

Discriminating the Relation Between Relations: The Role of Entropy in Abstract Conceptualization by Baboons (*Papio papio*) and Humans (*Homo sapiens*)

Joël Fagot

Center for Research in Cognitive Neurosciences

Edward A. Wasserman

Center for Research in Cognitive Neurosciences
and University of Iowa

Michael E. Young

University of Iowa

Two baboons (*Papio papio*) successfully learned relational matching-to-sample: They picked the choice display that involved the same relation among 16 pictures (same or different) as the sample display, although the sample display shared no pictures with the choice displays. The baboons generalized relational matching behavior to sample displays created from novel pictures. Further experiments varying the number of sample pictures and the mixture of same and different sample pictures suggested that entropy plays a key role in the baboons' conceptual behavior. Two humans (*Homo sapiens*) were similarly trained and tested; their behavior was both similar to and different from the baboons' behavior. The results suggest that animals other than humans and chimpanzees can discriminate the relation between relations. They further suggest that entropy detection may underlie same–different conceptualization, but that additional processes may participate in human conceptualization.

It would be very difficult . . . to determine how far animals exhibit any traces of . . . high mental powers. . . . If one may judge from various articles which have been published lately, the greatest stress seems to be laid on the supposed entire absence in animals of the power of abstraction, or of forming general concepts. (Darwin, 1871/1920, p. 84)

Charles Darwin penned these words 130 years ago in connection with his bold but then unsubstantiated claim that the difference in

mind between humans and other animals was one of degree and not of kind. What have we learned about the animal mind, in general, and about abstract conceptualization by animals, in particular, over the ensuing years?

First, it is still very difficult to determine the mental powers of nonhuman animals; however, a century of psychological science has made measurable progress in revealing some of the secrets of animal intelligence (Wasserman, 1993, 1997). The creative application of conditioning and computing methodologies now permits behavioral scientists to investigate animal intelligence with unprecedented power and precision.

Second, investigating the abstract conceptual abilities of nonhuman animals remains vitally important to understanding the evolution of human intelligence (Premack, 1983; Thompson, 1995). New behavioral investigations are prompting researchers to reconsider the common belief that animals entirely lack any abstract conceptual abilities.

We have recently discovered that, when they are trained and tested under similar experimental conditions, animals from widely different species can acquire a same–different concept. In one illustrative example (Wasserman, Hugart, & Kirkpatrick-Steger, 1995; Young & Wasserman, 1997, Experiment 1), pigeons were first taught to peck one button when they viewed an array of computer icons that comprised 16 copies of the same icon and to peck a second button when they viewed an array that comprised one copy of 16 different icons (a same–different discrimination task). These same and different training displays were created from one set of 16 computer icons. The pigeons were later tested with new same and new different displays that were created from a

Joël Fagot, Comparative Cognition Research Group, Center for Research in Cognitive Neurosciences; Edward A. Wasserman, Comparative Cognition Research Group, Center for Research in Cognitive Neurosciences and Department of Psychology, University of Iowa; Michael E. Young, Department of Psychology, University of Iowa.

Michael E. Young is now at the Department of Psychology, Southern Illinois University at Carbondale.

We thank Raymond Fayolle, France Lavergne, and Monique Chiambretto for their assistance in conducting the experiments. We acknowledge the Office of the Provost of the University of Iowa, the Center for Research in Cognitive Neurosciences of the National Center for Scientific Research in Marseille, the Human Frontier Science Program, Le Programme Interdépartemental "Origine de l'Homme, du Langage" of Le Centre National de la Recherche Scientifique, and National Science Foundation Grant 9904569 for their financial support to conduct the reported research. Data from Experiments 1 and 3 were previously presented at the March 2000 Conference on Comparative Cognition, Melbourne Beach, FL.

Correspondence concerning this article should be addressed to Joël Fagot, Center for Research in Cognitive Neurosciences, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France, or Edward A. Wasserman, Department of Psychology, University of Iowa, Iowa City, Iowa 52242-1407. Electronic mail may be sent to fagot@lnf.cnrs-mrs.fr or ed-wasserman@uiowa.edu.

second set of 16 computer icons that had never before been shown during discrimination training. Accuracy to the training stimuli averaged from 83% to 93% correct, and accuracy to the testing stimuli averaged from 71% to 79% correct; in each case, choice accuracy reliably exceeded the chance score of 50% correct. Such robust discrimination learning and stimulus generalization attest to the pigeon's acquisition of an abstract same-different concept (for more on the nature of this concept, see Wasserman, Young, & Nolan, 2000; Young & Wasserman, 1997; Young, Wasserman, & Dalrymple, 1997; Young, Wasserman, & Garner, 1997).

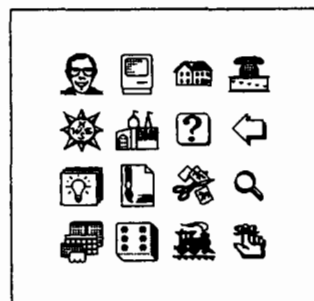
In a second illustrative example (Wasserman, Fagot, & Young, 2001), baboons were similarly trained and tested with the same visual stimuli. Accuracy to the training stimuli averaged 91% correct, and accuracy to the testing stimuli averaged 81% correct; in each case, choice accuracy reliably exceeded the chance score of 50% correct.

