

Transfer of visual identity matching-to-sample in two California sea lions (*Zalophus californianus*)

DAVID KASTAK

University of California, Santa Cruz, California

and

RONALD J. SCHUSTERMAN

California State University, Hayward, California
and *University of California, Santa Cruz, California*

In order to assess the abilities of two California sea lions to generalize an identity concept, both animals were taught a two-choice, visual matching-to-sample task. We hypothesized that initial identity-matching problems would be learned as conditional (if...then) discriminations but that an identity concept would emerge after training numerous exemplars of identity matching. After training with 15 two-stimulus identity matching-to-sample problems, transfer tests consisting of 15 novel problems were given to the animals. Pass-fail criteria were defined in terms of performance on Trial 1 of each test problem, performance on test trials compared with baseline trials, and performance on four-trial problem blocks. One sea lion passed on the second transfer test and the other passed on the third; both demonstrated successful generalization of an identity concept by all criteria used. A second experiment consisted of presentation of stimuli previously learned in a different context (arbitrary matching-to-sample). Both subjects immediately applied an identity concept to accurately solve these new problems. These tests conclusively demonstrate transfer of an identity matching rule in California sea lions.

The ease with which humans learn many complex tasks can be at least partially attributed to the ability to form concepts, through which entire categories of problems can be easily solved (see Premack, 1983). Interest in the capacities of nonhumans to form concepts has been an issue in comparative psychology from its Darwinian beginnings. Controversy over animal consciousness, thought, and "higher-order" mental abilities is still prevalent (see Griffin, 1992). A concept is a problem-solving strategy that is based on relations between stimuli rather than on particular aspects of individual problems. (For example, the ability to carry out a multiplication task depends not upon any specific multiplication problem but on the concept of multiplication.) Once a concept is learned through experience with a certain num-

ber of exemplars of a particular type of problem, it can be used to solve a potentially infinite number of problems from the same category.

The most common procedures used in laboratory tests for concept formation make some use of matching to sample (MTS; see Carter & Werner, 1978; Nissen, Blum, & Blum, 1948; Sidman & Tailby, 1982) or sameness-difference tasks (D'Amato & Colombo, 1985; Wright, Shyan, & Jitsumori, 1990). In a visual MTS procedure, the subject is presented with a sample stimulus followed by exposure to two or more comparisons. Choice of the correct comparison (S+) results in reinforcement, whereas choice of the incorrect comparison (S-) is not reinforced. Sameness-difference tasks are similar, except that the subject must respond to a pair of matching stimuli (rather than to individual comparison stimuli), and must withhold response to pairs of stimuli that are not identical.

To train and test for an animal's ability to form a sameness concept within an MTS context, a procedure called identity matching-to-sample (IMTS), in which sample and comparison stimuli are duplicates, is used. After being trained to match a variety of such duplicates, the subject can potentially learn an identity concept that will allow it to solve any number of novel problems.

Transfer or generalization of a sameness concept has been reported in a variety of subjects, including pigeons (Wright, Cook, Rivera, Sands, & Delius, 1988; Zentall & Hogan, 1974), monkeys (D'Amato & Colombo, 1985; Wright et al., 1990), chimpanzees (Nissen et al., 1948;

The experiments and preparation of this manuscript were financially supported by grants from the Office of Naval Research (Contract N00014-85-K-0244) and the Center for Field Research (Earthwatch) to R.J.S. The authors gratefully acknowledge the valuable assistance of Earthwatch and student volunteers in designing stimuli and conducting experiments. We would also like to thank Brigit Grimm and Michel Coulliais for help in the experimental design, and Colleen Reichmuth for helpful criticisms of an earlier version of this manuscript. D.K. is affiliated with the Department of Biology at the University of California, Santa Cruz, and R.J.S. is affiliated with the Departments of Psychology and Biology at California State University, Hayward, and with the Institute of Marine Sciences at the University of California, Santa Cruz. Correspondence should be addressed to R. J. Schusterman, Department of Psychology, California State University, Hayward, CA 94542.

Oden, Thompson, & Premack, 1988), a sea lion (Pack, Herman, & Roitblat, 1991), and a dolphin (Herman, Hovancik, Gory, & Bradshaw, 1989). It has been argued that dolphins and chimpanzees (presumably because of their large brains) easily form concepts, while smaller-brained animals (pigeons) have much more difficulty in doing so. Species intermediate in brain size (such as sea lions and monkeys) are thought to form concepts with more difficulty than dolphins and chimpanzees, but more easily than pigeons (Pack et al., 1991).

On the other hand, differences in training (of language, for instance; see Premack, 1983), manipulatory ability (such as in chimpanzees or monkeys; Oden et al., 1988), and procedure may account for what appear to be such drastic, species-specific differences in cognitive capacity. If the ability to form general rules to solve entire classes of problems is highly adaptive, then it ought to be found in a variety of (not necessarily large-brained) taxa. Further, if the behaviors tested in different species are judged to be similar, the abilities underlying the behaviors are likely to be general and not species specific.

