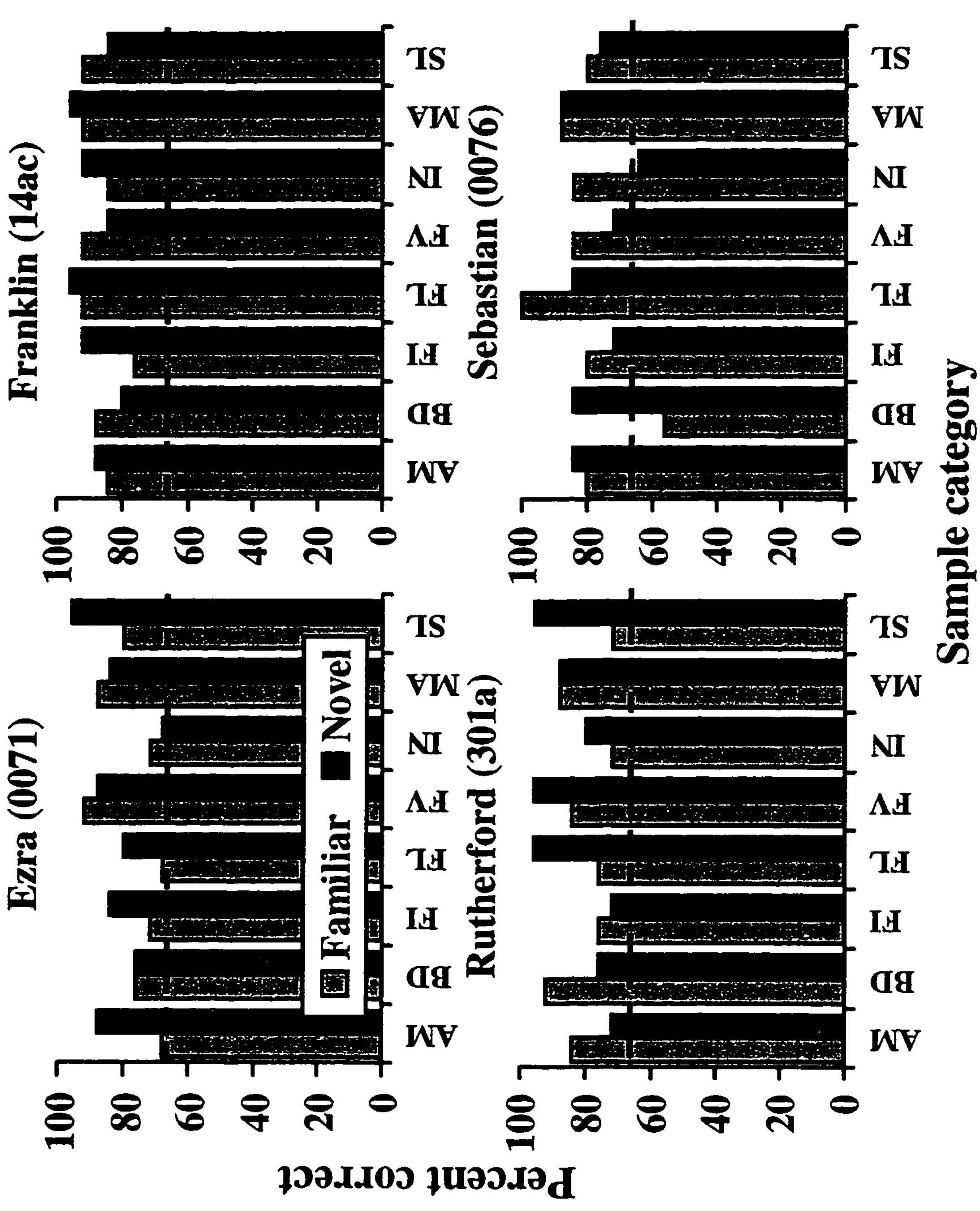


Figure 11. The results of the transfer test for Experiment 2 broken down by category for each subject. There were no significant differences between familiar (light gray) and novel (dark gray) stimuli for any subject on any category and performance on novel stimuli was above chance for all categories except insects for Sebastian. The dashed line represents chance based on the number of trials (68%).

systematically increasing the delays across sessions produced no significant decrease in accuracy, regardless of the length of delay used within the session. However, the longest delay used here was fairly short with respect to what other researchers have used with monkeys. Another significant aspect of the lack of delay effects in this experiment relates to the type of stimuli used here. As noted previously, many researchers using MTS and DMTS tend to use colors, forms, or some combination of the two. In the present experiment the delay had little effect on the monkeys ability to match multidimensional stimuli. Additionally, under delay conditions, the monkeys were able to transfer matching to novel stimuli. That is, they were able to remember the novel photographic stimuli for as long as 20s and then make the correct choice. Roberts, Mazmanian, and Kraemer report a decrease in memory for photographs as delays reach approximately 10 seconds for squirrel monkeys. However, they also asked the monkeys to match unobscured pictures with obscured pictures, and this may have been a factor in the decrease in accuracies above and beyond the length of the delay.

There were 64 stimuli in the sets used for this experiment, with most of the items appearing in nearly every session. After the first few days of sessions using these stimuli all the stimuli were familiar to the monkeys. Given a set of this size, there were 4,032 unique trial combinations, and the script files which controlled stimulus presentation across sessions were constructed so that each unique sample-match-distractor combination appeared only once in a set of 4032 trials. While it is possible that the monkeys may have become very familiar with the individual items that comprised the set it is doubtful they were able to memorize particular sample-match-distractor combinations.

More importantly, the monkeys showed no significant differences between familiar and novel stimuli on the transfer test. Once again, the overall percents correct for the first session on which the monkeys were presented with the novel stimuli were significantly above chance and not significantly different from one another. If the monkeys had been relying on memory to perform the MTS throughout the training with increasing delays, a disruption or decrease in performance would be predicted when novel stimuli were presented. This was not the case in this experiment. Perhaps, however, the monkeys attended to the novel stimuli more carefully because they were unfamiliar, resulting in successful transfer. If that were the case, one would predict a decrease in the percent correct for sessions following the initial presentation of the stimuli, since the stimuli would no longer be novel, but would presumably not be memorized either. Given the high, consistent responding throughout this experiment, however, this seems unlikely.

As was the case with Experiment 1, there was a great deal of variability in the accuracy between items within a category and no clear-cut effects of delay other than some suggestion that variability decreased (see Figures C1, C2, C3 ad C4 in Appendix C),. This, however, could simply be due to increasing familiarity with the stimuli.

Experiment 3

In this experiment the delay was varied within a session, so that the monkeys were unable to predict, from one trial to the next, how long the delay between sample and match-distractor pair would last. This procedure addressed the same questions about performance and transfer as a function of delay with photographic stimuli as addressed in Experiment 2.

There was an additional manipulation of the novel stimuli used in this experiment. Prior to the last set of stimuli, the items that made up a category were very variable in terms of the number of species represented within a category. For the final set in Experiment 3 the species were limited, thus significantly reducing the amount of within-category variability while maintaining across category variability. It was hypothesized that decreasing the variability within a category would reduce the effect of individual items within a category and increase the degree of within-category confusion, resulting in more errors on those trials on which the sample and distractor were from the same category.

Method

Subjects. The same four subjects who served as Experiment 1 and 2 served as subjects in Experiment 3.

Apparatus. The same apparatus used in experiment 1 and 2 was used here.

Stimuli. The transfer set of 64 stimuli used in Experiment 2 was the training set of stimuli in Experiment 3. Additionally, a new set of 32 stimuli was used as the transfer set for Experiment 3 (see Table 1). This new set of stimuli included the same categories as all previous sets, but whereas previous sets included a wide variety of category exemplars, this set included only a single species within a category. For example, previous sets contained a variety of insects as one of the categories. The insect category of the new set was comprised of different pictures of monarch butterflies only, with no other species represented. The breakdown of categories was as follows: amphibians=red tree frogs, birds=finches, fish=angelfish, flowers=daisies, fruits and vegetables=corn on the cob, insects=monarch butterflies, mammal faces=basset hound, shells=conch shells. Pictures within a category represented different individuals, and all members of a particular category were discriminable from one another when judged informally by humans.

Procedure. After the initial transfer test, shorter and longer delays were systematically added until the session consisted of trials with delays ranging from 4s to 40s in 4s increments (4s, 8s, 12s, 16s, 20s, 24s, 28s, 32s, 36s, and 40s). During this gradual introduction of multiple delays within a session, the number of trials for Ezra, Sebastian and Rutherford was gradually increased to 100 trials per session.

Franklin's delay range was increased by 4s for each delay (i.e. 4s to 40s became 8s to 44s, etc). This was done until Franklin's delay range within a session was 44s to 80s. At this point, he was placed on a delay range of 8s to 80s in 10s increments. The remaining monkeys were immediately placed on the 8s to 80s range of delays after six sessions (600 trials) of the 4s to 40s range of delays.

Each delay was randomly presented 10 times within a session. The monkeys were tested on this delay range for ten sessions (1000 trials) then given the transfer test using the novel set of 32 stimuli. The monkeys were then trained on this final set for 10 sessions (992 trials, comprised of all possible stimulus combinations).

Results

Acquisition. Acquisition of the set of stimuli presented using multiple delays in a session for each subject is shown in Figure 12. Scores remained in the 70%-80% range throughout the addition of delays during a session. There was a slight drop in scores for the last ten sessions when the delay range within a session was switched abruptly to a range of 8s-80s for three subjects. While Franklin received substantial interim training on these stimuli that the other monkeys did not receive, a one-way repeated measures ANOVA on the arcsine-transformed proportion correct data for these 10 sessions yielded a significant main effect of subject [$F(3,36)=5.66$, $p<0.003$). A Tukey's post hoc test revealed that Ezra's scores were significantly different from the other subjects. Franklin, Rutherford and Sebastian performed essentially the same on the last 10 sessions, despite the differences in training.

Percent correct. Figures 13, 14, 15 and 16 show performance for each category for each delay for each subject. The figure shows no relationship between accuracy on category and delay. This was further supported by the results of a chi-square analysis of the proportion correct for each category on each delay. None of the subjects showed a significant relationship between category and delay [Ezra – $X^2(63)=1.08$; Franklin - $X^2(63)=1.43$; Rutherford – $X^2(63)=1.05$; Sebastian – $X^2(63)=1.23$]. In light of this result, category and delay were further evaluated separately.

Figure 17 presents accuracy on each category for the 8s to 80s delay for each subject across the last 10 sessions without regard for delay. All scores were well above chance (60%). A one-way ANOVA on the arcsine-transformed proportion correct for

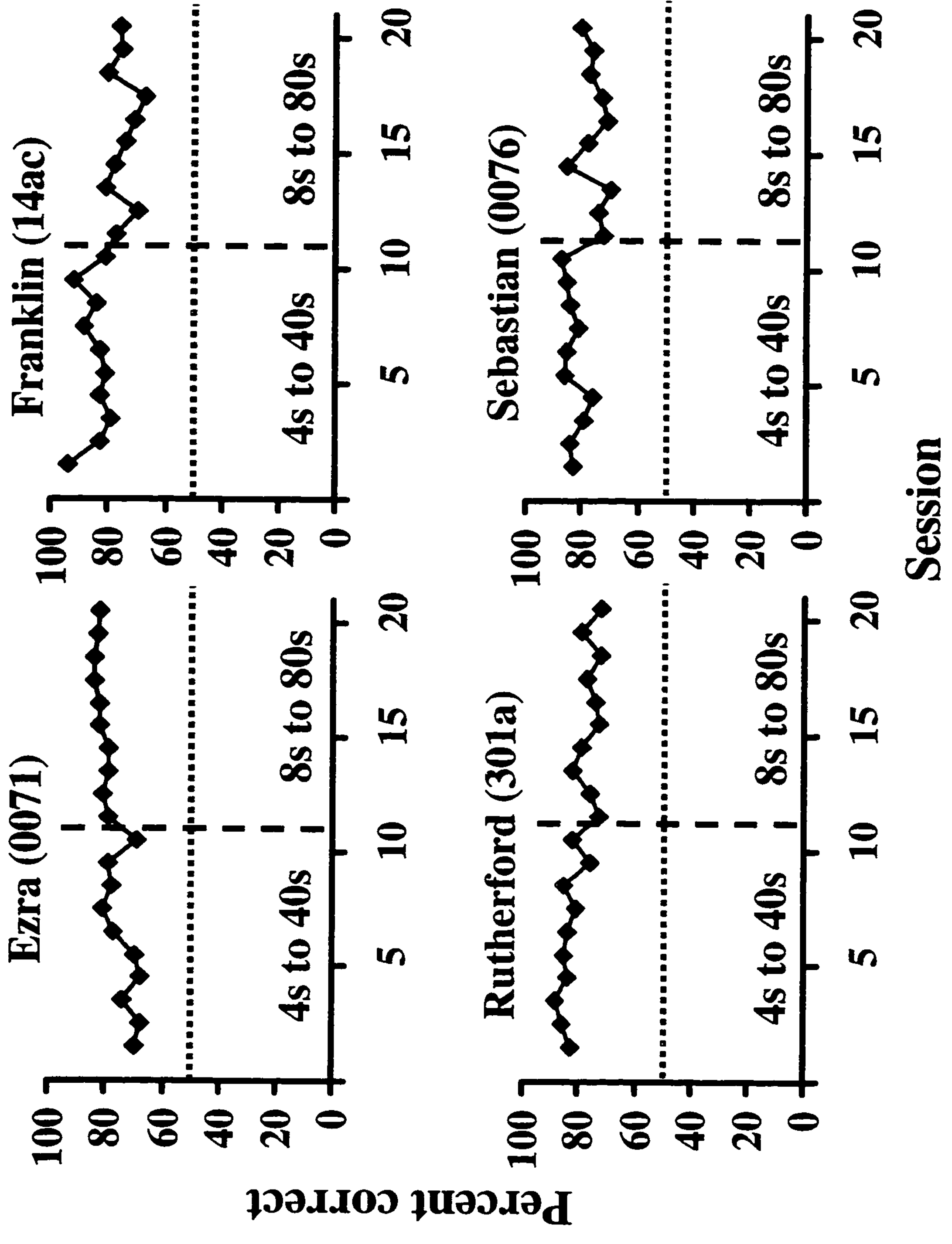


Figure 12. Overall percent correct by session for Experiment 3. The points to the left of the dashed vertical line indicate overall percent correct for each session while longer and shorter delays were systematically added to the session until a range of 4s-40s in 4s increments was achieved. At that point the delay range was switched to 8s-80s and the points to the right of the vertical dashed line indicate performance on that range. The same set of stimuli was used throughout this training.

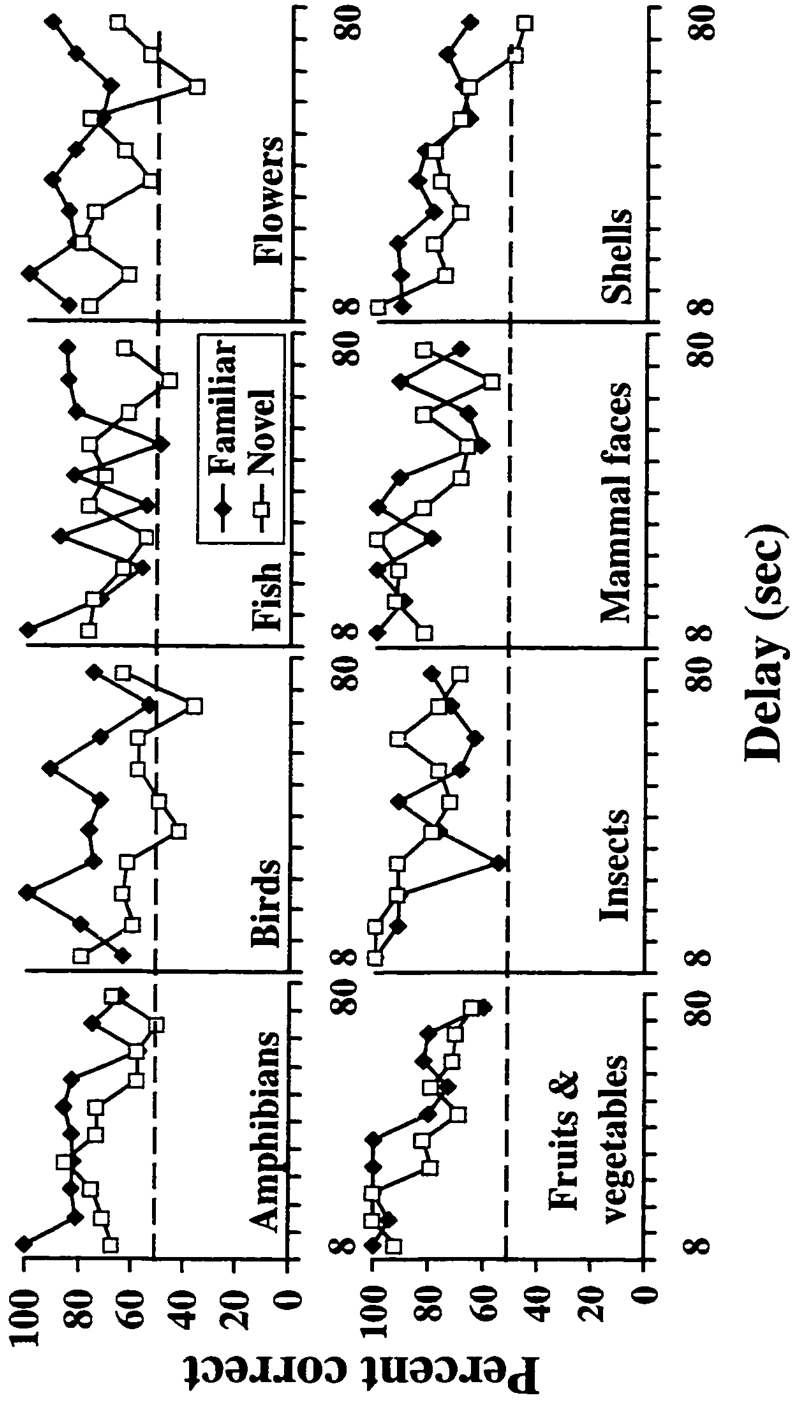


Figure 13. Performance on each category of stimuli for each delay for Ezra (0071). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel stimuli. The delay range was 8s to 80s.