It is a highly advanced intellectual feat for animals like pigeons and baboons to detect the sameness or differentness of a collection of visual stimuli and to make two distinctively different responses in order to report those same-different relations (Deliuss, 1994). An even more advanced feat would be for animals to match the relation between relations—in other words, to exhibit the essence of analogical reasoning (Premack, 1983; Thompson & Oden, 2000).

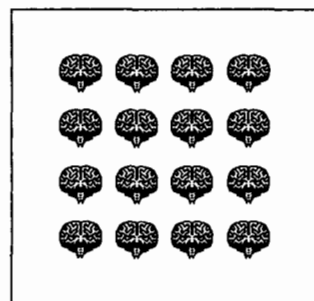
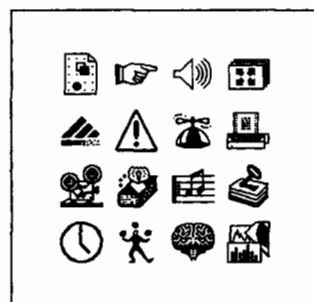
Consider the task that is schematized in Figure 1. Here, an animal is shown a sample display that depicts either all same icons or all different items. Following the presentation of the sample display, two choice displays are shown: the first of all different items and the second of all same items. The correct response is to select the choice display that involves the same relation among the items as was shown in the sample display.

In the case of the trial that is depicted in Figure 1, the correct response should be to the different choice display because the sample display involved icons that were also different from one another; these two displays are, in other words, alike in their differentness. Critically, none of the icons that was shown in the sample display was shown in either of the two choice displays. So, only the relations among the items in the sample and the choice displays are available to guide discriminative responding; the particular items themselves are of no informational importance. If a nonhuman animal were able to learn this so-called relational matching-to-sample task, then this learning would constitute what some theorists consider to be the strongest evidence of abstract conceptual behavior (Thompson, 1995). To date, only human beings and chimpanzees have successfully learned relational matching-to-sample (see Thompson, Oden, & Boysen, 1997, for an illustrative example; see Thompson & Oden, 2000, for a review of the literature).

A same-different relational matching task was presented to baboons in the present study. Our initial objective was to see if the baboon could successfully solve that task (which would be unprecedented for a nonape species) and to explore the cognitive substrate of this ability. To that aim, we were interested in whether the baboon could generalize its matching-to-sample behavior to sample displays that contained completely novel visual items. Successful generalization to novel sample displays would provide clear evidence that these monkeys had indeed learned a highly abstract and broadly applicable concept.



Sample (displayed first)



Choice (displayed second)

Figure 1. An example of a 16-icon different sample array (top box) and examples of the 16-icon choice arrays (bottom two boxes). The correct response for this trial should be to the different choice display because the sample display involved icons that were also different from one another. The actual stimuli appeared as white-on-black images.

Because we found just such successful discrimination learning and stimulus generalization based on the relation between relations, we proceeded more deeply to explore the cognitive basis of this relational matching behavior. Our follow-up research investigated (a) the effects of the number of items in a sample (and a test) display and (b) the effects of the mixture of same icons and different items in a sample display on relational matching-to-sample behavior. That follow-up research suggested that the ba-

boons were strongly controlled by the variability of the items that were shown in the sample stimulus displays.

Finally, because our experimental methods were novel, for comparative purposes, we trained and tested human beings on the same discrimination tasks that we had given to baboons. These behavioral comparisons between humans and baboons showed both similarities and differences in the cognitive processes of abstract conceptual behavior.

Experiment 1

Can baboons learn a relational matching-to-sample task? If so, then can they generalize that learning to new exemplars? We addressed these two questions in our first experiment. Humans, too, served as research participants in our first experiment, so that we could better judge the behavioral achievements of our nonhuman primates.

Method

Subjects

We studied 1 male (i.e., B03) and 1 female (i.e., B08) adult Guinea baboon (*Papio papio*) who lived in the same social group at the animal facility of the Center for Research in Cognitive Neurosciences, Marseille, France. The 2 baboons were naïve with respect to the relational (same-different) matching-to-sample problem. However, the baboons had previously been used in the Wasserman et al. (2001) same-different discrimination learning project, which involved the same types of pictorial stimuli as in the present study. In the earlier study, the baboons viewed only one display at a time, and they were required to make one report response when that display involved 16 identical items and to make a different report response when that display involved 16 nonidentical items.

As a result of their prior participation in several other experiments, the baboons were also familiar with the matching-to-sample procedure involving identical sample and choice stimuli and with the joystick system used in this experiment (see Fagot & Deruelle, 1997; Fagot, Kruschke, Dépy, & Vauclair, 1998). The baboons were not food deprived, but they received their food ration after the completion of the training or testing sessions.

For comparative purposes, we also trained and tested one woman (H01) and one man (H02) under similar circumstances. The humans (*Homo sapiens*) volunteered to participate, and they were naïve with respect to the general purpose of the experiment.