Several factors important in the interpretation of results obtained from generalized identity matching tests have recently been addressed (Dube, McIlvane, & Green, 1992; Kastak & Schusterman, 1992; Schusterman & Kastak, in press), and the reader is referred to these articles for specifics. First, and probably most important, is the "exclusion" effect, which provides an explanation other than that of concept formation that is applicable to several studies reporting transfer of an identity rule in monkeys (D'Amato & Colombo, 1985), pigeons (Holmes, 1979), a dolphin (Herman et al., 1989), and a sea lion (Pack et al., 1991). During testing, the correct and incorrect comparisons are often chosen such that the correct comparison is a novel stimulus and the incorrect comparison is familiar (i.e., previously trained). Consequently, a correct choice does not necessarily occur because of an identity relation between it and the sample. The subject could instead "exclude" the familiar incorrect comparison as a choice because it has already been related to another sample [see Schusterman, Gisinier, Grimm, & Hanggi (1993) for a discussion of exclusion in marine mammals' conditional discriminations]. The response is made by process of elimination rather than by a more complex cognitive ability such as identity matching (i.e., the subject need not learn a rule governing a general class of stimulus-stimulus relations in order to use exclusion). In fact, exclusion is often applied immediately upon presentation of novel relations (Schusterman et al., 1993), while it is generally agreed that formation of a concept requires some degree of training. To ensure that the subject has no opportunity to respond on the basis of exclusion, novel test stimuli should only be paired with other novel stimuli (Dube et al., 1992; Kastak & Schusterman, 1992; Schusterman & Kastak, in press).

In addition, novelty itself can affect test results if individual animals respond in unusual and disruptive ways to presentation of novel test stimuli (Holmes, 1979; Pack

et al., 1991). In order to properly assess generalization, an experimenter needs to determine how the subject responds on the very first exposure to novel identity problems. Since the disturbance from such an effect can drastically interfere with the interpretation of first-trial data, any disruption brought about by test novelty should be eliminated. One successful way to avoid a novelty effect is to provide the subject with a very large pool of both training and test stimuli (Holmes, 1979; Wright et al., 1988). Presumably, the subject habituates to the novelty associated with a test situation by being repeatedly exposed to a large number of novel stimuli of the same general type prior to the test.

In the present study, we report the results from IMTS training and testing with two California sea lions, members of a species considered by Pack et al. (1991) to be intermediate in concept-forming ability. The tests comprised the initial phase of a more comprehensive experiment designed to demonstrate stimulus equivalence (see Sidman & Tailby, 1982). Specifically, *reflexivity* (the ability to conditionally relate a stimulus to itself) was to be examined. Reflexivity is a property of equivalence which seems to be fundamentally tied to symbolic or referential aspects of language in humans. Currently, the question of whether any nonhuman animal can demonstrate equivalence is open to some debate (Hayes, 1989), so success in a California sea lion might shed some light on cognitive abilities in taxa only distantly related to humans (see Schusterman & Kastak, 1993).

Because of the effects of novelty and exclusion, the results of the only previously reported IMTS tests involving this species (Pack et al., 1991) are inconclusive. In order to avoid these interfering effects detrimental to other matching experiments, the following provisions were included in the experimental design:

1. All of the initial training problems (except one), as well as all of the test problems, were arranged in sets of two stimuli, and novel stimuli were always paired with other novel stimuli. Thus, exclusion could not be used to guide choices.

2. A large number of stimuli were used in training, in accordance with the hypothesis that many such stimuli would facilitate concept formation and habituate the subjects to test novelty (Wright et al., 1988).

3. A large pool of test stimuli was used in order to allow an accurate assessment of a large sample of first-trial data. The use of first-trial data is necessary for interpretation of generalization or transfer of the identity concept, since testing in extinction typically disrupts the subject's behavior (Oden et al., 1988; Schusterman & Kastak, 1993; Wright et al., 1988).

4. In order to assure that conditionality was maintained during testing procedures (see Dube et al., 1992), each comparison stimulus had an equal probability of appearing as the S+ and the S-, thus ensuring that the subjects made their choices on the basis of the relationship between samples and comparisons and not merely on the basis of particular aspects of the comparison stimuli.

5. Provisions were made whereby, upon failure of any transfer test, another test would be given, and the previous test stimuli would be incorporated into the baseline of learned stimulus relations. We hypothesized that the identity-matching experience gained by completing one test would facilitate successful performance on any subsequent test(s).

GENERAL METHOD

Subjects

The subjects of these experiments were two female California sea lions (*Zalophus californianus*): Rocky, a fifteen-year-old, and Rio, a six-year-old, both housed at Long Marine Laboratory in Santa Cruz, California. Both sea lions had previously participated in extensive arbitrary MTS testing (see Schusterman et al., 1993). Experimental sessions were conducted on an average of five days per week, and ordinarily between the hours of 9:00 a.m. and noon. The animals were fed between 4 and 6 kg of freshly thawed cut herring and capelin per day, one third of which was typically consumed during sessions.

Apparatus

The apparatus was made up of a set of hinged wooden panels containing three window-fronted boxes in which the stimuli were placed (Schusterman et al., 1993). The middle board, housing the sample window, was 120×120 cm and the side boards were 120 cm high × 61 cm wide. The stimulus boxes were 30×30 cm, and 10 cm deep, covered by sliding opaque doors that allowed the stimuli to be placed inside, out of view of the subject. During the experiments, placement of the stimuli in the boxes was done by two assistants seated behind the boards, who received instructions from the experimenter via radio headphones. The two comparison stimuli were placed in their respective boxes simultaneously, so the subject could not be cued to the correct choice by the timing of its placement. A third assistant seated behind the boards delivered a piece of fish as reinforcement for correct responses. Stimuli used in Experiment 1 consisted of black shapes painted on white backgrounds (see Figure 1), on 30×30 cm pieces of plywood. Stimuli used in Experiment 2 were three-dimensional "junk" objects constructed of wood, steel, and/or plastic, painted black, and also presented against a white background.

Procedure

At the beginning of a trial, the sample stimulus was exposed for approximately 4 sec before the two side doors were opened, revealing the correct (S+) and incorrect (S-) comparisons. No observing response was required, and the sample remained exposed during presentation of the comparisons (simultaneous condition).