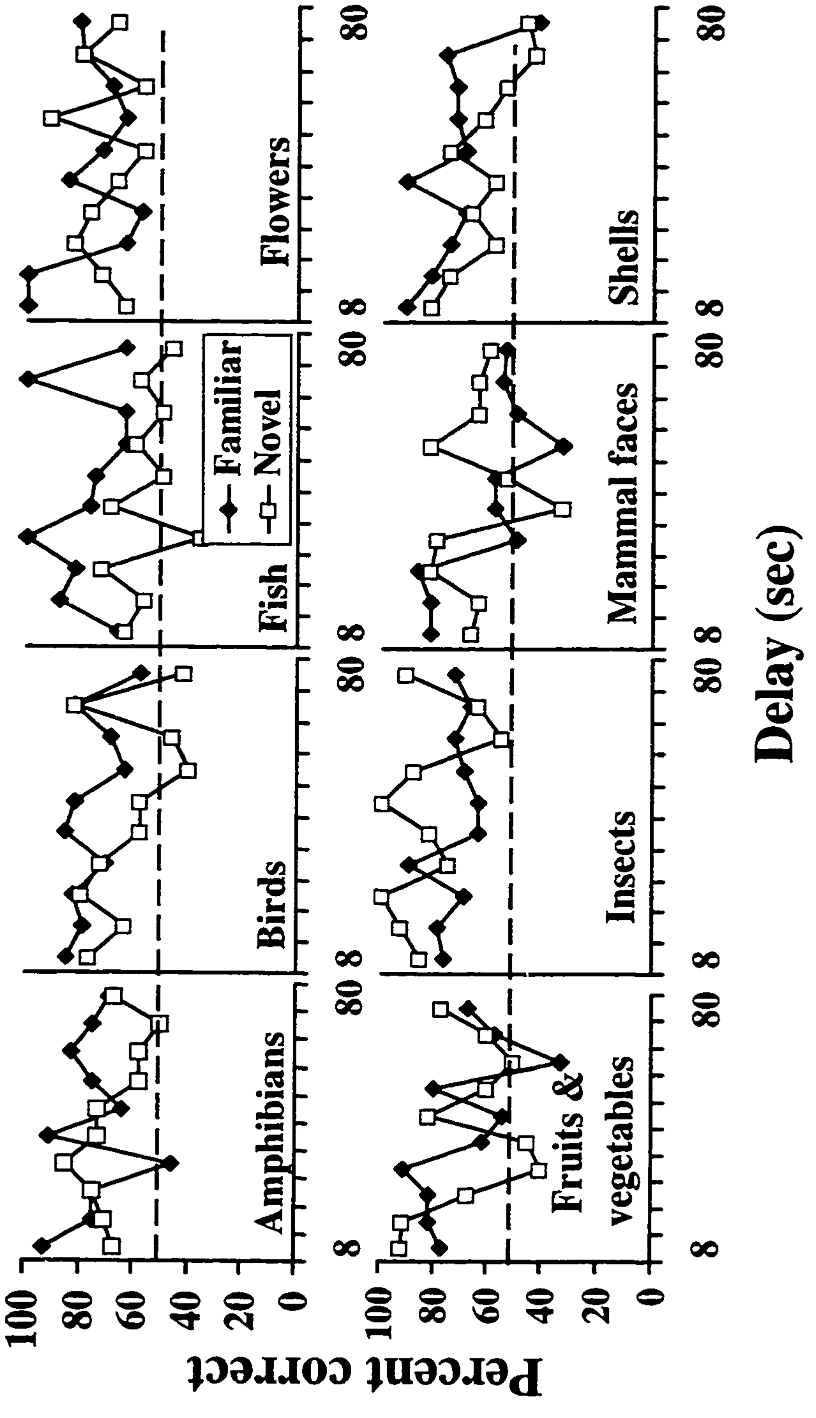


Figure 14. Performance on each category of stimuli for each delay for Franklin (14ac). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel stimuli. The delay range was 8s to 80s.

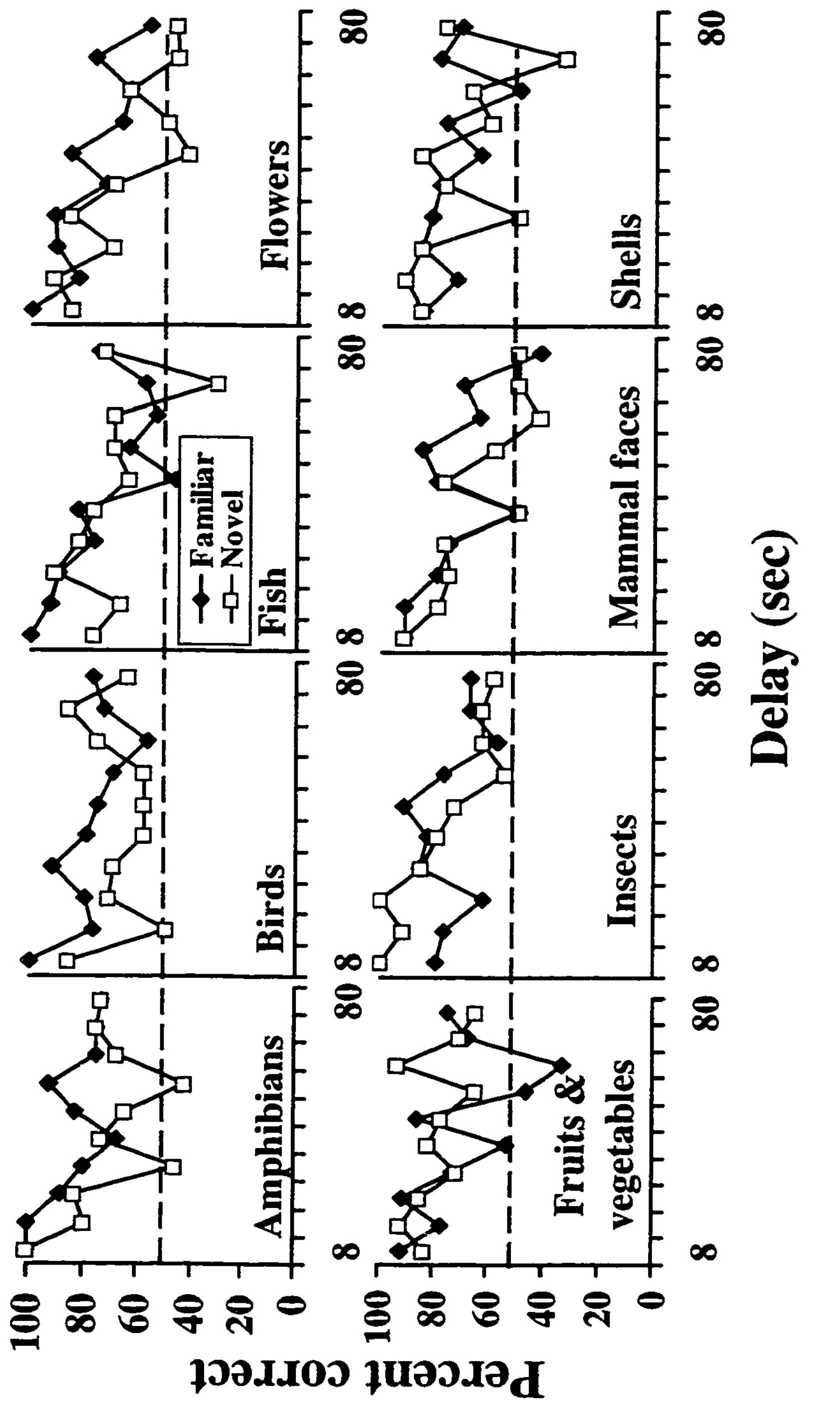


Figure 15. Performance on each category of stimuli for each delay for Rutherford (301a). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel stimuli. The delay range was 8s to 80s.

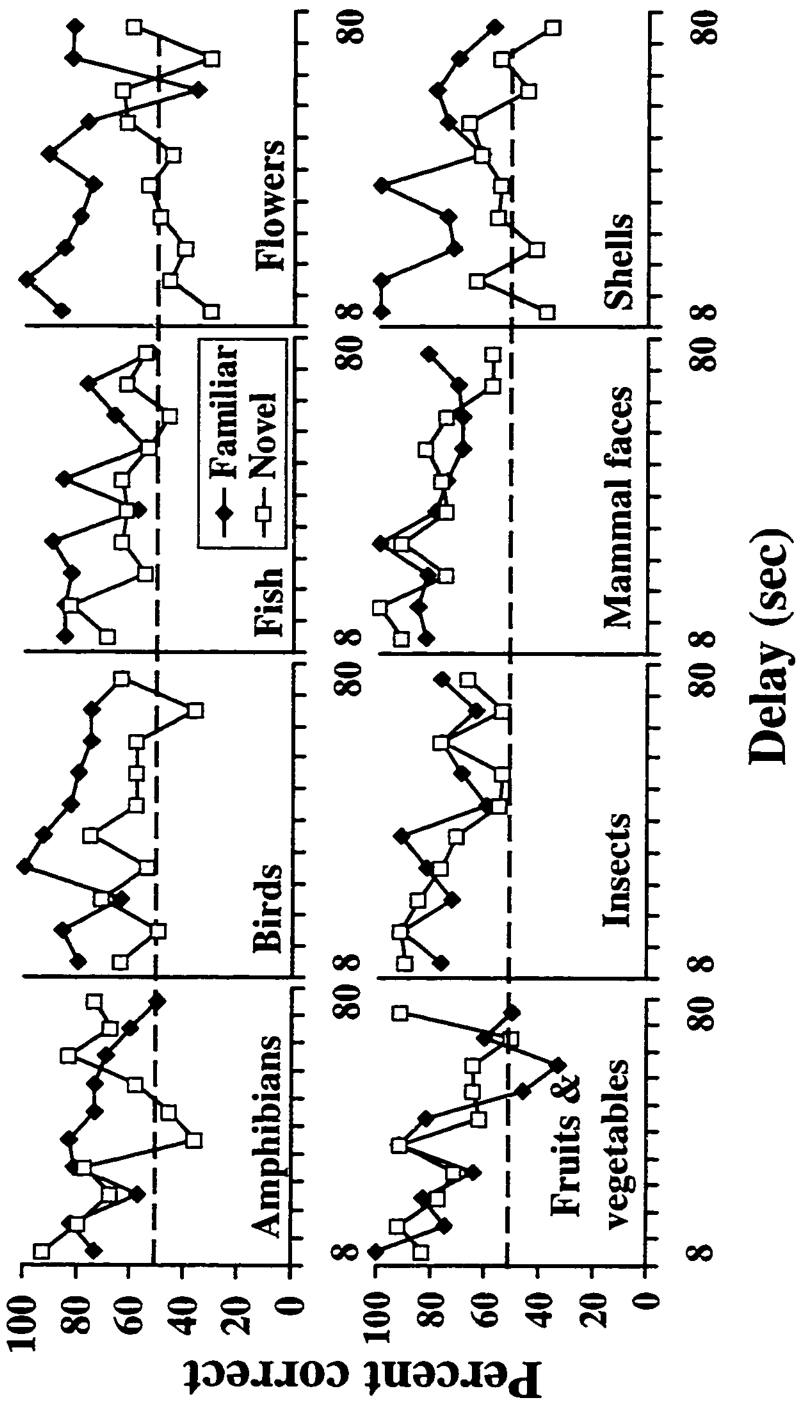


Figure 16. Performance on each category of stimuli for each delay for Sebastian (0076). The closed diamonds show accuracy on familiar stimuli and the open squares show accuracy on novel stimuli. The delay range was 8s to 80s.

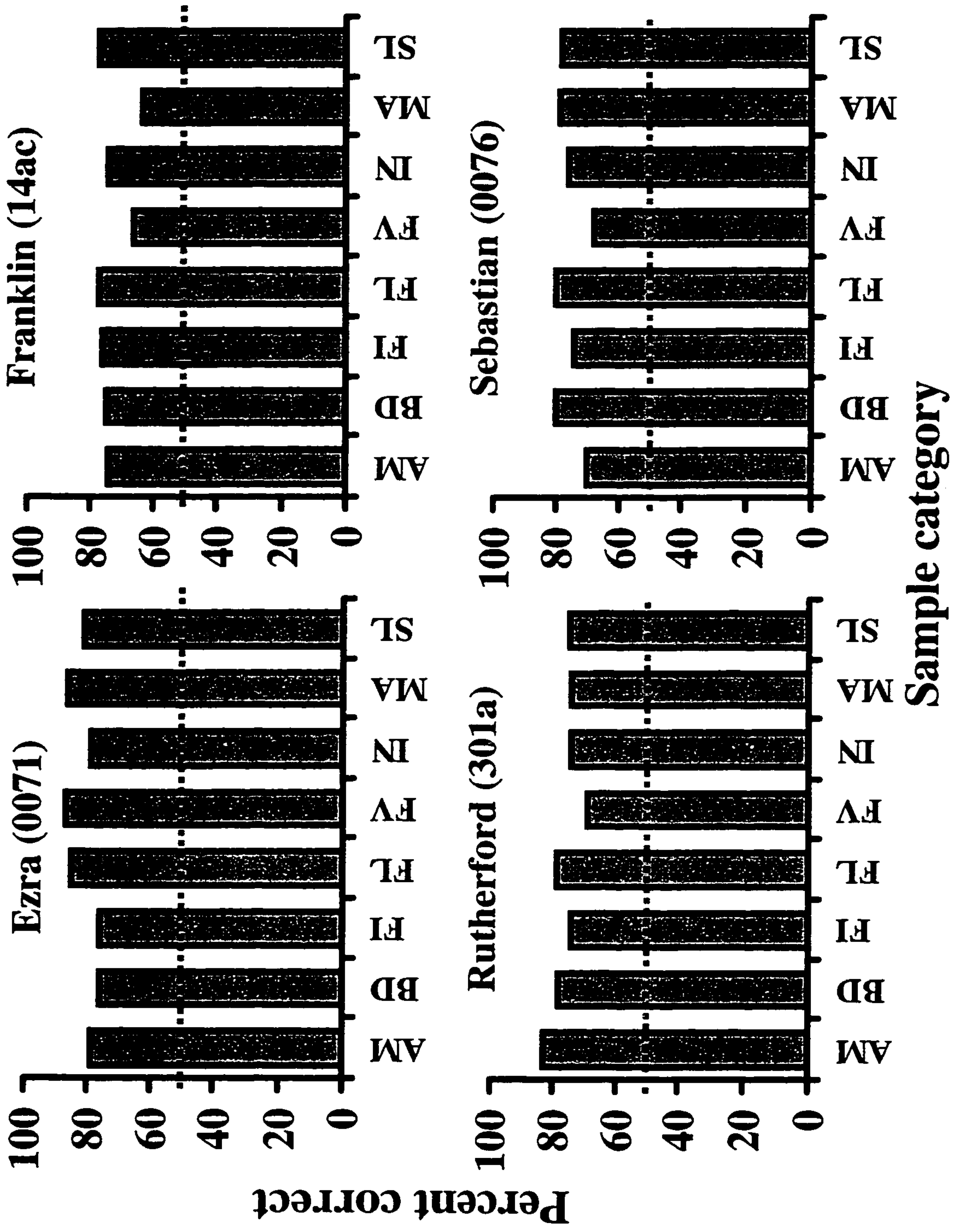


Figure 17. Overall percent correct on each category of familiar stimuli in Experiment 3 without regard for delay.

each category by session showed no significant differences between categories for any of the subjects [Ezra – *F*(7,72)=1.82; Franklin – *F*(7,72)=1.67; Rutherford – *F*(7,72)=1.34; Sebastian – *F*(7,72)=1.30].

All monkeys, however, showed a significant main effect of delay [Ezra – *F*(9,90)=4.16, *p*<0.001; Franklin – *F*(9.90)=3.92, *p*<0.001; Rutherford – *F*(9.90)=6.29, *p*<0.001; Sebastian – *F*(9,90)=4.23, *p*<0.001]. Table 6 lists the results of Tukey's post-hoc tests on delay. With few exceptions, accuracy on shorter delays was significantly different from accuracy on longer delays, which is clearly illustrated in Figure 18 (closed diamonds).

Within- and between-category performance. The middle two columns for each graph in Figure 4 show accuracy on same and different trials. A paired t-test showed that overall percent correct for trials on which the distractor was an item from the same category as the sample was not significantly different from the overall accuracy of trials on which the distractor was from a different category, *t*(3)=-0.95.

Transfer test. Due to the large number of delays, all the trials of the final set of training were used to determine whether or not the monkeys were able to transfer the identity match to novel stimuli across a wide range of delays. The right-most columns on each panel of Figure 4 show the overall percent correct for the last session of the familiar stimuli and the first session of novel stimuli. Both familiar and novel scores were all above chance for a 100 trial session (60%), but performance on novel stimuli was significantly different from performance on familiar stimuli according to a paired t-test, *t*(3)=9.13, *p*<0.05. Also, as the right-most column in each panel of Figure 5 shows,

	Familiar set		Transfer test	
Subject	Delay	Different from	Delay	Different from
Ezra	8s	56s, 64s, 80s	8s	56s, 64s, 72s, 80s
				16s 56s, 64s, 72s, 80s
				24s 56s, 64s, 72s, 80s
Franklin	8s	48s, 56s, 64s, 80s	8s	32s, 40s, 48s, 56s,
	16s	48s, 80s, 64s, 56s		64s, 72s, 80s
			16s	40s, 48s, 56s, 64s, 80s
			24s	48s, 64s, 80s
Rutherford	8s	40s, 48s, 56s, 64s, 72s, 80s	8s	32s, 40s, 48s, 56s,
	16s	40s, 64s		64s, 72s, 80s
	24s	64s		16s 40s, 56s, 64s, 72s, 80s
			24s	40s, 56s, 64s, 72s, 80s
Sebastian	16s	56s, 64s, 80s	8s	24s, 48s, 56s, 64s,
	40s	64s, 80s		72s, 80s

Table 6. Summary of post-hoc analyses on the significant main effect of delay for each subject

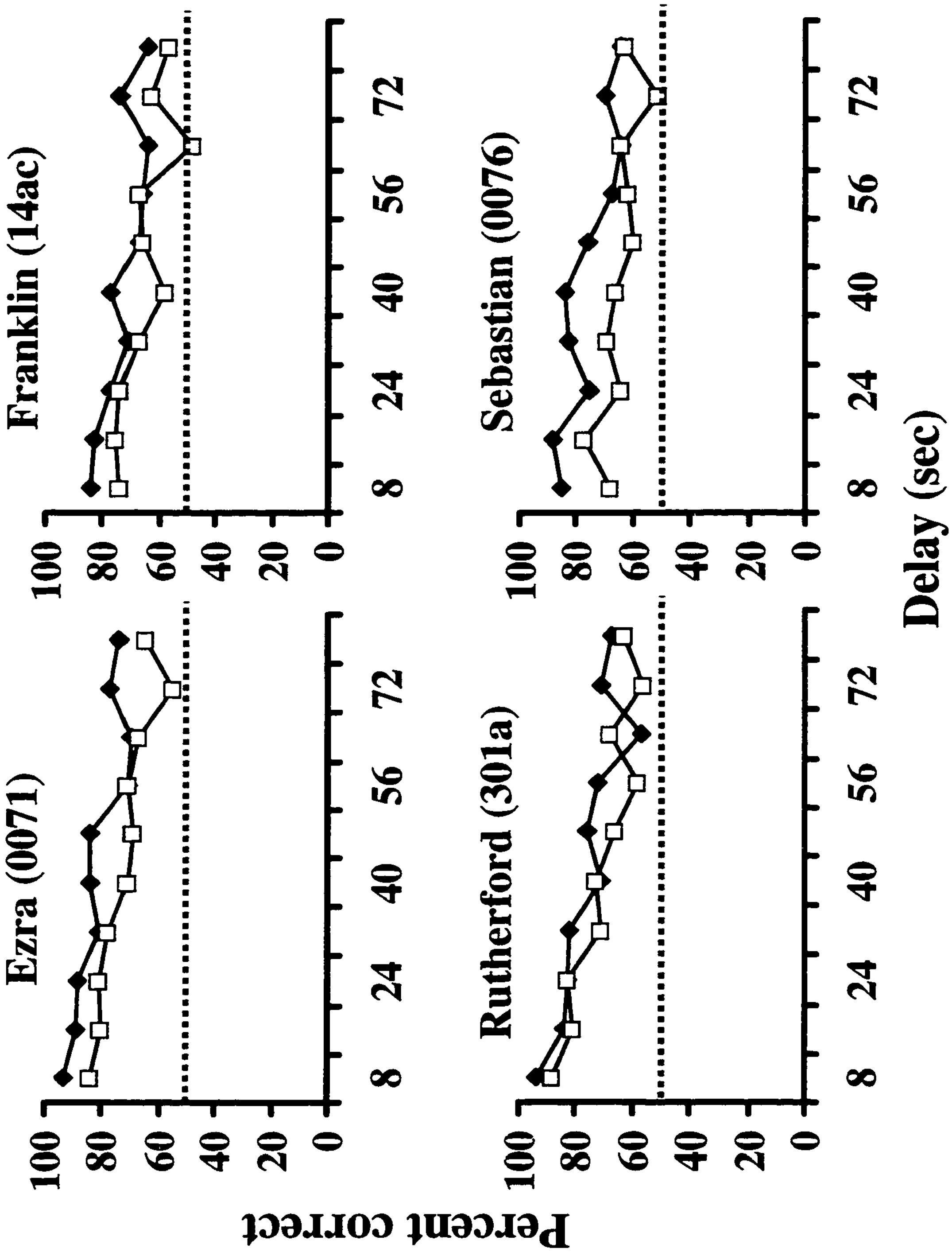


Figure 18. Overall percent correct as a function of delay for Experiment 3. The closed diamonds show performance on the familiar set of stimuli and the open squares show performance on the novel set of stimuli.

performance dropped to chance for three monkeys for the first presentation of each sample stimulus. Sebastian's scores remained similar to those of his other two transfer tests.