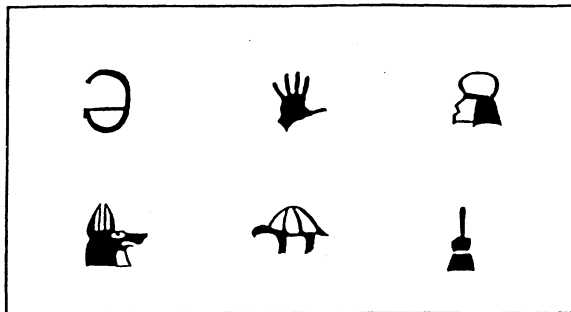


Figure 1. A sample of six stimulus configurations used in Experiment 1 for both Rocky and Rio.

After an interval of 2–4 sec, the subject was released from station (directly in front of the center box) in order to point at its choice. A response was defined by the nose of the subject breaking the plane formed by the front of the stimulus box. Agreement between two judges regarding correct and incorrect responses was nearly perfect. Correct responses were rewarded with a piece of fish. All trials were balanced for left and right correct choices, and all responses were differentially reinforced. Intertrial intervals ranged from 10–25 sec. For both testing and training phases of each experiment, a *problem* was defined as two sets of matching stimuli (for example, the stimuli *I* and *triangle*). Baseline trials were all trials in which a previously trained stimulus served as sample and correct comparison. Reference to performance on a test trial means performance on a particular trial (e.g., Trial 1) for *all* test problems (e.g., the percentage of correct responses on Trial 1 of all test problems).

EXPERIMENT 1

Method

Exclusion phase. Rio's initial training problem of two sets of stimuli consisted of exclusion trials—that is, correct stimuli were always presented as alternative comparisons to a familiar stimulus [see Schusterman et al. (1993) for a detailed discussion of "errorless" learning of MTS via exclusion in sea lions]. During this phase of the experiment, Rio's responses would probably be determined by the exclusion of the familiar, nonmatching comparison as a possible choice rather than by an identity relation between sample and correct comparison. Each session consisted of 40 exclusion trials (in which the familiar stimulus was the incorrect comparison) and 20 baseline trials. In addition, probe trials (two per session) pitting novel comparisons against each other were administered in order to assess performance on trials in which exclusion could not be used. In contrast to this procedure, exclusion was not used in any phase of Rocky's training.

Trial-and-error training phase. Rio's 14 additional training problems were all taught in a conventional manner (trial and error). Forty trials of each 60-trial session consisted of pairings of two training stimuli, and 20 were baseline trials (previously learned stimuli, kept in pairs). Novel stimuli were always introduced and trained in pairs; two distinct samples and their corresponding comparisons were trained together, not one at a time as sample and S+. When a criterion of 90% correct responses (36 out of 40 total trials, or 18 out of 20 for each stimulus) in two consecutive sessions was met, the training stimuli were incorporated into the baseline of previously learned stimulus relations. Hence, after the acquisition of each problem, the baseline expanded in size by two stimuli. Rocky's entire training set of 15 problems was trained in this fashion.

Reshuffling. Prior to the reshuffling phase, training problems were kept in pairs, both members of a pair appearing as sample and correct (or incorrect) comparison. During reshuffling, the problems were split up so that any stimulus could appear together with any other stimulus in a single MTS trial. For example, the first identity problem consisted of the stimulus relations *I-I* and *triangle-triangle*, and the second problem was *fish-fish*, *hook-hook*. During the trial-and-error training phase, when *fish* appeared as the S+, *hook* always appeared as the S-. Likewise, when *hook* appeared as the S+, *fish* always appeared as the S-. During reshuffling, however, *fish* and *triangle*, *hook* and *I*, *fish* and *I*, etc., could all appear as alternative comparisons. There was a total of 840 trials (30×28, because the initial two pairings weren't included).

The rationale for reshuffling was threefold: (1) the additional experience on a large number of training trials was thought to facilitate concept formation (see Wright et al., 1988); (2) any dependence on the context of training would be eliminated, so that the subject's choice would depend only on the relation between sample and correct comparison, and not on the incorrect compar-

ison with which it was learned; and (3) by breaking up the training problems, we attempted to minimize control by other, potentially correlated, features of the training sets. Conceivably, Rocky and Rio could solve certain problems by attending to certain categorical features of the sample and comparison stimuli (e.g., by matching solid to solid or hollow to hollow). By reshuffling, any tendency to respond in this manner would be extinguished, since reshuffled problems would cut across all possible categories.

Testing. Rio's initial test pool consisted of 30 novel stimuli grouped into 15 problems. These problems were superimposed on a baseline of previously trained stimuli. The three test sessions each consisted of 10 unique test trials randomly placed among 50 baseline trials. As in training and baseline, all trials were balanced for left and right correct choices, and all responses were differentially reinforced.

The second test for Rio consisted of 30 additional novel stimuli, also arranged into 15 problems. Four training sessions were run in the interval between the first and second tests, which was just long enough to allow for the construction of 15 additional sets of stimuli. In these four sessions, the Test 1 stimuli were used as baseline stimuli, although they were never trained to any set criterion. During Test 2, experimental trials were grouped into blocks of four consecutive trials, each of the same problem, balanced for side and for S+ (i.e., each stimulus in a problem appeared twice as a correct comparison and twice as an incorrect comparison, as well as twice on each side). Each of the six sessions consisted of 20 test trials (5 blocks) superimposed on a new baseline that included all the stimuli from the first test in addition to training stimuli.

For Rocky, both Test 1 and Test 2 followed the design for Rio's Test 2. In Rocky's third test, the ratio of test trials to baseline trials was increased (40 test trials to 20 baseline trials in each of the three sessions). In addition, the four trials of each test problem were not run consecutively but were distributed randomly throughout the session. The Test 3 baseline consisted of Test 1, Test 2, and training stimuli.

Assessment. Results from the transfer tests were analyzed in several ways: (1) first-trial performances were compared with chance (50%); (2) test trials were divided into two groups—Trials 1–4 (the first block of four trials) and Trials 5–8 (the second block of four trials) for each problem; (3) Trial 1 was also analyzed separately due to its importance in demonstrating rapidity of transfer; (4) performance on each set of trials was compared with baseline performance for the particular session in which it appeared; and (5) as a separate analysis, each of the problems was treated as a four-trial test to be either passed or failed. Since problems were administered in blocks of four trials, there were two separate tests for each problem: Trials 1–4 formed the first test, and Trials 5–8 the second test. A pass was defined as three or four out of four trials correct, with anything less constituting a failure.

Results and Discussion

Training. Rio's first training problem was learned with little difficulty. She was correct on 36 out of 40 trials (90%) in which novel stimuli were correct choices during Session 1, and on 314 out of 320 trials (98%) for the entire training period (eight sessions). However, initial probe trials pitting two novel comparisons against each other were responded to at chance levels (50% on Days 1 and 2, and 70% on Days 3–5). Totaled, Rio's performance on these trials was slightly, though not significantly, better than chance (12 of 16 correct; two-tailed binomial test, $p > .05$). These data are consistent with the exclusion interpretation of Schusterman et al. (1993) for sea lions in an arbitrary MTS task.

Reshuffling. Neither animal showed a decrement in performance during the reshuffling phase (99% correct responses for Rio and 93.3% correct responses for Rocky), even though in many cases, stimuli from the same category (judged by obvious similarities such as size, orientation, etc.) appeared as both correct and incorrect comparisons in the same trial. If the incorrect comparison happened to fall into a different perceptual category than the sample, a correct choice in an IMTS test might be based on a categorical relationship between the sample and the correct comparison rather than on an identity relationship. Reshuffling trials with two similar comparisons were no more difficult for the animals to solve than presumably less difficult trials in which comparisons were extremely different in appearance. This evidence supports the conclusion that during this phase of training both animals either were responding on the basis of an identity relationship or had learned "if...then" rules for every sample/S+ combination (Carter & Werner, 1978).

Test 1. Rio's performance on the first two trials (for 15 problems) of Test 1 was 60%—not significantly better than chance (two-tailed binomial test, $p > .05$). Her performance on Trials 3–6 rose to levels above 90%, suggesting either rapid learning or concept formation. Initial evidence that Rio was treating novel problems differently than familiar ones came by way of an untrained response by Rio—in the form of tactile contact (with vibrissae) and occasional vocalizations—to the presentation of some novel stimuli. Combined with her poor performance on Trials 1 and 2, this strongly suggested that presentation of novel samples disrupted Rio's testing behavior. These reactions (novelty responses) were rarely observed on baseline trials. Rio very quickly began to treat test stimuli as baseline stimuli and to perform at baseline levels; by the third exposure of each stimulus, the novelty effect seemed to have worn off.

In order to make a comparison between test trial and baseline trial behavior, the novelty response was quantified in relation to trial type. During the test, Rio exhibited a novelty response on 15 out of the 30 novel sample presentations, while on an equal-sized sample of baseline trials, she exhibited the novelty response only 6 times—a significant difference [$\chi^2(1) = 5.9$, $p < .05$].

Although Rocky, unlike Rio, showed no apparent reaction to presentation of novel stimuli as samples, it is reasonable to assume that novel stimulus presentation was disruptive for her as well. Interestingly, in previous work in arbitrary matching-to-sample, rather than reacting to the novel sample itself, Rocky occasionally refused to respond to novel stimuli as *comparisons* (Schusterman et al., 1993).

Rocky's performance on Test 1 was ambiguous. Although she performed at a level of 87% correct on the first trial (for 15 problems), her response accuracy quickly fell to chance levels on the several trials that followed.

Test 2. In contrast to Test 1, in which Rio showed novelty reactions to the sample stimuli on 15 of the 30 initial presentations, in Test 2 she showed this reaction on

a significantly lower proportion of novel sample presentations [10%, or 3 out of 30; $\chi^2(1) = 11.4, p < .05$]. In addition, for a random sample of 30 baseline trials during Test 2, the novelty response was exhibited twice, not a significantly different level from that of test trials [$\chi^2(1) = .218, p > .05$].

On Trial 1 of Test 2, Rio's performance was significantly better than chance [80%, or 12 out of 15; $p < .05$, two-tailed binomial test]. In addition, improvement was shown on Trials 2–4 (93.3%) and on Trials 5–8 (95%). Figure 2 shows Rio's performance level on Test 2 compared with baseline performance. Chi-square tests performed on an equal number of test trials (Trials 1–4) and baseline trials showed no significant difference between performance on the two [$\chi^2(1) = 2.143, p > .05$]. This result, in conjunction with the data on novelty responses, indicates that the disruption of performance for Rio on Test 1 was ameliorated during Test 2; once novelty was neutralized as a performance factor, successful IMTS generalization was demonstrated.

Overall, the results of Rocky's second test (not shown) were comparable to those of Test 1—her performance on Trial 1 was 46.7%. She improved somewhat on Trials 2–4 of the block, to a level of 93.3% correct responses, which is indicative of within-problem learning. However, an unexplained decrement in performance to 66.7% and 60% on Trials 5 and 6 suggests that Rocky was using a strategy other than identity matching to solve these problems.

Test 3. Rocky's performance on Test 3 is shown in Figure 3. First-trial performance (for 15 novel problems) was strong (73.3%), but not significantly better than chance ($p > .05$, two-tailed binomial test). A chi-square test on the first four-trial block (54 out of 60 correct, or 90%) versus an equal number of baseline trials, however, showed no significant difference in performance [$\chi^2(1) = 1.034, p > .05$]. As in Rio's case, performance on Trials 5–8 (93.3%) equaled that of baseline.