As with the familiar set, a chi-square analysis was used to assess the relationship between category and delay. The data for the novel set are also graphed in Figures 13, 14, 15 and 16 (open squares). As the graphs suggest, the chi-square analysis showed no significant relationship between performance on the various categories as a function of delay [Ezra – $\underline{X}^2(63)=0.83$; Franklin - $\underline{X}^2(63)=1.89$; Rutherford – $\underline{X}^2(63)=1.77$; Sebastian – $\underline{X}^2(63)=1.59$].

Figure 19 presents overall percent correct data for each subject on the familiar and novel sets. Performance on the novel set was significantly different at the $\underline{p}<0.05$ level from performance on the familiar set [Ezra – $\underline{t}(9)=8.59$; Franklin – $\underline{t}(9)=2.43$; Rutherford – $\underline{t}(9)=2.33$; Sebastian – $\underline{t}(9)=7.46$].

Two-way 2 X 8 ANOVAs on the arcsine-transformed proportions correct for set by category showed a significant interaction between set and category for all monkeys [Ezra – $\underline{F}(7,144)=2.16$, $\underline{p}<0.041$; Franklin - $\underline{F}(7,144)=3.49$, $\underline{p}<0.029$; Rutherford – $\underline{F}(7,144)=2.19$, $\underline{p}<0.038$; Sebastian – $\underline{F}(7,144)=5.08$, $\underline{p}<0.001$]. A simple effects analysis showed that Ezra performed significantly different on novel amphibians, birds, and flowers. Franklin performed significantly different on novel amphibians, fish, and sea life. Rutherford performed significantly different on novel amphibians and flowers, and Sebastian performed significantly different on novel birds, flowers, and shells. In addition, the post-hoc analysis showed significant differences in performance on various

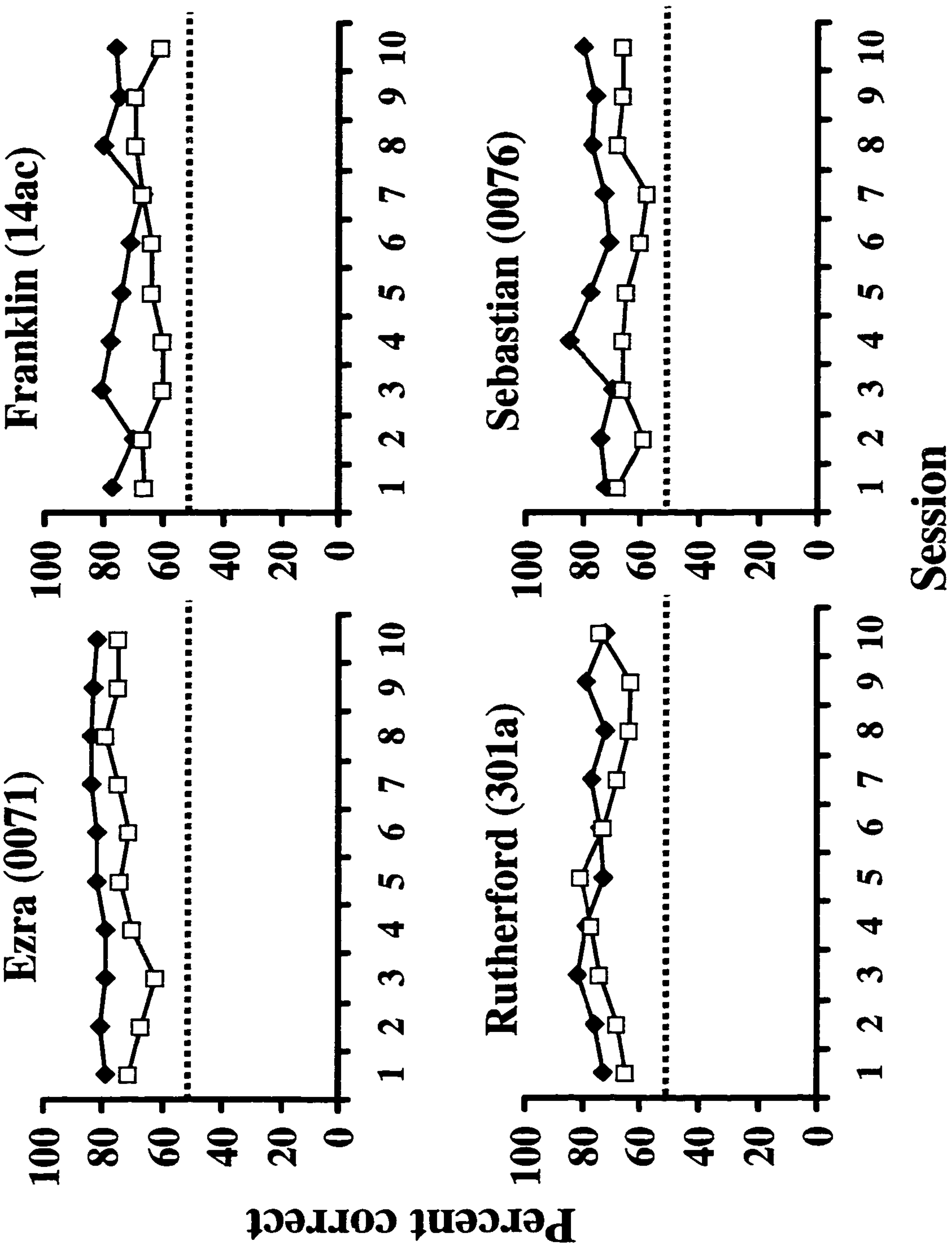


Figure 19. Overall percent correct for each session for the familiar (closed diamonds) and novel stimulus sets (open squares) for Experiment 3 on the 8s to 80s range of delays. Scores on familiar and novel stimuli were significantly different from each other for all subjects.

categories of the novel set of stimuli for Franklin and Sebastian. Figure 20 presents a side-by-side comparison of accuracy on each category for familiar and novel stimuli.

A two-way 2 X 10 ANOVA on the arcsine-transformed proportion correct for each delay as a function of stimulus set showed significant main effects of set for all monkeys [Ezra – $F(1,180)=29.18$, $p<0.001$; Franklin – $F(1,180)=11.56$, $p<0.001$; Rutherford – $F(1,180)=4.00$, $p<0.047$; Sebastian – $F(1,180)=40.34$, $p<0.001$] and delay for all subjects [Ezra – $F(7,180)=7.54$, $p<0.001$; Franklin – $F(7,180)=5.91$, $p<0.001$; Rutherford – $F(7,180)=9.88$, $p<0.001$; Sebastian – $F(7,180)=5.54$, $p<0.001$] Post-hoc analyses indicated that overall accuracy tended to decrease as a function of delay. A summary of the results of this post-hoc analysis is also presented in Table 6.

Finally, the right-hand columns in each panel of Figure 4 show the overall percent correct for same and different trials for the final set of novel stimuli tested with the 8s to 80s range. A paired t-test showed a significant difference between same and different, $t(3)=-6.34$, $p<0.05$. Additionally, when performance on same trials across single delay MTS and the familiar and novel sets of Experiment 3 were evaluated using a one-way ANOVA on the arcsine transformed proportion correct, there was a significant difference, $F(7,11)=3.50$, $p<0.05$. A similar analysis on different trials was not significant, $F(7,11)$, 0.41. Tukey's post-hoc analysis revealed that performance on same trials for the novel set of Experiment 3 was significantly lower than performance on same trials for the other two conditions.

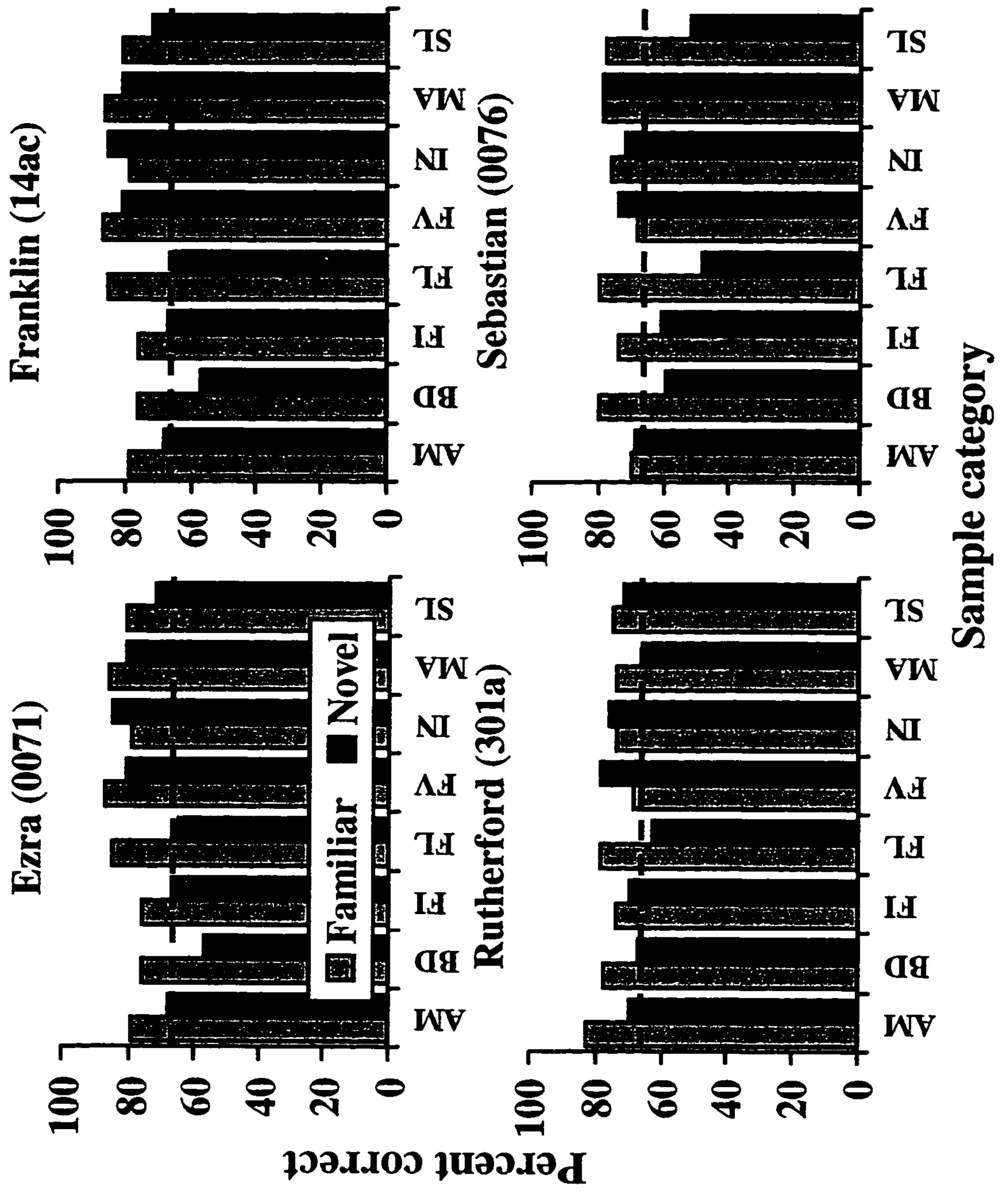


Figure 20. The results of the transfer test for Experiment 3 broken down by category. Light gray bars show percent correct for familiar stimuli and dark gray bars represent percent correct for novel stimuli. The upper dotted line represents chance based on 100 trials.

Discussion

Unlike the results of Experiment 2, both delay and stimulus novelty had significant effects on the results of this experiment, and both produced marked decreases in accuracy. The accuracy functions for the 10 delays used in this experiment produced a nearly monotonic function, suggesting that if additional, longer delays were added, a further decrement in performance would occur. This is very much in keeping with delay effects found in pigeons and monkeys under the same experimental conditions (i.e. delays vary across trials within a session). However, unlike pigeons that were able to sustain comparably short delays before performance fell to chance levels, the monkeys in this experiment were able to perform above chance for delays up to 80s, even with unfamiliar stimuli.

Given that, however, the relatively poor performance on the novel set of stimuli is of interest. The identity match was robust on MTS and DMTS of up to 20s, in that there were no significant differences between performance on familiar and novel stimuli on the transfer tests. Additionally, performance on the familiar stimuli of this experiment remained between 70% and 80%. However, performance changed significantly when novel stimuli were presented with delays ranging from 8s to 80s within a session..

In the previous experiments there was little evidence that the monkeys consistently remembered some categories better than others, but within stimulus sets certain items were more discriminable. In the present experiment, the within-category variability was decreased in the hopes of high, consistent performance on all the items of a category rather than variability across individual items. It was hypothesized that greater within-category similarity would support better transfer, particularly if the categories were categories with which the monkeys were already familiar. There is evidence that this manipulation had the desired effect. Unlike Experiments 1 and 2, there was a significant difference between trials on which the sample and distractor stimuli were from the same category as opposed to when they were from different categories. Also, all four monkeys performed poorly on the same categories in this experiment, although the poorest performance was on different individual items within a category (see Figures D1 and D2 Appendix D). The poorest performance for all four animals on the transfer test was on birds and flowers. Three of the monkeys also performed below chance on fish as well. There was a much higher degree of similarity across subjects as to what categories they performed below chance on when presented with novel stimuli.

Additionally, while there appears to be a confound between the delay, the decreased within-category variability of the stimuli, or simply the novelty of the stimuli under these conditions, and that one or all may have contributed to the decreased performance on the transfer test and the significant difference between same and different trial types, it is possible to rule all but one variable out. An analysis of performance on different trials across the three delay conditions and stimulus sets reveals no significant differences in performance. When performance on same trials is compared in a similar fashion, there is a significant difference and further analysis reveals that performance on same trials for the novel set of stimuli in Experiment 3 were significantly different than performance on same trials for the other two conditions. The novel set used in Experiment 3 was explicitly designed to limit within-category variability. These results suggest that the monkeys committed more errors because the sample and distractor stimuli on these trials were highly similar to one another.

General discussion

A few methodological issues should be addressed regarding these results. Much of the early work with the MTS paradigm looked at procedural factors that influenced acquisition of the identity concept. Evidence suggests there are several factors that influence performance on MTS and delayed MTS tasks that have nothing to do with the subject's ability to remember the sample stimulus.

In the present experiments, the duration of the sample stimulus was under the control of the subject, and across all three experiments, all subjects showed very little variability in response times to the sample stimulus, averaging approximately half a second of viewing time, even for novel stimuli where one would expect an increase. Several studies have indicated that the duration of the sample stimulus is a potentially important factor, with longer sample durations leading to more accurate performance (D'Amato and Worsham, 1972; Devine, Jones, Neville, and Sakai, 1977; Herzog, Grand, and Roberts, 1977; Blough, 1996) . This appears to be more true of naive subjects than experienced subjects, however. Two of the subjects in these experiments participated in several same-different experiments wherein the sample was presented for a fixed period of time during which the subject was not allowed to touch it. The other two subjects participated in several list-learning experiments, and were accustomed to touching stimuli when they appeared on the screen. All were well acquainted with the apparatus and were experienced subjects in that they had participated in numerous experiments. It may be the case that like D'Amato's experienced subjects, the duration of the sample stimulus was an irrelevant factor in performance, especially since the monkeys in the present experiments performed above chance even on difficult tasks.

In addition, D'Amato (1973) proposed the temporal discrimination hypothesis (THD) to account for error patterns in MTS and delayed MTS that have nothing to do with memory. This theory suggests that the temporal sequences of samples as well as the size of the stimulus set play important roles in performance. In short, the closer in time the same stimulus appears as a sample and a distractor affects performance, and that longer periods of time between presentations of the same stimulus result in better accuracy (Mason and Wilson, 1974; Worsham, 1974). Larger sets naturally contribute to this result. It is doubtful that this was a factor in the present research, since the stimulus sets, numbering either 32 or 64 images, were substantially larger than those used in these studies, which averaged between 6 and 12 items.

The results of these three experiments address several issues regarding the identity concept in delayed MTS as well as addressing non-human primate memory for complex stimuli. First, monkeys learned the identity matching concept when trained to perform a simultaneous MTS task using a set of 32 complex stimuli. Having done so, they were able to transfer the identity matching concept to a novel set of similar, complex stimuli. When a delay was introduced, the monkeys were able to maintain performance, and showed no decrement in accuracy across increasing delays. Performance remained at baseline levels when a transfer test with similar novel stimuli was performed. Finally, when presented with multiple delays in a session, overall accuracy decreased as a function of delay. The monkeys were able to transfer the identity matching concept under the multiple delay within a session condition though performance on the novel set also showed the same decrement in performance as a function of delay.

These results are interesting when taken within the context of previous research, particularly research done by Roberts and Mazmanian (1988), and by Wasserman and associates (1988, 1992). As mentioned previously, Wasserman, Kiedinger, and Bhatt (1988) found that pigeons, when forced to make a choice between two items, tended to make more errors when items were from the same category than when they were from different categories. That was not the case in the present experiments, until the final set of stimuli. The category of the distractor had no effect on accuracy on sample categories at all. In that respect, these results agree with Roberts and Mazmanian's (1988) finding that their non-human subjects learned nothing about the individual species contained in the comparison stimuli when learning to discriminate between birds and a variety of animals. Additionally, the specific pairing of stimuli in this experiment had a significant effect on accuracy in much the same way it had an effect in the results found by Devine et al. (1977) and Devine, Burke, and Rohack (1979). Though the type of stimuli they used was markedly different than those used here, they also found a decrease in percent correct when stimuli from a particular category were paired together, particularly when those stimuli were very similar in appearance (i.e. black and white shapes).

Given the differences in the nature of the stimuli between those experiments and the ones presented here, it is clear that very little in the way of categorization was occurring, and that items that humans would place into particular categories were not grouped together. In the same there was no reason for the subjects to learn anything about the species of the negative stimuli in Roberts and Mazmanian (1988), there was no reason for the subjects in the present set of experiments to learn anything about the stimuli as far as which category it belonged to was concerned. Given that the category of the distractor stimulus had little effect on performance until the last set of stimuli, and the amounts of variability in performance on individual items composing a category even in Experiment 3, it is likely the monkeys were attending to physical features of the stimuli in order to perform the task rather than displaying conceptually mediated behavior.

The MTS paradigm could be used to train categorization of stimuli or equivalence classes via differential reinforcement, though Herrnstein (1990) claims that this procedure does not produce categorization as much as it reveals previously formed concepts. These experiments often use MTS or DMTS, and associate either food and no food, a particular food type, or a different ratio of reinforcement with particular stimuli (Edwards, et al, 1982; Urcuioli and DeMarse, 1994; Urcuioli, DeMarse, and Lionello, 1998). In a variation of this procedure, different delays of reinforcement could be associated with particular stimuli as well (Astley and Wasserman, unpublished manuscript). It would be of interest to see whether or not associating particular categories of stimuli with particular delays would produce any differences in responding to particular categories of stimuli, since there is evidence that pigeons can learn the DMTS task when particular delays are associated with particular stimuli (Perkins, Lydersen, and Beaman, 1973).

An examination of performance on categories for all three transfer tests used in the present studies, however, offers some suggestion that, according to Herrnstein's argument, the monkeys may have come into these experiments with pre-formed concepts. It was noted in the discussion of Experiment 1 that all four animals did particularly well on the transfer tests for fruits and vegetables and mammal faces. For all three experiments all four monkeys continued to perform consistently well on fruits and vegetables regardless of the delay, stimulus novelty, or within-category variability. Similarly, three of the monkeys performed consistently well on insects, and on mammal faces (though not the same three in both cases). As previously mentioned, of the categories used in this study, fruits and vegetables and mammal faces are the two with which the monkeys would have had some real-world experience.

Since the main focus of the present experiments was to assess the effects of delay on acquisition and transfer of the identity concept, another experiment that could be run would involve performing a series of transfer tests, each with a novel set of stimuli. The delay condition would be counterbalanced across subjects and across days so that on one test day one monkey might receive a simultaneous transfer test, followed by a multiple delay transfer test the next day. Another monkey would receive the same tests but in a different order. This would address potential order-effects confounds in the present research, in that all four subjects progressed from a simultaneous MTS, to a single delay MTS, to a multiple delay MTS, with substantial acquisition training occurring prior to the first two delay conditions. Since the four subjects used in these experiments are now trained on a variety of MTS and DMTS procedures, this would be a logical next step.

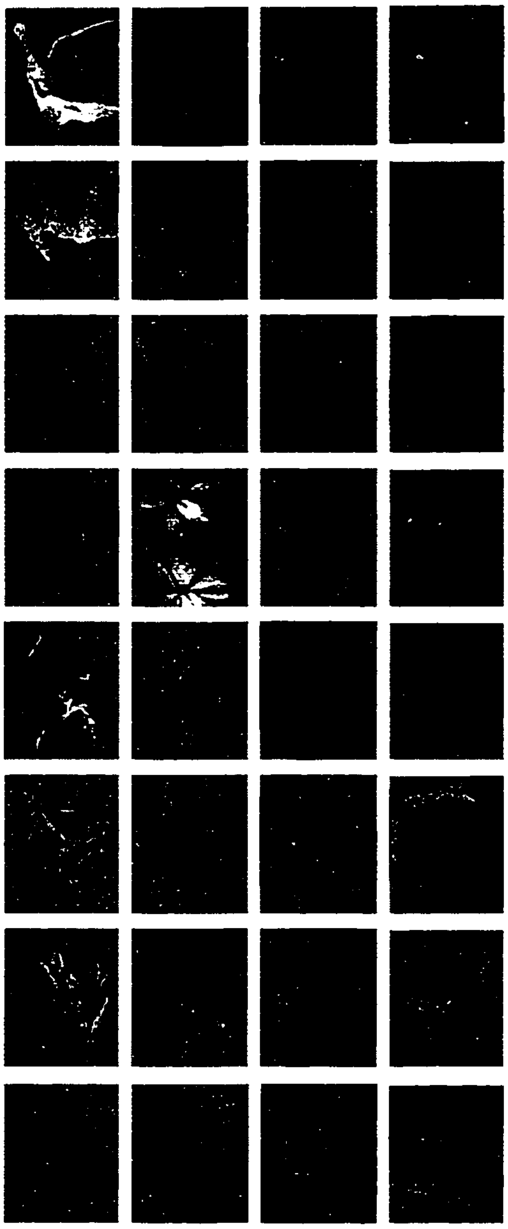
Similarly, conducting the same set of studies using carefully manipulated kaleidoscope images may shed some light on what aspects of the stimulus the monkeys were using to perform the task. Since kaleidoscope images can be created and the proportion of colors, angles, and overall complexity could potentially be quantified, systematically manipulating these features and examining acquisition as well as how discriminable particular pairs of stimuli are could provide valuable insight into what information the monkeys extract from the stimulus in order to perform the task.

Finally, the matching-to-sample procedure was chosen for these experiments as a precursor to future research questions addressing issues in social perception. First, the identity match will be used to investigate memory for species as opposed to memory for individuals in a delayed MTS task. In the species task samples of a particular sex and species will be paired with distractors of the same sex but a different species. In the individual task, samples of a particular sex and species will be paired with distractors of the same sex and species but pictures of different individuals will be used. Delays will be varied within a session. One question of importance in a study such as this is whether or not it is easier to remember and discriminate between different species, or between individuals within a species. A social cognition perspective would predict that since macaque monkeys live in social groups, they should remember and discriminate between individuals of their own species with a higher degree of accuracy than for other species. However, since it would be possible for the monkeys to be matching on the basis of individuals in either the species or individual task, a non-identical MTS task using the same variables would need to be conducted as well.

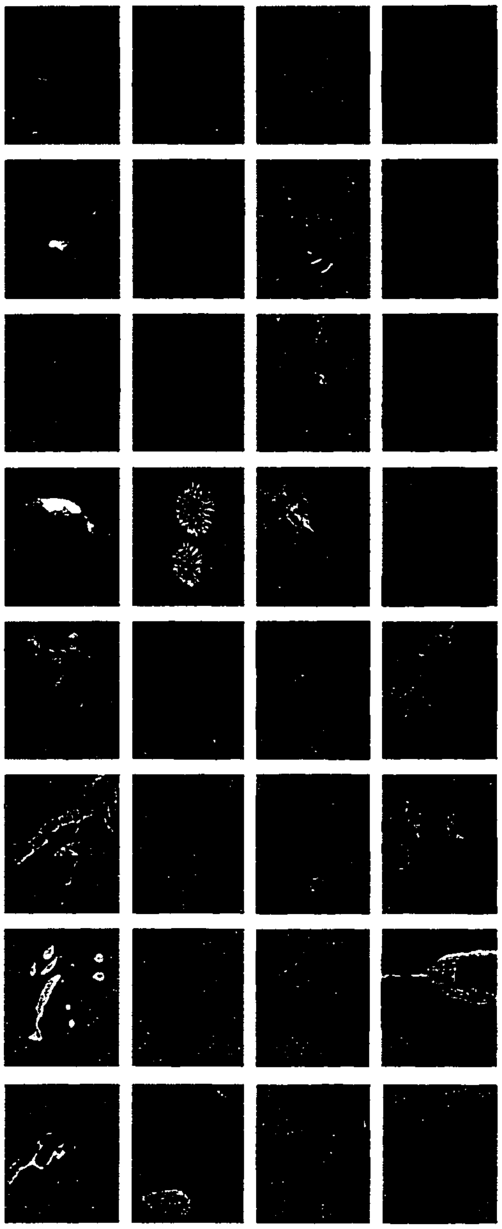
Additionally, should it be possible to train a non-identical match, and there is some indication that it might be (Neiworth and Wright, 1994; Dasser, 1987) the experiments with the MTS procedure could make use of the possibility of presenting the monkeys with "ambiguous" trials. For example, in an experiment looking at what features of a stimulus that contains social information might be controlling responding, a male rhesus monkey would be presented as the sample. The comparison stimuli in an experiment such as this would be a male from another species, and a female rhesus monkey. Each comparison is a correct categorical match to the sample, and patterns of responding over trials would provide insight into what social features are relevant to the monkeys. The delay could also figure into this as well to provide information about the cognitive world of the monkey.

Appendix A: Stimuli

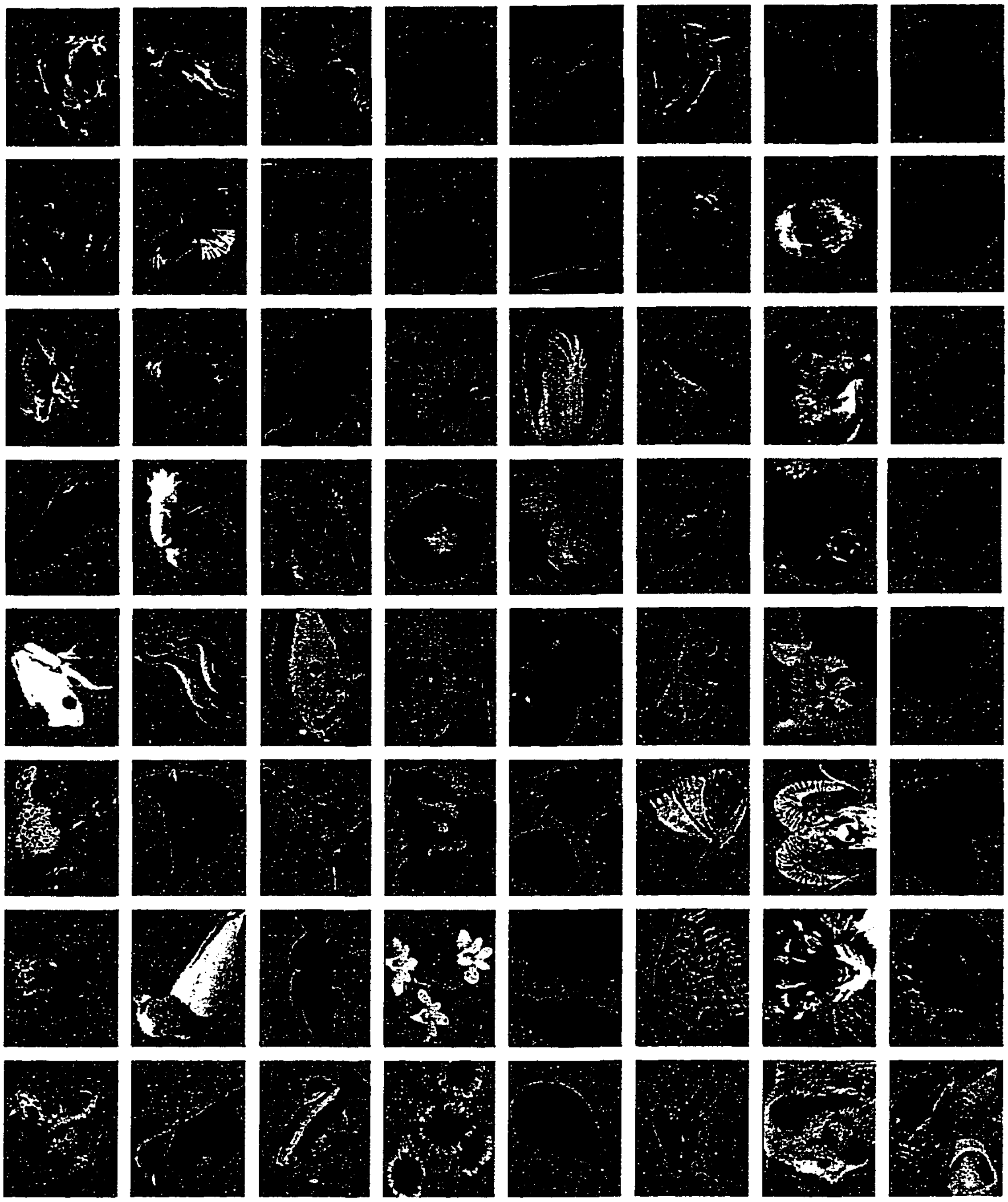
Stimulus sets I, II, III, IV and V are presented here. Sets I – IV were used in Experiment 1, sets III and IV were used in Experiment 2, and sets III – V were used in Experiment 3 (see Table 1). 16-bit color versions of the images were used in the experiments.



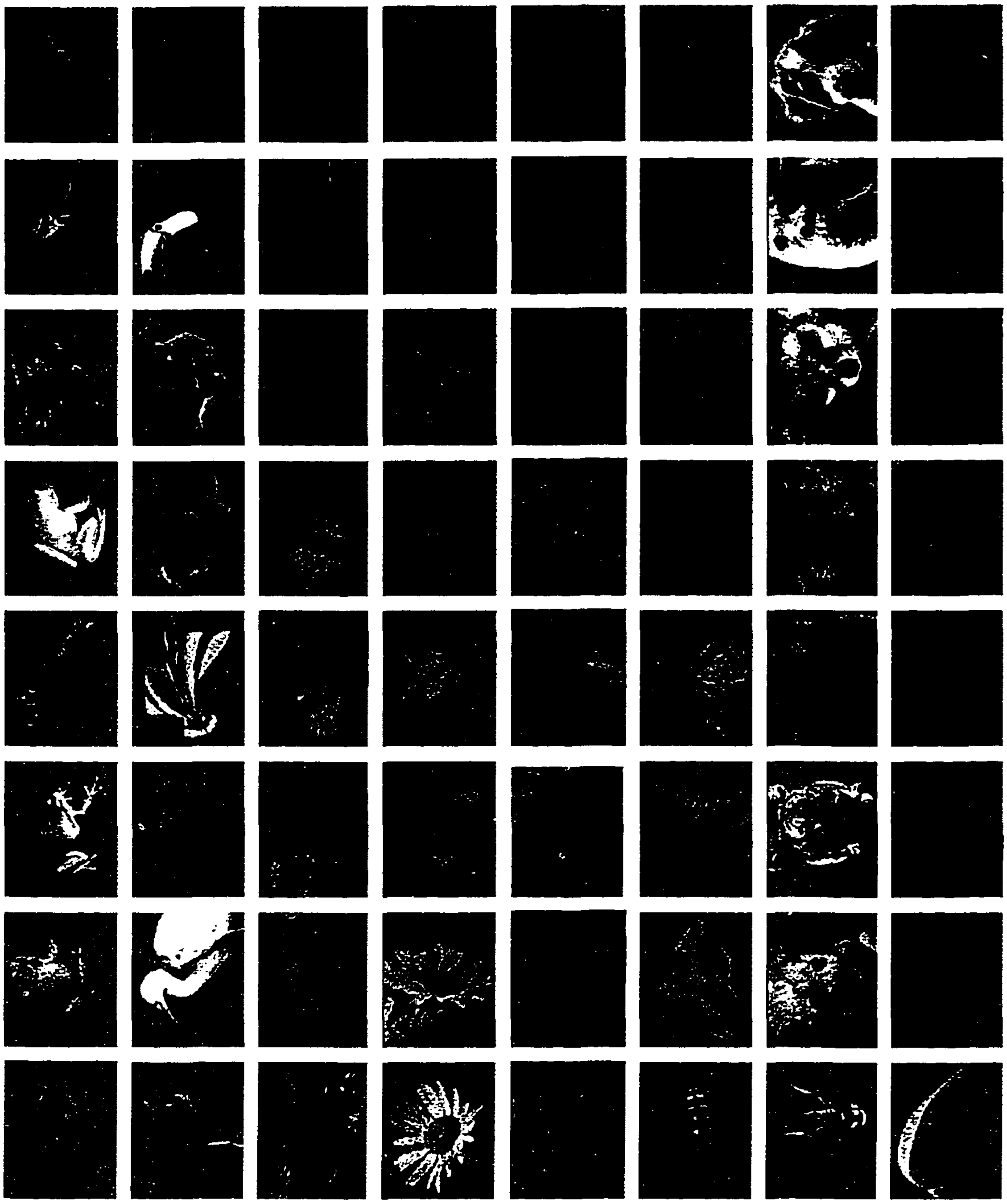
Set I



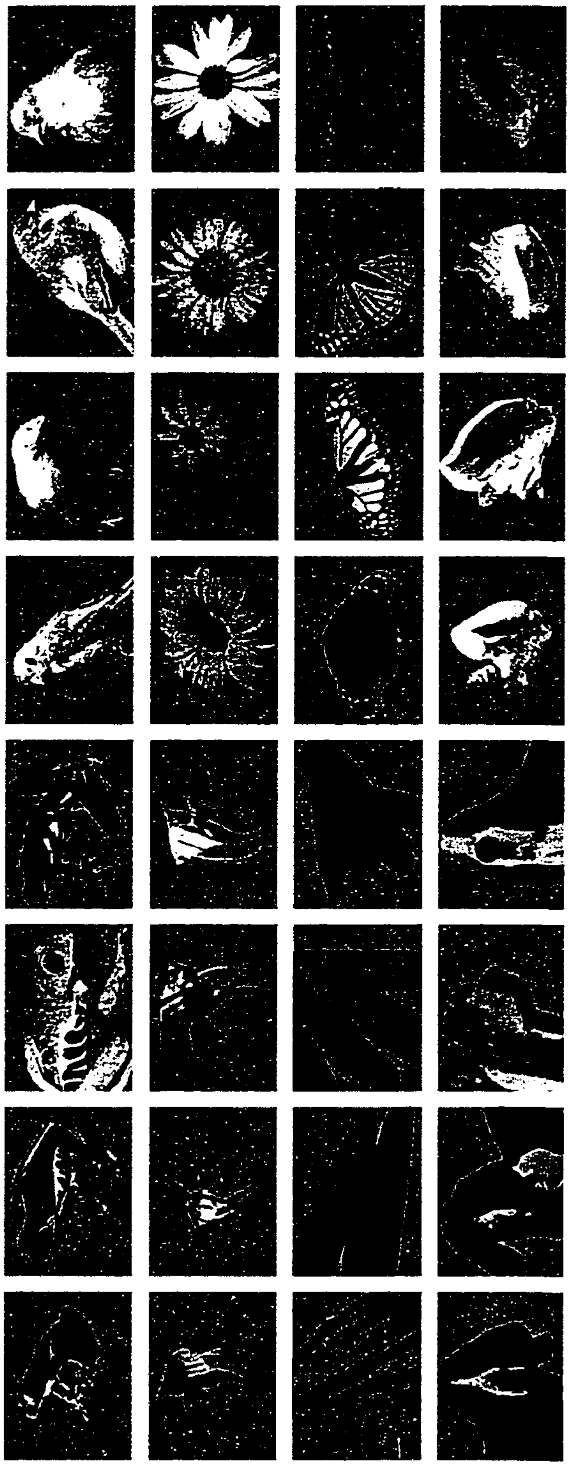
Set II



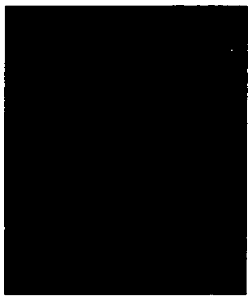
Set III



Set IV



Set V



Start stimulus

Appendix B: Performance on individual items in Experiment 1

Additional information about performance on individual items during acquisition and on the transfer test is presented in Appendix B.