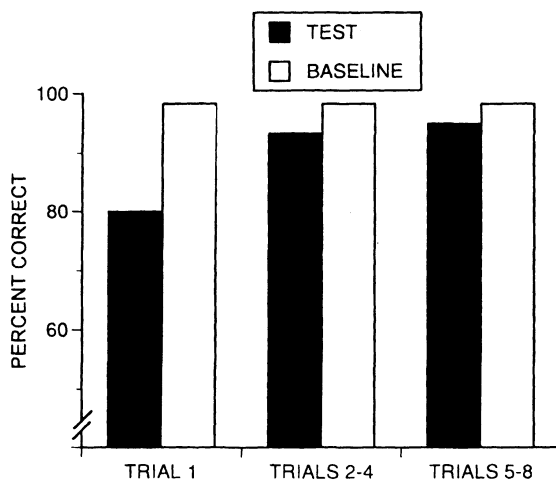


Figure 2. Performance by Rio on Test 2, Experiment 1.

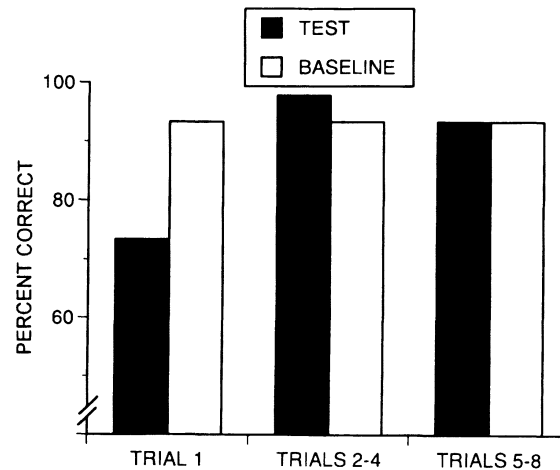


Figure 3. Performance by Rocky on Test 3, Experiment 1.

Pass-Fail analysis. Pass-fail analyses of Rio's Test 2 and Rocky's Test 3 are shown in Figures 4 and 5. Rio passed all 15 two-stimulus problems in the first block (Trials 1–4) and 14 out of 15 problems in the second block (Trials 5–8). Of the successful problems, on the first block, 9 were passed with four correct responses out of four trials and on the second block, 13 were passed with four out of four correct. Rocky passed 14 problems on the first block (11 passed with four out of four correct) and 15 problems on the second block (again, 11 passed with four out of four correct).

The advantage of a pass-fail analysis is that dependence on first-trial data is reduced. In spite of diminished novelty reactions to novel stimulus presentation, the first appearance of a particular sample may still disrupt in other ways. This is reflected in a slight decrement in first-trial performance even during successful generalization tests. In a pass-fail analysis, there is no need to exclude potentially illuminating data because of such novelty effects. In addition, this measure of transfer is independent of baseline performance because the two trial types are not compared; each problem becomes a test which, on its own merits, is either passed or failed.

EXPERIMENT 2

In order to assess the degree to which generalization of a matching rule would occur with stimuli previously encountered in a nonidentity context, both Rocky and Rio were retrained on an arbitrary MTS task on which each subject had received extensive prior training (Schusterman et al., 1993), and they were then tested on IMTS using the familiar stimuli from this task. The tests were essentially the same as those in Experiment 1 except for the nature of the stimuli used: instead of novel test stimuli, each animal was tested with familiar stimuli.

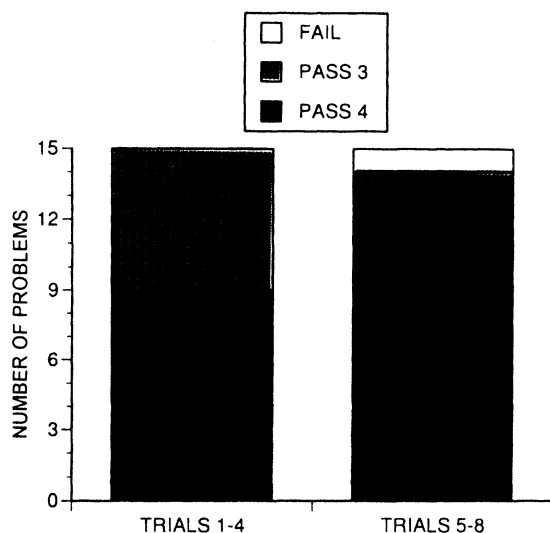


Figure 4. Pass-fail analysis for Rio on Test 2, Experiment 1. Bars indicate the number of problems passed at each level (4/4 trials correct; 3/4 trials correct) or failed.

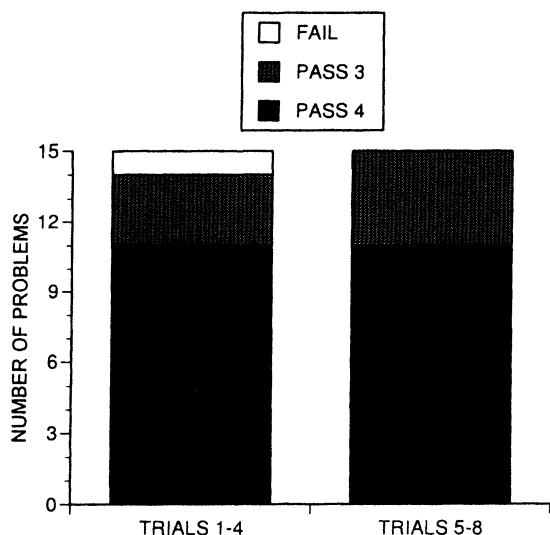


Figure 5. Pass-fail analysis for Rocky on Test 3, Experiment 1. Bars indicate the number of problems passed at each level (4/4 trials correct; 3/4 trials correct) or failed.