LIST OF APPENDIX B FIGURES

Page

Figure B1. Performance on each stimulus for each category across all four blocks of acquisition for Ezra (0071)........88

Figure B2. Performance on each stimulus for each category across all four blocks of acquisition for Franklin (14ac)........89

Figure B3. Performance on each stimulus for each category across all four blocks of acquisition for Rutherford (301a)........90

Figure B4. Performance on each stimulus for each category across all four blocks of acquisition for Sebastian (0076)........91

Figure B5. Overall percent correct on individual stimuli in the fourth block of training. The eight different patterns correspond to the eight categories of stimuli used in the experiment........92

Figure B6. Ezra's (0071) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. The dashed line designates chance........93

Figure B7. Franklin's (14ac) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. The dashed line designates chance........94

Figure B8. Rutherford's (301a) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. The dashed line designates chance........95

Figure B9. Sebastian's (0076) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. The dashed line designates chance........96

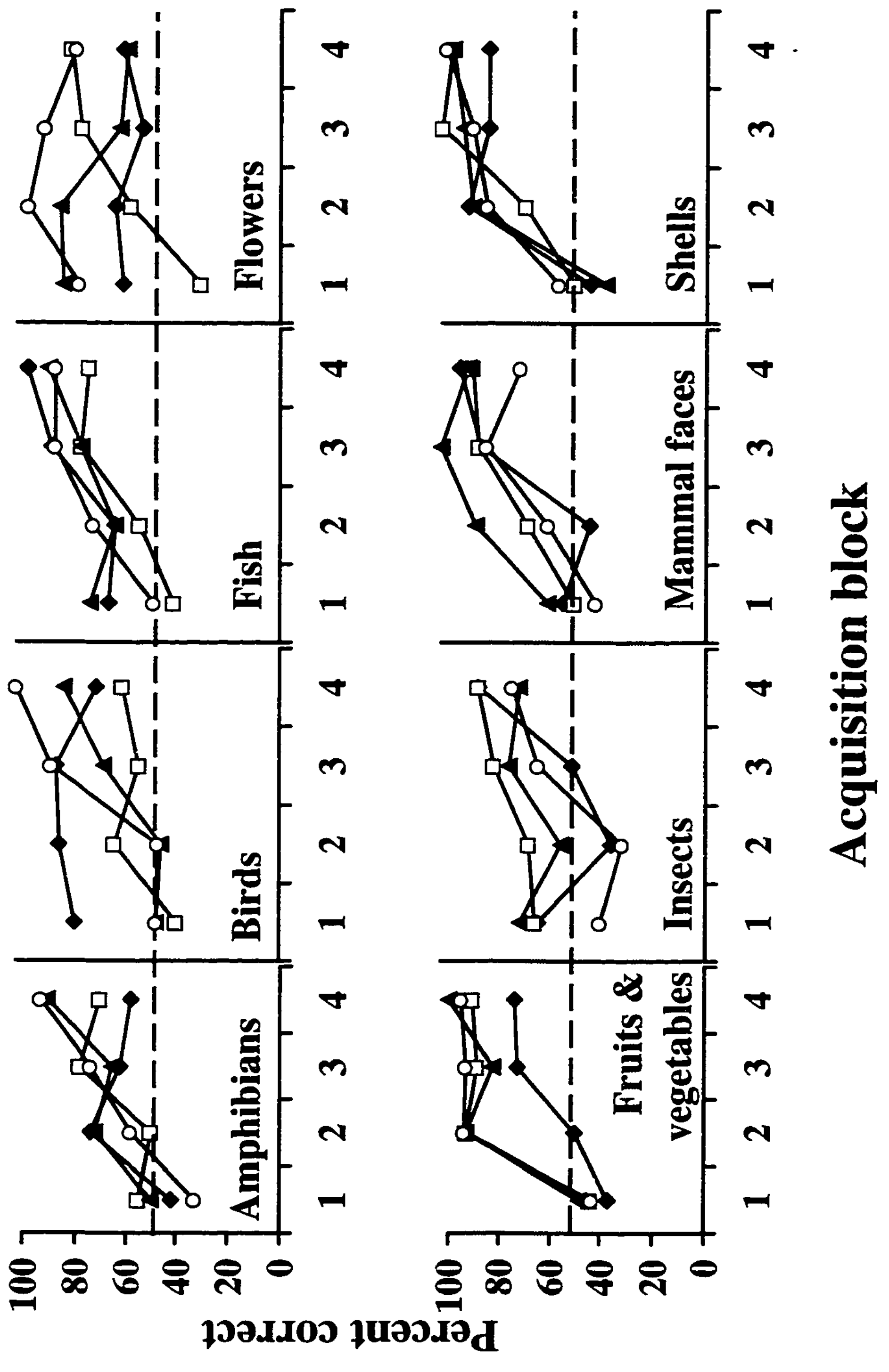


Figure B1. Performance on each stimulus for each category across all four blocks of acquisition for Ezra (0071).

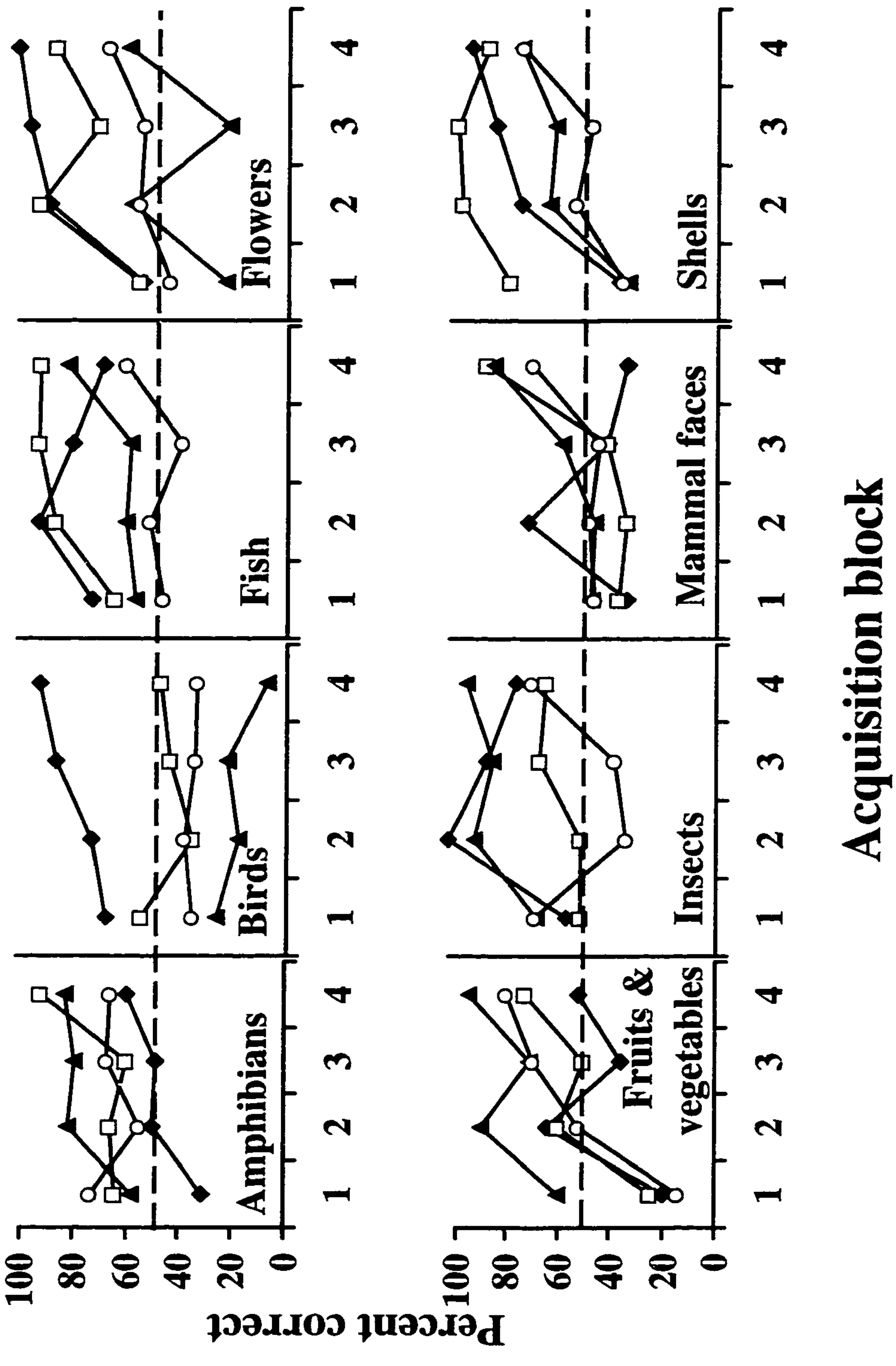


Figure B2. Performance on each stimulus for each category across all four blocks of acquisition for Franklin (14ac).

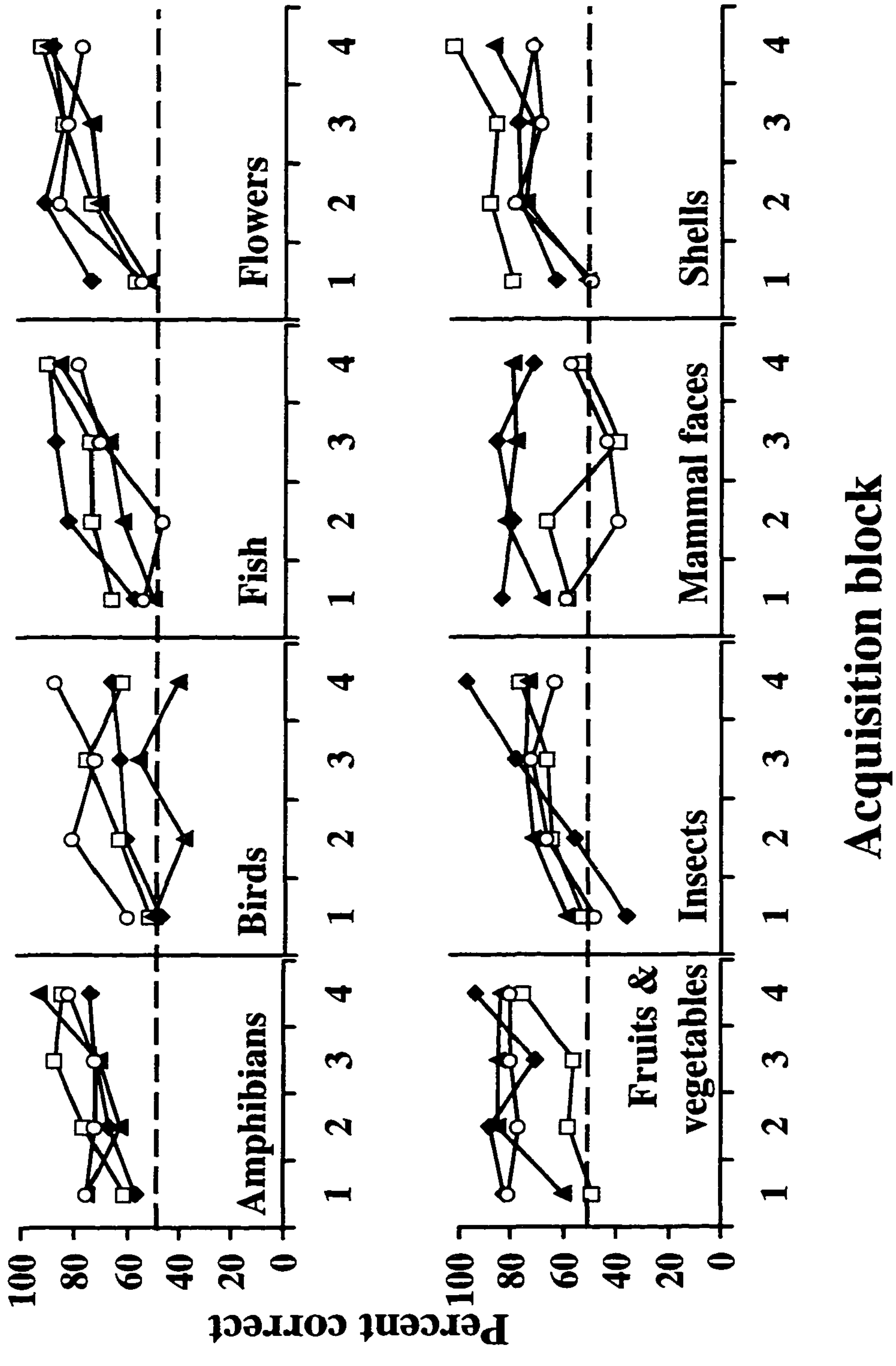


Figure B3. Performance on each stimulus for each category across all four blocks of acquisition for Rutherford (301a).

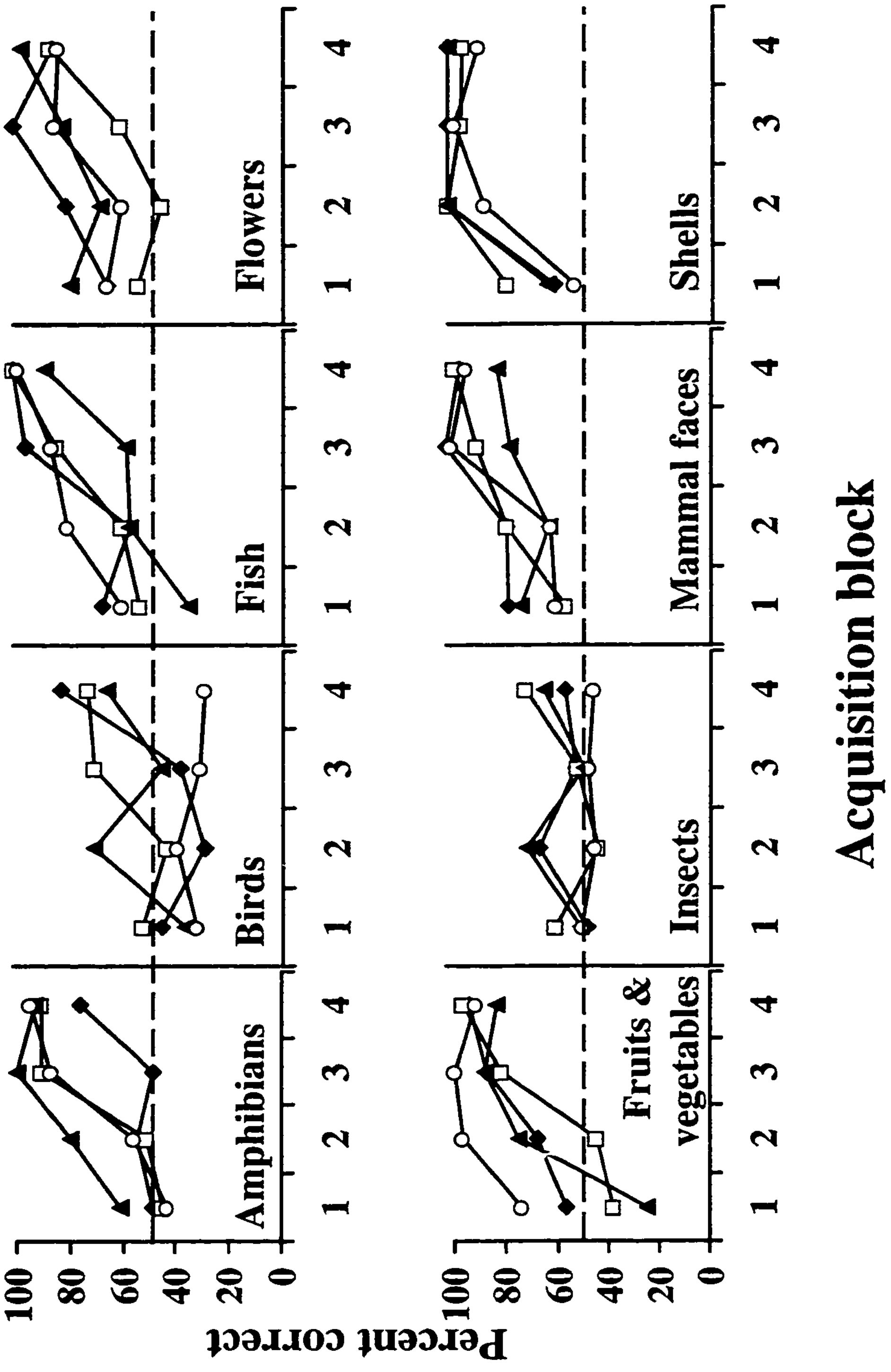


Figure B4. Performance on each stimulus for each category across all four blocks of acquisition for Sebastian (0076).