Method

Training. After completing tests for identity generalization, both Rio and Rocky were reintroduced to an arbitrary MTS procedure using previously learned stimulus relations. All of the relations learned in this context consisted of pairings of three-dimensional stimuli, as opposed to the planometric stimuli used in Experiment 1. Between Experiments 1 and 2, Rio learned an additional 22 arbitrary relations in a period of about six months. Rocky was retrained to a criterion of 90% correct on the arbitrary relations she had previously acquired (approximately 2 weeks of training). Experiment 2 involved placing each stimulus from a previously learned arbitrary relation into the context of an IMTS pro-

cedure. For instance, the arbitrary relation of *ring*-*bat* could produce two possible identity trials: *ring* as sample and correct comparison; and *bat* as sample and correct comparison. Each animal was presented with 10 problems of two stimuli each.

Test trials occurred in two blocks of four (for a total of 8 trials per problem) and were distributed randomly throughout the session. Except for an initial 10 trials with previously learned IMTS stimuli as a warm-up, the entire experimental session consisted of the animal matching these previously seen stimuli in an IMTS context. The stimuli used in half of the problems had previously appeared as comparisons only. Those used in the other half consisted of stimuli that had previously appeared only as samples. No further adjustments for novelty were made, since all samples were familiar stimuli. Each stimulus appeared four times as a correct comparison and four times as an incorrect one; correct choices were balanced for side. Sessions consisted of 40 experimental trials.

Results and Discussion

The number of correct responses for Trials 1-8 of the 10 arbitrary to identity-matching problems are shown for Rio and Rocky in Figure 6. First-trial performances by both animals were strong (80% for Rio and 70% for Rocky), and both quickly improved to levels of > 80%. This level, however, was not consistently maintained. Either of two explanations may account for lowered performance on some trials relative to transfer performance in Experiment 1: (1) it may have been a sampling phenomenon due to the small number of problems; or (2) it may have been due to the different experimental histories of the stimuli involved (arbitrary vs. identity matching). Interestingly, in another experiment in which one of the comparisons was a physical (identical) match to a particular sample and the other was a correct (previously learned) arbitrary match, both Rocky and Rio showed a strong tendency to choose the previously learned arbitrary match (Spada, Kastak, & Schusterman, 1994).

Pass-Fail analysis. As shown in Figure 7, Rocky passed 10 out of 10 problems on the first block of four trials, 7 of them being passed with four out of four trials correct. On the second block, she passed 8 out of 10 problems, six being passed with four out of four trials correct.

Rio passed 7 out of 10 and 8 out of 10 problems on the first and second blocks, respectively, but not as many problems were passed with four out of four trials correct. Though not quite as robust as those in Experiment 1, these data reflect generalization of an identity concept extending to stimuli previously matched in a different context. The results indicate that Rocky and Rio show reflexivity among elements previously related only to dissimilar or nonmatching stimuli. Lower performance levels overall may be attributed to the different experimental histories of the stimuli used. The stimuli tested in Experiment 1 were completely novel, while those in Experiment 2 had a long history of use in arbitrary or "symbolic" matching tasks (and were completely familiar). The subjects' expectancies of being presented with nonidentical stimuli (i.e., the previously learned "correct comparisons") during this experiment probably interfered with identity matching. If, as our unpublished data (Spada et al., 1994) suggest, memory for these arbitrary relations is applied much more readily than a sameness

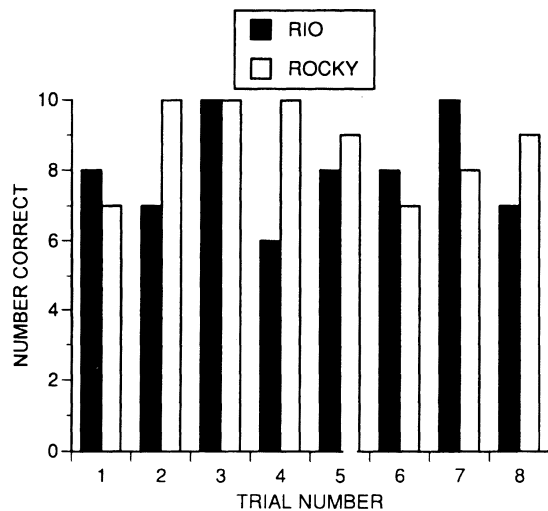


Figure 6. Total number correct (for 10 problems) on Trials 1-8 by Rocky and Rio in Experiment 2. Trial 1 corresponds to the first exposure in an identity context of each of 10 novel problems.

concept, interference because of these expectancies would probably have been just as disruptive, if not more so, than the novelty effect described earlier.

GENERAL DISCUSSION

Both sea lions in this study transferred an identity concept to completely novel problems in a visual MTS test. There are several factors that probably influenced the strong performance by the subjects. First, a large number of exemplars was provided. This is thought to have facilitated sameness-concept formation in at least two other studies (Wright et al., 1988; Wright et al., 1990). Considering that each animal had previous and extensive experience of matching stimuli that were not physically identical, the extended training phase probably allowed them to determine which previously used matching strategies (e.g., matching on the basis of categories or other perceptual similarities) were incorrect, so that the proper strategy could be utilized (see Harlow, 1959). Second, the reshuffling phase, in which the specific problems were broken up, reduced any reliance on particular attributes or categories, strategies known to be used by other animals (Herrnstein, Loveland, & Cable, 1976), as well as by at least one of the sea lions (Rio) involved in this study in the context of an arbitrary MTS task (Schusterman et al., 1993). For example, consider an animal who responds on the basis of two categories of stimuli, solid and hollow. If, during training, the two alternative comparisons are always solid versus hollow, the subject should have little difficulty responding correctly. If during the reshuffling phase, however, two different stimuli that fall into the solid category appear as comparisons, the categorization strategy is no longer useful. As reinforcement probabilities change, improper strategies used to solve initial problems are gradually ex-

tinguished, giving way to more accurate ones. In identity matching, it becomes advantageous for the subject to solve problems based on one factor only—namely, the identity relation between each sample and correct comparison.