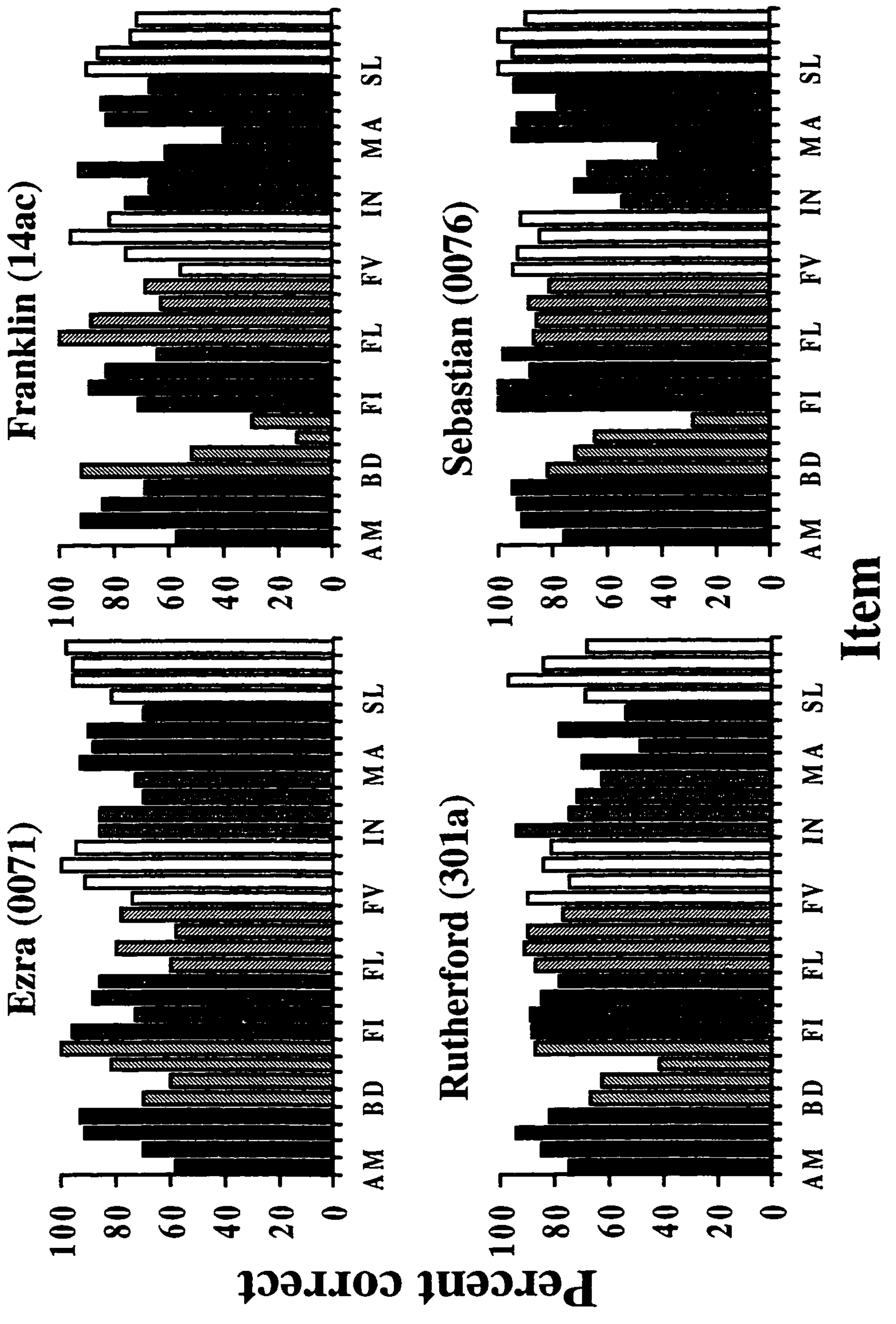


Figure B5. Overall percent correct on individual stimuli in the fourth block of training. The eight different patterns correspond to the eight categories of stimuli used in the experiment.

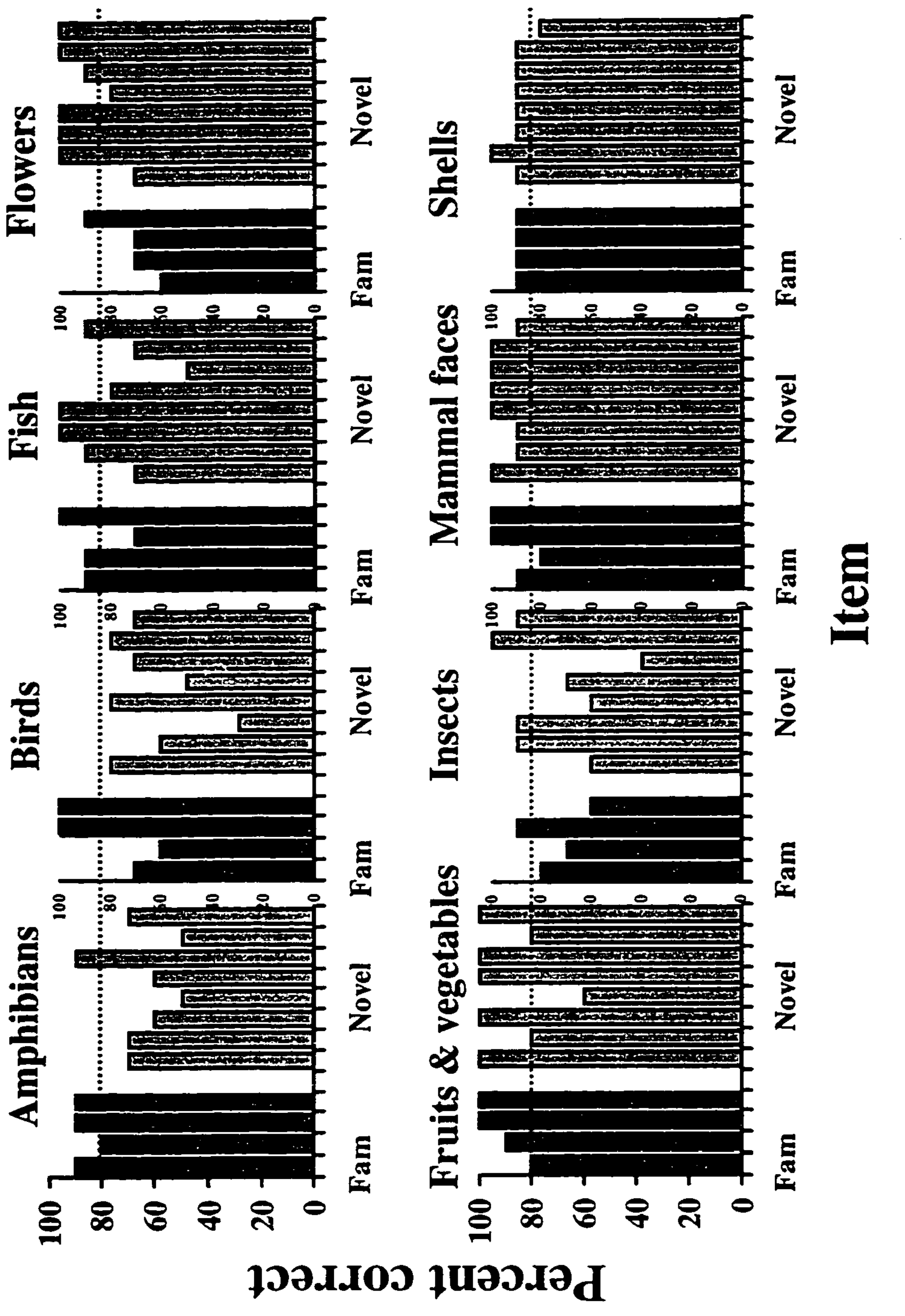


Figure B6. Ezra's (0071) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. The dashed line designates chance.

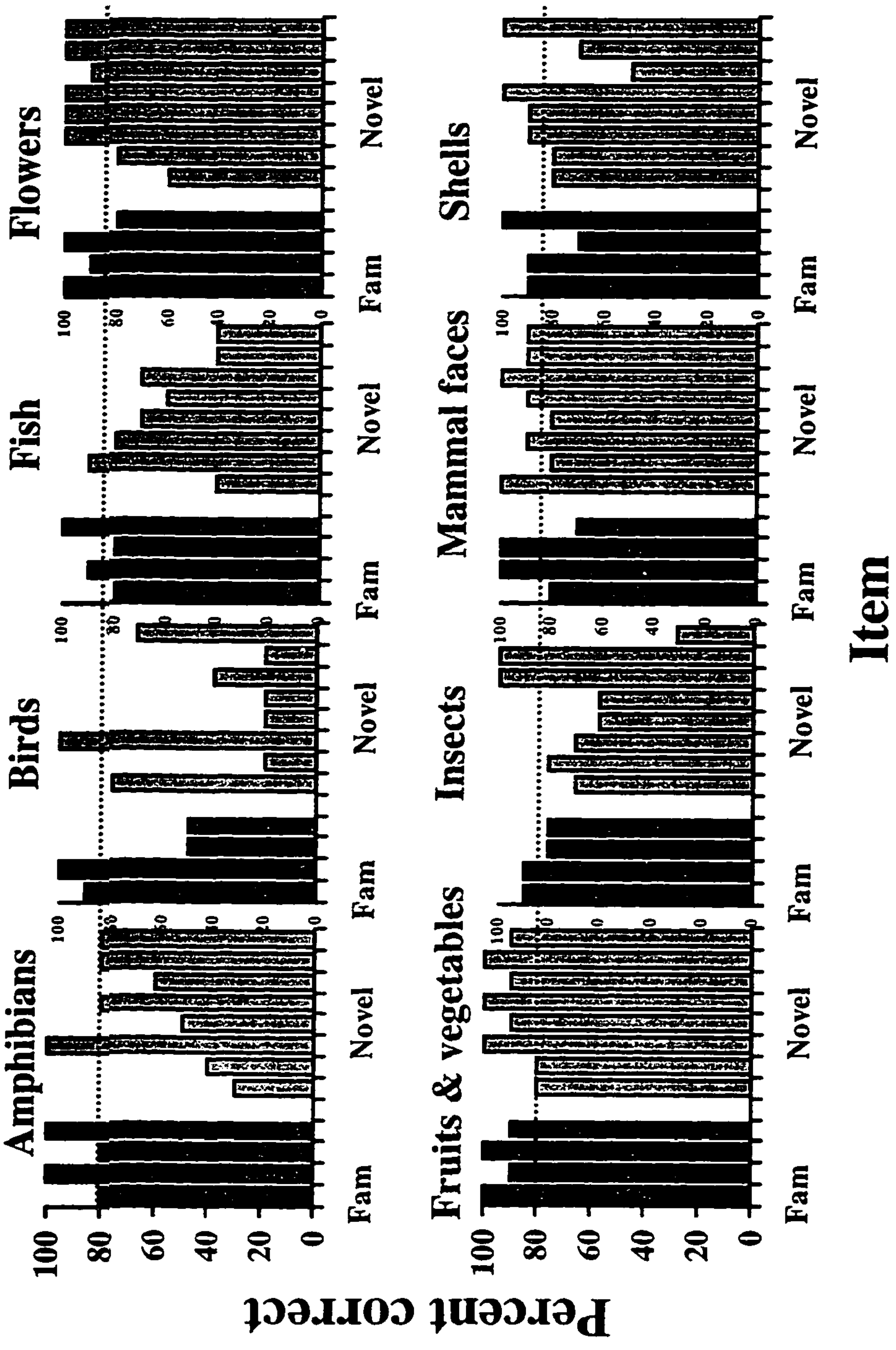


Figure B7. Franklin's (14ac) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. The dashed line designates chance.

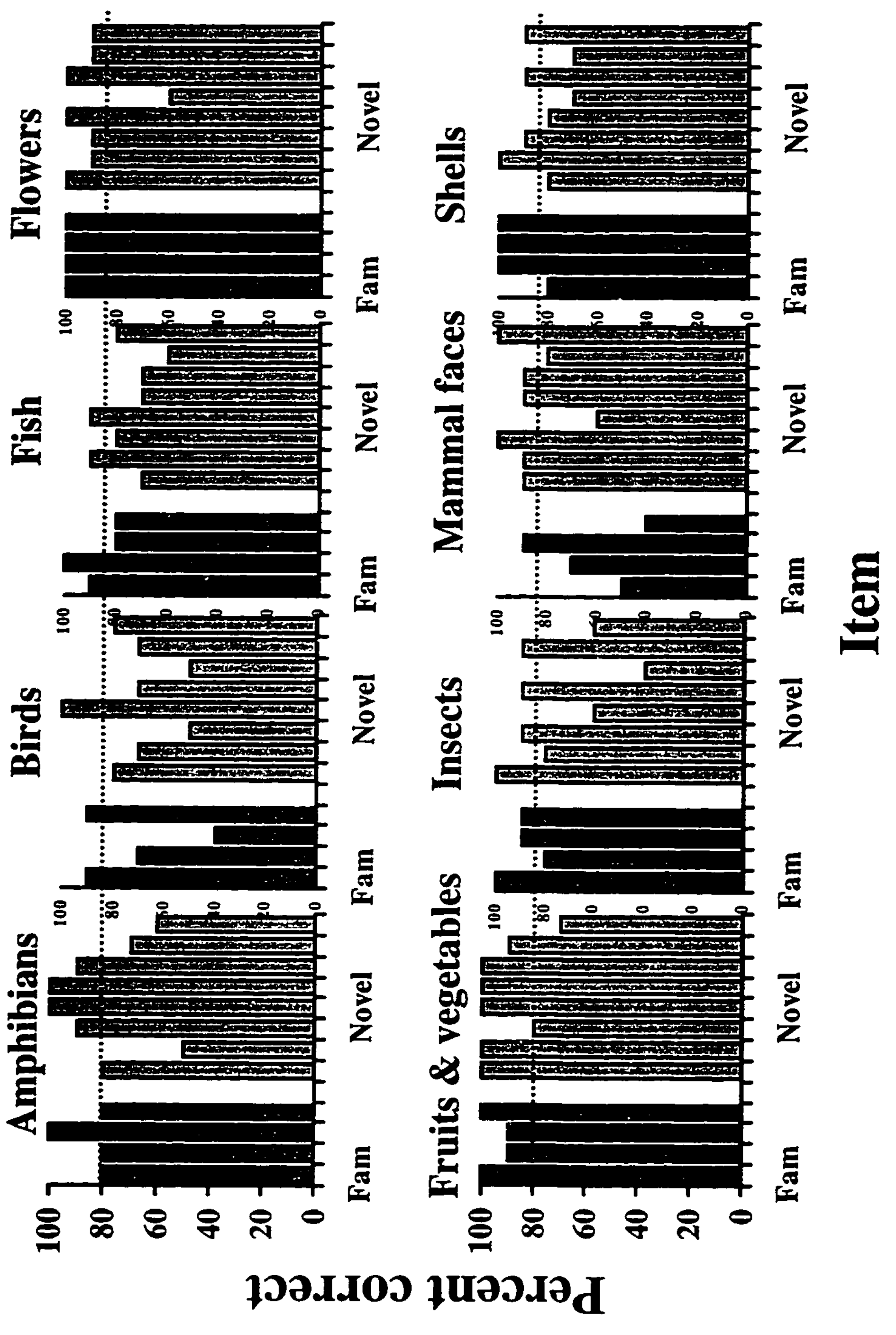


Figure B8. Rutherford's (301a) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. The dashed line designates chance.

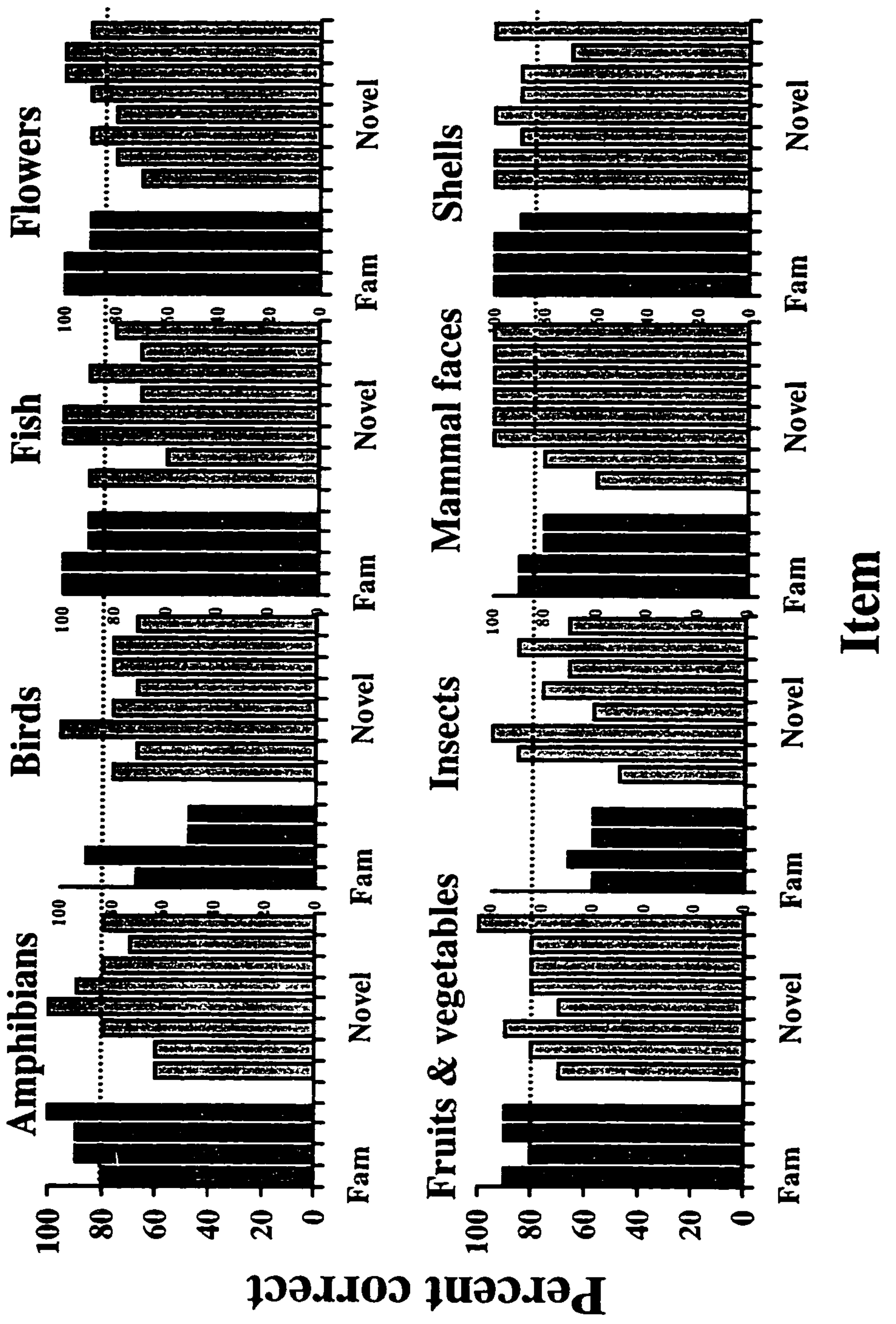


Figure B9. Sebastian's (0076) percent correct for individual items for the transfer test. Ten trials of each stimulus were used to calculate percents correct. Γhe dashed line designates chance.

Appendix C: Performance on individual items in Experiment 2

Additional information for Experiment 2. Table C1 shows the slopes and y-intercepts for acquisition of each delay. The figures show the mean score on each delay for each category with error bars showing the range of extreme scores for items within the category.

LIST OF APPENDIX C TABLES AND FIGURES

Page

Table C1. Slopes and y-intercepts for the overall acquisition of each delay. * indicates significant difference at the $p<0.05$ level....................................98

Figure C1. The average percent correct for the stimuli in each delay for Ezra (0071). The error bars indicate the minimum and maximum percent correct for items in that category..99

Figure C2. The average percent correct for the stimuli in each delay for Franklin (14ac). The error bars indicate the minimum and maximum percent correct for items in that category..100

Figure C3. The average percent correct for the stimuli in each delay for Rutherford (301a). The error bars indicate the minimum and maximum percent correct for items in that category...........................101

Figure C4. The average percent correct for the stimuli in each delay for Sebastian (0076). The error bars indicate the minimum and maximum percent correct for items in that category..102

Figure C5. Overall percent correct on individual stimuli on criterion sessions for the 20s delay for Ezra, Franklin, and Rutherford, and for the 16s delay for Sebastian..103

	Ezra		Franklin		Rutherford		Sebastian			
Delay	**slope**	**y-int**	**slope**	**y-int**	**slope**	**y-int**	**slope**	**y-int**	**t(slope)**	**t(Y-int)**
Sim	1.41	77	1.88	83	0.04	74	3.08	72	2.21	15.13
0s	2.79	77	1.77	73	0.6	80	7.03	73	3.54 *	9.95
1s	0.61	72	2.53	72	4.06	77	3.13	83	0.13	6.06
2s	1.05	73	3.59	76	-4.69	95	0.99	77	1.59	7.77
4s	0.31	76	0.84	83	6.09	69	1.25	85	1.97	9.00
8s	0.16	81	0.04	82	1.56	85	1.36	70	1.55	7.46
12s	0.07	78	0.00	86	1.48	68	0.57	78	1.17	5.74
16s	0.08	77	1.06	63	9.38	78	0.04	84	1.04	18.35

Table C1. Slopes and y-intercepts for the overall acquisition of each delay. * indicates significant difference at the $p<0.05$ level.