Finally, the method of testing used in Experiment 1 allowed the subjects to habituate to the continued presentation of novel stimuli. Previous studies have had either to rely on data collected after the first exposure to novel test stimuli or to consider first-trial performance a failure because of disruptive, neophobic responses (D'Amato, Salmon, & Colombo, 1985; Pack et al., 1991). Only one subject in this study (Rio) exhibited an overt response to novelty. This response, however, was manifested not as avoidance of novel stimuli, but rather as intense investigation. The correlation between this response and performance on test trials, and the subsequent improvement following the loss of a novelty response, constitute powerful evidence for a disruption in performance based on stimulus novelty. Thus, the differing results of Rio's first and second tests suggest an interpretation based not upon a delay in forming a sameness concept until Test 2, but upon other reactions to the novel stimuli during a test session. For example, the ratio of test trials to baseline trials in Test 1 was 1:5, and the test trials were randomly interspersed among baseline trials. If Rio could predict the occurrence of baseline trials, she could be consistently reinforced (her baseline performance was automatic and virtually without error), without having to pay much attention to test trials. The same type of reaction has been observed in IMTS tests with human children. If given a choice between novel and familiar sample stimuli, they invariably choose the familiar samples, especially when the intervals between trials are long (Saun-

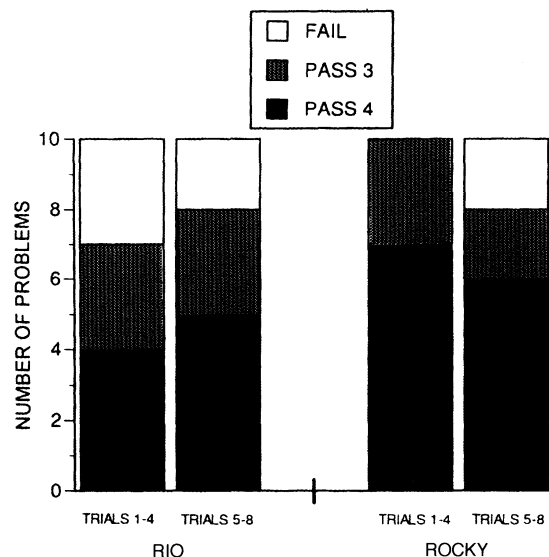


Figure 7. Pass-fail analysis for Rio and Rocky in Experiment 2. Bars indicate number of problems passed at each level (4/4 trials correct; 3/4 trials correct) or failed.

ders & Sherman, 1986). In other words, problems with a high probability of reinforcement were preferred over novel ones by children. We can reasonably assume that Rio was reacting in a similar manner.

One solution to this problem was to increase the number of test trials relative to baseline trials, so that Rio could not predict the occurrence of a long sequence of baseline trials during the test. In order to be consistently reinforced, she would have to pay greater attention to appropriate aspects of novel stimuli—that is, to the identity relation between the sample and the correct comparison.

In contrast to Rio, Rocky showed no reaction upon exposure to novel stimuli, but still showed poor transfer initially. However, in previous arbitrary MTS tests, Rocky displayed a neophobic response in the traditional sense, by refusing to respond to novel comparisons (Schusterman et al., 1993). It was logical to believe, therefore, that novel-sample presentation within the context of her IMTS experiment would have a similarly disruptive effect without necessarily being reflected in overt behavior.

The differences in the degree of transfer may be accounted for in several ways. It may be that the more drastic change in Rocky's test format relative to that of Rio was sufficient to allow demonstration of Rocky's matching ability. It is equally likely, however, that individual differences may account for the apparent differences in the rate of transfer. Age, for instance, is a possible factor that might be reflected in differences in cognitive ability. It is also possible that Rio, raised by humans as one of the subjects of an imprinting experiment (Schusterman, Gisiner, & Hanggi, 1992), was simply more adaptable than Rocky to the testing situation.

It is impossible to determine the exact point (or range of problems) at which concept formation occurred for either animal, but, in all probability, both were using an identity concept to solve problems by the end of their training phases. The exact number of exemplars needed to train the identity concepts is unclear, but estimates from other studies range from two—in the case of Oden et al.'s (1988) chimpanzees (using a combination of visual and tactile matching)—to 152 for pigeons in visual MTS (Wright et al., 1988). In a study on visual matching in *Tursiops truncatus* (Herman et al., 1989), the subject received at least 11 and as many as 15 exemplars before being presented pairings of novel objects (see Herman et al.'s Experiment 3, p. 131). The sea lions in this study seem to compare favorably with *Tursiops*, in requiring a somewhat intermediate number of exemplars (i.e., lying between those required by pigeons and those required by chimpanzees).

Perhaps the more intriguing question to ask is: To what degree can a matching concept be transferred? The answer may provide insight into the relative abstractness of the concept. Iversen, Sidman, and Carrigan (1986), for instance, demonstrated that context plays an important role in MTS tests. In experiments designed to test the role of sample position, the matching performance of monkeys in a two-choice matching task suffered when

the sample was shifted from the center to one of the two sides. The results suggest that matching performance may not transfer even to novel arrangements of the testing stimuli. Exploratory studies with Rocky and Rio have yielded nearly identical results.