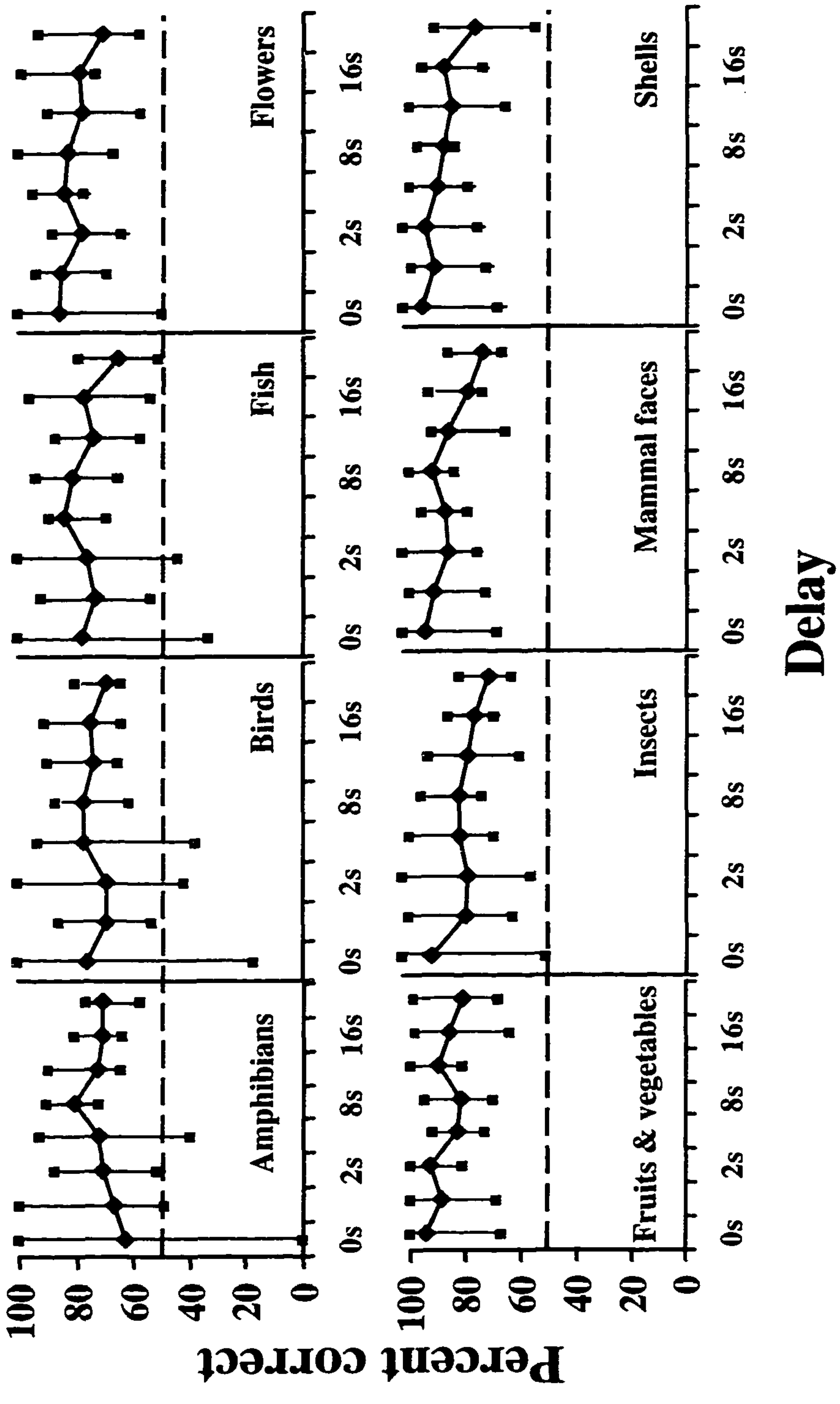


Figure C1. The average percent correct for the stimuli in each delay for Ezra (0071). The error bars indicate the minimum and maximum percent correct for items in that category.

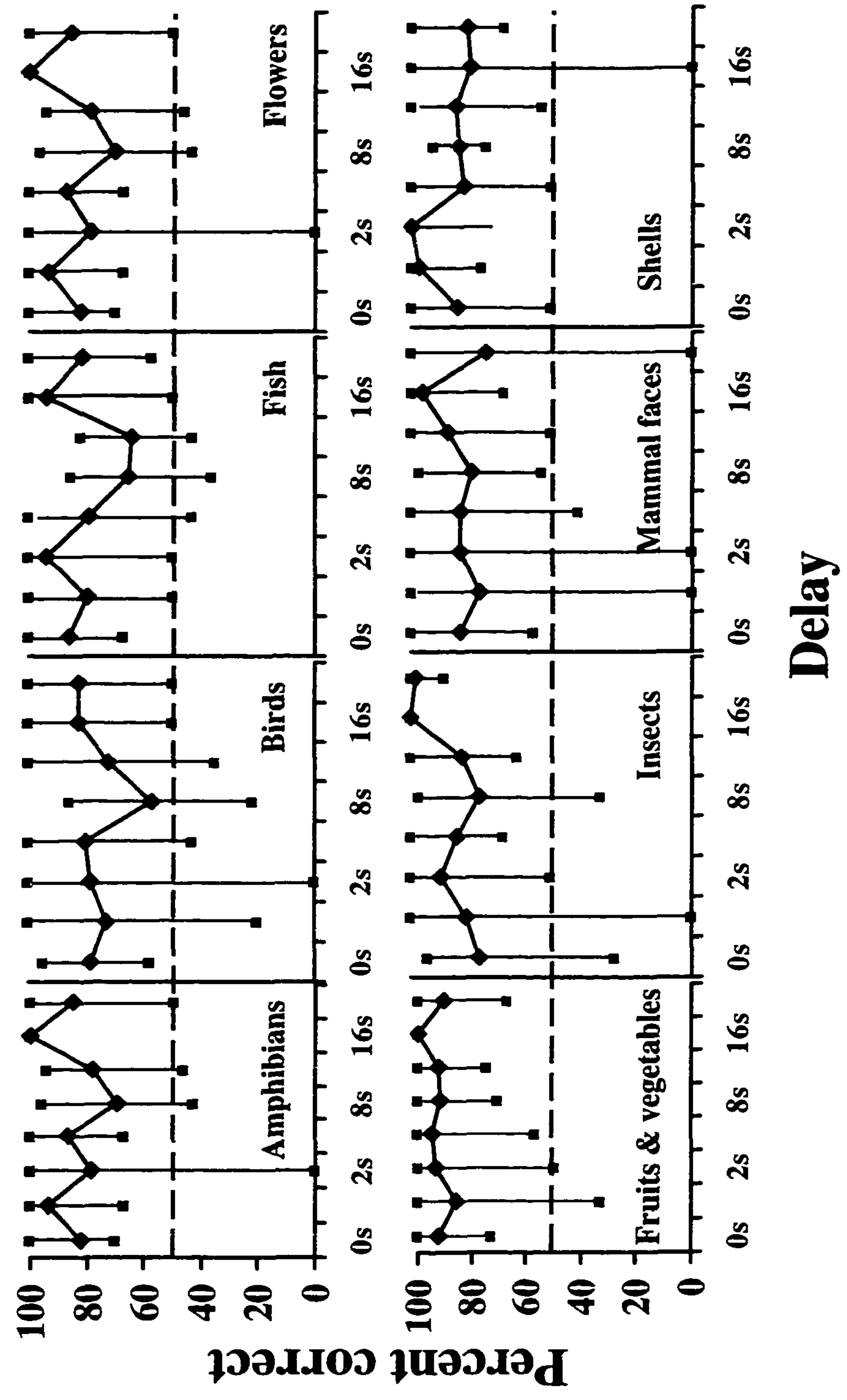


Figure C2. The average percent correct for the stimuli in each delay for Franklin (14ac). The error bars indicate the minimum and maximum percent correct for items in that category.

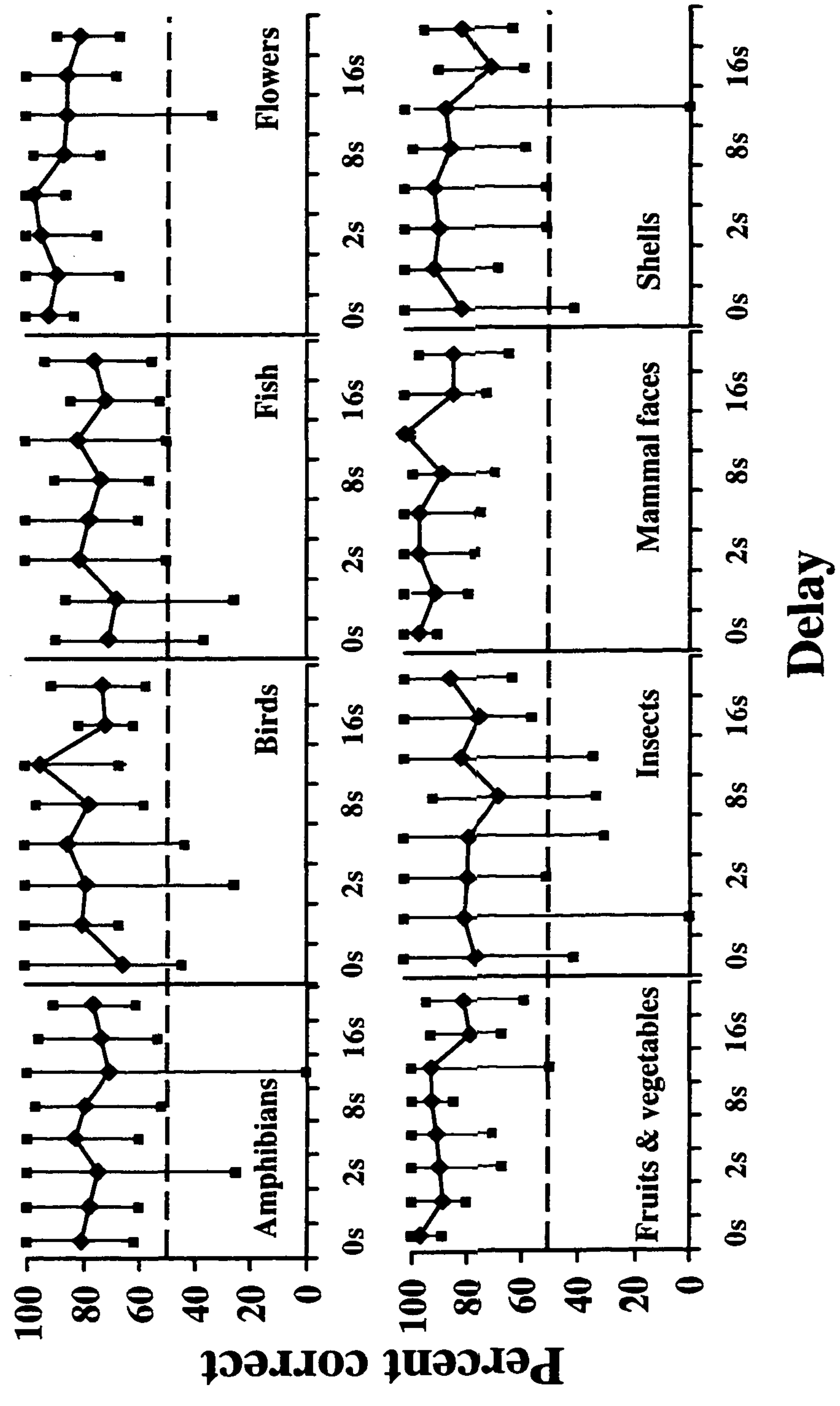


Figure C3. The average percent correct for the stimuli in each delay for Rutherford (301a). The error bars indicate the minimum and maximum percent correct for items in that category.

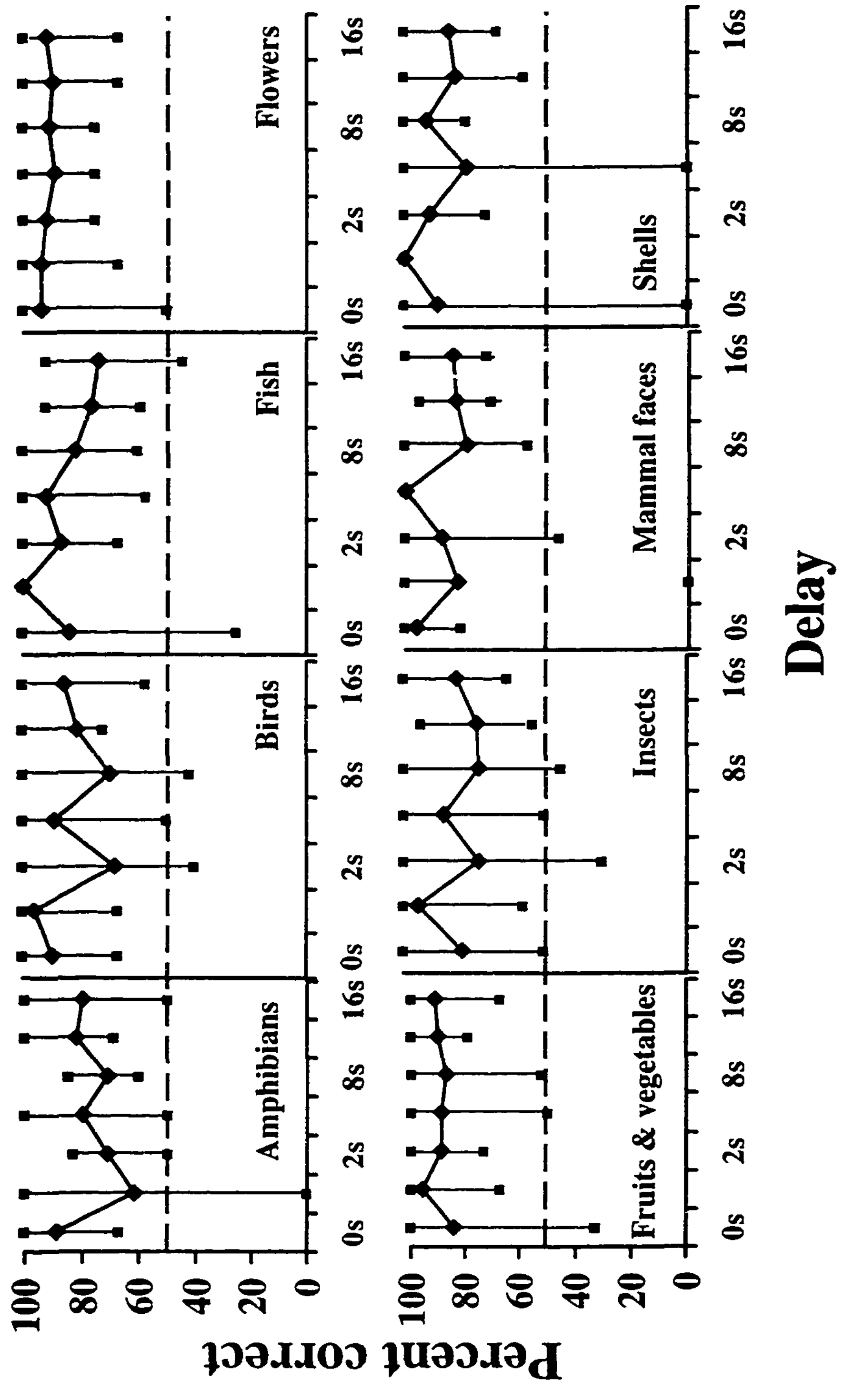


Figure C4. The average percent correct for the stimuli in each delay for Sebastian (0076). The error bars indicate the minimum and maximum percent correct for items in that category.

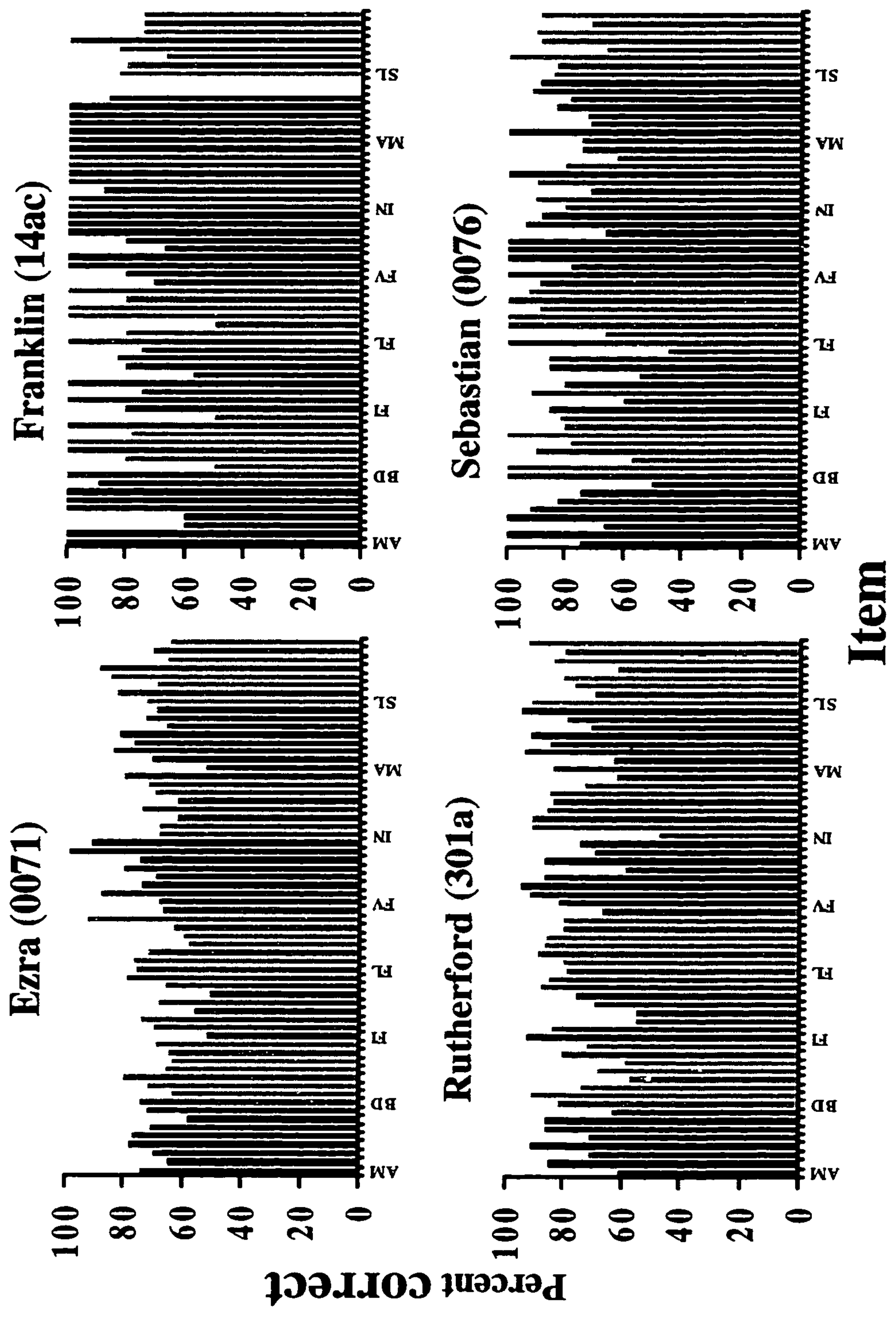


Figure C5. Overall percent correct on individual stimuli on criterion sessions for the 20s delay for Ezra, Franklin, and Rutherford, and for the 16s delay for Sebastian.