Further, D'Amato and Colombo (1989) tested transfer of IMTS performance from static stimuli to dynamic (flashing) stimuli in cebus monkeys, also with negative results. They concluded that the matching concept may be limited to the precise context under which it was learned, although additional training might produce positive results in a variety of contexts. Experiments under way at our laboratory are currently testing the sea lions' abilities to transfer matching to a novel context.

Both Rocky and Rio demonstrated successful transfer in Experiment 2, which arranged a slightly less stringent test of the abstractness of the identity concept. This experiment provided evidence that identity matching is not limited to the types of stimuli with which the ability was trained. Stimuli formerly used in other matching contexts were immediately matched using an identity rule. These stimuli were all three-dimensional objects, whereas the original IMTS stimuli were two-dimensional. The fact that successful generalization of identity matching to both two- and three-dimensional stimuli was observed contrasts somewhat with a previous conjecture that three-dimensional stimuli might be necessary to demonstrate this ability in marine mammals (Herman et al., 1989; Pack et al., 1991).

It is probable that, in addition to California sea lions, many other species of nonhuman animals are capable of immediate transfer at levels approximating those of chimpanzees. However, modification of the testing procedure to allow for analysis of first-trial data in the case of "neophobic" animals may be a requirement. In any case, the issues may be clarified by a standardization of procedures and analysis.

REFERENCES

- 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.
- D'AMATO, M. R., & COLOMBO, M. (1985). Auditory matching-to-sample in monkeys (*Cebus apella*). *Animal Learning & Behavior*, **13**, 375-382.
- D'AMATO, M. R., & 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., 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.
- DUBE, W. V., MCLIVANE, W. J., & GREEN, G. (1992). An analysis of generalized identity matching-to-sample test procedures. *Psychological Record*, **42**, 289-300.
- GRIFFIN, D. R. (1992). *Animal minds*. Chicago: University of Chicago Press.
- HARLOW, H. F. (1959). Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2, pp. 492-537). New York: McGraw-Hill.
- HAYES, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, **51**, 385-392.
- HERMAN, L. M., HOVANCIC, J. R., GORY, J. D., & 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**, 124-136.
- HERRNSTEIN, R. J., LOVELAND, D. H., & CABLE, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**, 285-302.
- HOLMES, P. W. (1979). Transfer of matching performance in pigeons. *Journal of the Experimental Analysis of Behavior*, **31**, 103-114.
- IVERSEN, I. H., SIDMAN, M., & CARRIGAN, P. (1986). Stimulus definition in conditional discriminations. *Journal of the Experimental Analysis of Behavior*, **45**, 297-304.
- KASTAK, D. A., & SCHUSTERMAN, R. J. (1992). Comparative cognition in marine mammals: A clarification on match-to-sample tests. *Marine Mammal Science*, **8**, 414-417.
- NISSEN, H. W., BLUM, J. S., & BLUM, R. A. (1948). Analysis of matching behavior in chimpanzee. *Journal of Comparative & Physiological Psychology*, **41**, 62-74.
- ODEN, D. L., THOMPSON, R. K. R., & PREMACK, D. (1988). Spontaneous transfer of matching in infant chimpanzees, *Pan troglodytes*. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 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 & Behavior*, **19**, 37-48.
- PREMACK, D. (1983). The codes of man and beasts. *Behavioral & Brain Sciences*, **6**, 125-167.
- SAUNDERS, R. R., & SHERMAN, J. A. (1986). Analysis of the "discrimination failure hypothesis" in generalized matching and mismatching behavior. *Analysis & Intervention in Developmental Disabilities*, **6**, 89-107.
- SCHUSTERMAN, R. J., GISINER, R., GRIMM, B. G., & HANGGI, E. B. (1993). Behavioral control by exclusion and attempts at establishing semanticity in marine mammals using matching-to-sample paradigms. In H. L. Roitblat, L. M. Herman, & P. Nachtigall (Eds.), *Language and communication: Comparative perspectives* (pp. 249-275). Hillsdale, NJ: Erlbaum.
- SCHUSTERMAN, R. J., GISINER, R., & HANGGI, E. B. (1992). Imprinting and other aspects of pinniped-human interactions. In H. Davis & D. Balfour (Eds.), *The inevitable bond: Examining scientist-animal interactions* (pp. 334-356). New York: Cambridge University Press.
- SCHUSTERMAN, R. J., & KASTAK, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *Psychological Record*, **43**, 823-839.
- SCHUSTERMAN, R. J., & KASTAK, D. (in press). There is no substitute for an experimental analysis of marine mammal cognition. *Marine Mammal Science*.
- SIDMAN, M., & TAILBY, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, **37**, 5-22.
- SPADA, E. C., KASTAK, D., & SCHUSTERMAN, R. J. (1994). *Long-term memory for arbitrary relations in two California sea lions (Zalophus californianus)*. Paper presented at the annual meeting of the Animal Behavior Society, Seattle.
- WRIGHT, A. A., COOK, R. G., RIVERA, J. J., SANDS, S. F., & DELIUS, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning & Behavior*, **16**, 436-444.
- WRIGHT, A. A., SHYAN, M. R., & JITSUMORI, M. (1990). Auditory same/different concept learning by monkeys. *Animal Learning & Behavior*, **18**, 287-294.
- ZENTALL, T. R., & HOGAN, D. E. (1974). Abstract concept learning in the pigeon. *Journal of Experimental Psychology*, **102**, 393-398.

(Manuscript received April 5, 1993;
revision accepted for publication December 9, 1993.)