Appendix D: Performance on individual items in Experiment 3

Additional individual information for Experiment 3. The two figures show overall accuracy on individual items for familiar and novel stimuli for each subject. Overall scores were calculated without regard for delay.

LIST OF APPENDIX D FIGURES

Page

Figure D1. Overall item accuracy for familiar stimuli presented in Experiment 3.......105

Figure D2. Overall item accuracy for novel stimuli presented in Experiment 3..........106

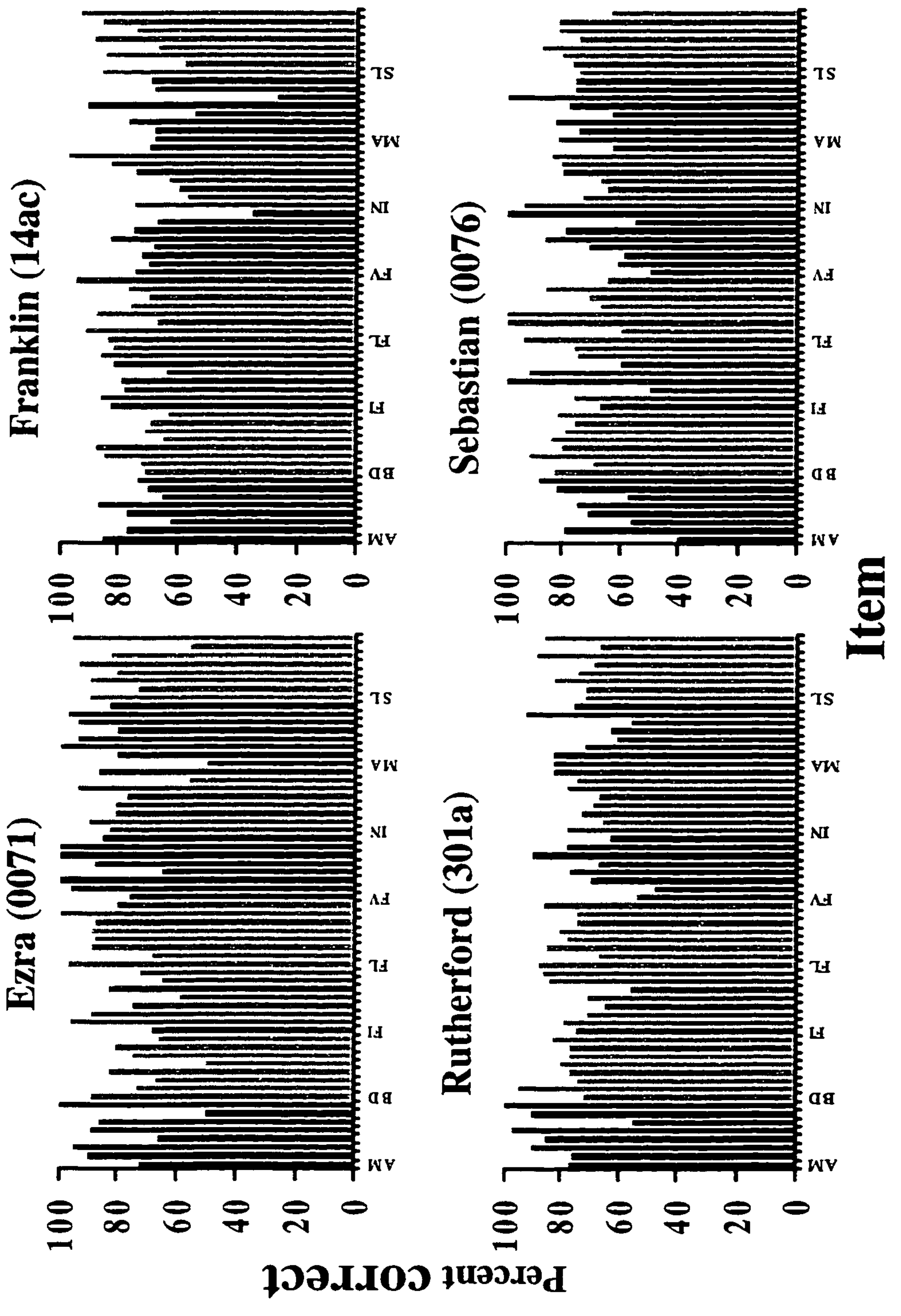


Figure D1. Overall item accuracy for familiar stimuli presented in Experiment 3.

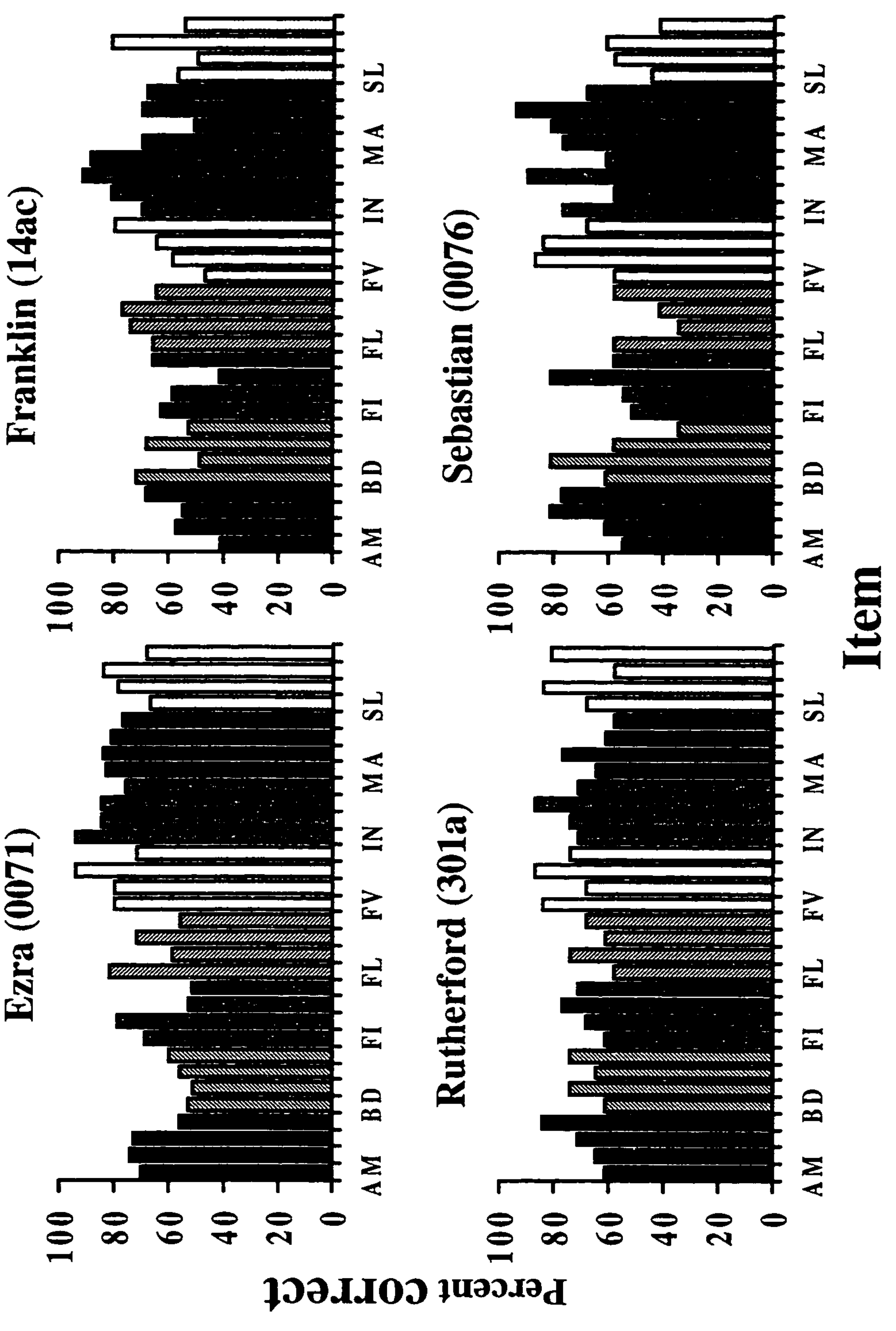


Figure D2. Overall item accuracy for novel stimuli presented in Experiment 3.

References

Astley, S., & Wasserman, E. (1992). Categorical discrimination and generalization in pigeons: All negative stimuli not created equal. *Journal of Experimental Psychology: Animal Behavior Processes, 18 (2),* 193-207.

Bauer, R., & Steele, T. (1985). Short-term memory for haptic cues in monkeys (Macaca mulatta). Animal Learning and Behavior, 13 (3), 291-302.

Berryman, R., Cumming, W., & Nevin, J. (1963). Acquisition of delayed matching in the pigeon. Journal of the Experimental Analysis of Behavior, 6(1), 101-107.

Bhatt, R. & Wright, A. (1992). Concept learning by monkeys with video picture images and a touch screen. Journal of the Experimental Analysis of Behavior, 57, 219-225.

Blough, D. (1996). Error factors in pigeons discrimination and delayed matching. Journal of Experimental Psychology: Animal Behavior Processes, 22 (1), 118-131.

Brown, J., Brown, A., & Poulson, C. (1997). Evaluating the identity concept using MTS procedures. In Baer, D., & Pinkston, E. (Eds.). Environment and Behavior. Boulder, CO: Westview Press.

Carter, D. E., & Werner, T. J. (1978). Complex learning and information processing by pigeons: A critical analysis. Journal of the Experimental Analysis of Behavior, 29, 565-601.

Chen, S., Swartz, K. B., & Terrace, H. S. (1997). Knowledge of the ordinal position of list items in rhesus monkeys. Psychological Science, 8, 80-86.

Cumming, W., & Berryman, R. (1961). Some data on matching behavior in the pigeon. Journal of the Experimental Analysis of Behavior, 4, 281-284.

D'Amato, M. (1971). Sample familiarity and delayed matching in monkeys. Psychonomic Science, 25 (3), 179-180.

D'Amato, M. (1973). Delayed matching and short-term memory in monkeys. In G.H. Bower (Ed.). The psychology of learning and motivation (Vol. 7). New York: Academic Press.

D'Amato, M., & Colombo, M. (1989). On the limits of the matching concept in monkeys (Cebus apella). Journal of the Experimental Analysis of Behavior, 52, 225-236.

D'Amato, M. R. & Van Sant, P. (1988). The person concept in monkeys (Cebus apella). Journal of Experimental Psychology: Animal Behavior Processes, 14(1), 43-55.

D'Amato, M. & Worsham, R. (1972). Delayed matching in the capuchin monkey with brief sample durations. Learning and Motivation, 3, 304-312.

D'Amato, M. R., Salmon, D. P., & Colombo, M. (1985). Extent and limits of the matching concept in monkeys (Cebus apella). Journal of Experimental Psychology: Animal Behavior Processes, 11, 35-51.

D'Amato, M. R., Salmon, D. P., Loukas, E., & Tomie, A. (1986). Processing of identity relations and conditional relations in monkeys (Cebus apella) and pigeons (Columbia livia). Animal Learning and Behavior, 14, 43-55.

Dasser, V. (1987). Slides of group members as representations of the real animals (Macaca fascicularis). Ethology, 76, 65-73.

Devine, J., Burke, M., & Rohack, J. (1979). Stimulus similarity and order as factors in visual short-term memory in nonhuman primates. Journal of Experimental Psychology: Animal Behavior Processes, 5 (4), 335-354.

Devine, J., Jones, L., Neville, J., & Sakai, D. (1977). Sample duration and type of stimuli in delayed MTS in rhesus monkeys. Animal Learning and Behavior, 5 (1), 57-62.

Estes, W. (1994). Classification and Cognition. Oxford University Press: New York.

Fetterman, J. (1996). Dimensions of stimulus complexity. Journal of Experimental Psychology: Animal Behavior Processes, 22 (1), 3-18.

Fujita, K. (1987). Species recognition by five macaque monkeys. Primates, 28, (3), 353-366.

Herman, D., & Gordon, J. (1974). Auditory delayed matching in the bottlenose dolphin. Journal of the Experimental Analysis of Behavior, 21, 19-26.

Herman, L. M., Gory, J. D., Hovancik, J. R., & Bradshaw, G. L. (1989). Generalization of visual matching by a bottlenosed dolphin (Tursiops truncatus): Evidence for invariance of cognitive performance with visual and auditory materials. Journal of Experimental Psychology: Animal Behavior Processes, 15 (2), 124-136.

Herrnstein, R. J. (1979). Acquisition, generalization, and discrimination reversal of a natural concept. Journal of Experimental Psychology: Animal Behavior Processes, 5, 116-129.

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. Cognition, 37, 133-166.

Herrnstein, R. J. & Loveland, D. H. (1964). Complex visual concept in the pigeon. Science, 16, 549-551.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 2, 285-302.

Herzog, H., Grant, D., & Roberts, W. (1977). Effects of sample duration and spaced repetition upon delayed MTS in monkeys (Macaca arctoides and Saimiri sciureus). Animal Learning and Behavior, 5 (4), 347-354.

Iverson, I. (1993). Acquisition of MTS performance in rats using visual stimuli on nose keys. Journal of the Experimental Analysis of Behavior, 59, 471-482.

Kastak, D., & Schusterman, R. J. (1994). Transfer of visual identity MTS in two California sea lions (Zalophus californianus). Animal Learning and Behavior, 22 (4), 427-435.

Keller, F., & Schoenfeld, W. (1950). Principles of Psychology. New York: Appleton-Century-Crofts.

Kojima, T. (1979). Discriminative stimulus context in MTS of Japanese monkeys. Japanese Psychological Research, 21 (4), 189-194.

Malone, D., Tolan, J., & Rogers, C. (1980). Cross-modal matching of objects and photographs in the monkey. Neuropsychologia, 18, 693-697.

Mason, M., & Wilson, M. (1974). Temporal differentiation and recognition memory for visual stimuli in Rhesus monkeys. Journal of Experimental Psychology, 103 (3), 383-390.

Moss, M., Rosene, D., & Peters, A. (1988). Effects of aging on visual recognition. Neurobiology of Aging, 9, 495-502.

Nakagawa, E. (1993). Relational rule learning in the rat. Psychobiology, 21 (4), 293-298.

Oden, D. L., Thompson, R. K. R., Premack, D. (1988). Spontaneous transfer of matching by infant chimpanzees (Pan troglodytes). Journal of Experimental Psychology: Animal Behavior Processes, 14 (2), 140-145.

Pack, A. A., Herman, L. M., & Roitblat, H. L. (1991). Generalization of visual matching and delayed matching by a California sea lion (Zalophus californianus). Animal Learning and Behavior, 19 (1), 37-48.

Perkins, D., Lydersen, T., & Beaman, D. (1973). Acquisition under mixed-delay and multiple-delay MTS. Psychological Reports, 21, 634-640.

Premack, D. (1983). Animal cognition. The Annual Review of Psychology, 34, 351-362.

Rapp, P. & Amaral, D. (1991). Recognition memory deficits in a sub-population of aged monkeys resemble the effects of medial temporal lobe damage. Neurobiology of Aging, 12, 481-486.

Roberts, R. A. & Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys and people. Journal of Experimental Psychology: Animal Behavior Processes, 14 (3), 247-260.

Roberts, W., Mazmanian, D., & Kramer, P. (1987). Memory for picture fragments in monkeys and humans. Canadian Journal of Psychology, 41 (1), 1-19.

Roitblat, H. L., Penner, R. H., & Nachtigall, P. E. (1990). MTS by an echolocating dolphin (Tursiops truncatus). Journal of Experimental Psychology: Animal Behavior Processes, 16 (1), 85-95.

Sands, S. A., Lincoln, C. E., & Wright, A. A. (1992). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. Journal of Experimental Psychology: General, 3, 368-389.

Schrier, A. M. & Brady, P. M. (1987). Categorization of natural stimuli by monkeys (Macaca mulatta): Effects of stimulus set size and modification of exemplars. Journal of Experimental Psychology: Animal Behavior Processes, 13, 136-143.

Schrier, A. M., Angarella, R., & Povar, M. L. (1984). Studies of concept formation by stumptailed monkeys: Concepts of humans, monkeys and letter A. Journal of Experimental Psychology: Animal Behavior Processes, 10, 564-584.

Swartz, K., Chen, S., & Terrace, H. (1991). Serial learning by Rhesus monkeys: I. Acquisition and retention of multiple four-item lists, Journal of Experimental Psychology: Animal Behavior Processes, 17 (4), 396-410.

Thompson, R. K. R. (1995). Natural and relationtional concepts in animals. In H. Roitblat & J. Meyer (Eds.). Comparative Approaches to Cognitive Science. Cambridge, MA: MIT Press.

Thompson, R. K. R., & Oden, D. L. (in press). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. Cognitive Science.

Urcuioli, P., & DeMarse, T. (1994). On the relationship between differential outcomes and differential sample responding in matching-to-sample. Journal of Experimental Psychology: Animal Behavior Processes, 20 (3), 249-263.

Urcuioli, P., DeMarse, T., & Lionello, K. (1998). Transfer of performance to new comparison choices following differential outcome matching-to-sample. Animal Learning and Behavior, 26 (2), 139-153.

Wasserman, E., Kiedinger, R., & Bhatt, R. (1988). Conceptual behavior in pigeons: Categories, subcategories and pseudocategories. Journal of Experimental Psychology: Animal Behavior Processes, 14 (3), 235-236.

Winer, B. (1962). Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co.

White, K., & Cooney, E. (1996). Consequences of remembering: Independence of performance at different retention intervals. Journal of Experimental Psychology: Animal Behavior Processes, 22 (1), 51-59.

Worsham, R. (1974). Temporal discrimination factors in the delayed MTS task in monkeys. Animal Learning and Behavior, 3 (2), 93-97.

Wright, A., Santiago, H., & Sands, S. (1984). Monkey memory: Same/different concept learning, serial probe acquisition, and probe delay effects. Journal of Experimental Psychology: Animal Behavior Processes, 10 (4), 513-529.