Apparatus

The baboons were studied inside an experimental enclosure (68 × 50 × 72 cm) facing an analog joystick, a metal touch pad, and a 14-in. color monitor that was driven by a Pentium 133 PC computer. On the front of the enclosure was a view port, a hand port, and a food dispenser; the food dispenser delivered 190-mg banana-flavored Noyes food pellets into the enclosure in accordance with the prevailing contingencies of reinforcement. Manipulation of the joystick induced isomorphic displacements of a cursor on the monitor screen. The distance between the view port and the monitor was 49 cm.

Human participants were tested with a joystick system that was in all respects identical to that of the baboons, except that the monitor screen and the joystick were placed on a table. To equate the viewing distance between species, the humans had to place their chins on a small bar that was located 49 cm from the monitor screen.

Stimuli

Seventy-two highly distinguishable computer icons, each averaging 1.1×1.1 cm ($0.9 \times 0.9^\circ$ of visual angle), were chosen as the total item

pool and were later randomly sorted into three sets of 24: Set 1, Set 2, and Set 3. The icons appeared as white-on-black pictures on a black background. Sets 1 and 2 were the same as those used by Wasserman et al. (2001); Set 3 contained 24 completely novel icons. Two types of stimulus arrays were created from each of these icon sets: same arrays and different arrays. The same arrays were created by the random selection of a single icon from the designated stimulus set (e.g., Set 1), which was repeated 16 times and randomly distributed in the 16 cells of an unmarked 4×4 matrix; the different arrays were composed of 16 different icons, which were randomly selected from a common stimulus set and randomly distributed in the 4×4 matrix. A thin white square line (7.5 cm wide × 7.5 cm high) framed the stimulus presentation matrix on the otherwise black screen.

Procedure

General procedure. The baboons were isolated from their social group and individually placed into the test apparatus. The experiment involved multiple sessions per day of many trials that involved a two-alternative forced-choice matching-to-sample procedure. The baboons initiated trials by contacting the touch pad with one hand. That action induced the display of a circular white cursor (0.5 cm in diameter) in the center of the monitor, along with a white 0.5×0.5 cm square fixation point (FP) stimulus that was randomly placed 1.5 cm above or below the cursor. By manipulating the joystick, the baboons were required to place the cursor on the FP stimulus to initiate the presentation of the same or different sample stimulus array. This procedure was adopted to ensure that the animals were attending to the screen at the beginning of each trial (Wilde, Vauclair, & Fagot, 1994). Once the cursor was on the FP, a same or a different sample array appeared for 500 ms in either the left or the right half of the monitor (Figure 1, top); randomly locating the sample stimulus array to the left or to the right of the FP was also adopted to ensure the animals' continued attention. Immediately after sample stimulus offset, two choice arrays appeared on the vertical axis of the screen (Figure 1, bottom). One of these two choice stimuli was a same array; it involved a single icon that was repeated 16 times in the 16-cell matrix. The other choice stimulus was a different array; it involved 16 different icons that were randomly located in the 4×4 matrix. The location of the same and different choice arrays, either at the top or bottom of the screen, randomly varied across trials.

Of utmost importance for our research was the fact that the stimulus set that was used to create the sample stimuli differed from the stimulus set that was used to create the choice stimuli. This difference meant that correct choice responding could not be based on the identity of the individual icon(s) in the sample and the choice arrays; such identity was impossible. Correct responding could only be based on the relation of the icons in the sample and the choice arrays.

The choice stimuli remained on the screen for a maximum of 10 s, during which the baboons could make a single choice response by moving the cursor into contact with one or the other choice array. Failures to make a choice response in the allotted time were rare during training and never occurred during testing. Correct responses were followed by a high tone and one food pellet; incorrect responses were followed by a low tone and a 7-s time-out. After food delivery or time-out, the next trial could be initiated by contact of the touch pad following a 6-s intertrial interval.

The procedure for the human participants was the same as that for the baboons, except that the reinforcement procedure involved the 800-ms display of the French words *vrai* or *faux* (true or false) immediately after one of the two choice arrays had been selected. For all of the subjects, either baboons or humans, Experiment 1 involved a training phase followed by a testing phase.

Training. Training consisted of multiple 96-trial sessions, each containing 48 trials in which the sample array showed the "same" relation and 48 trials in which the sample array showed the "different" relation. The baboons received from two to four training sessions per day with training taking place 5 days per week. During training, half of the sample stimuli

were made from Stimulus Set 1; the other half of the sample stimuli were made from Stimulus Set 2. The icons that were used to create the choice arrays were randomly selected from Set 2 whenever the sample was created from Set 1; the icons that were used to create the choice arrays were randomly selected from Set 1 whenever the sample was created from Set 2. Training sessions were repeated until the subjects reached 80% correct across both same and different displays for two consecutive sessions. A correction procedure was adopted during training; whenever the baboons made an error, the identical trial was immediately repeated, with the constraint that no more than five correction trials in a row could be presented. Correction trials were not scored. The order of same trials and different trials within a session was randomized and varied from one session to the next and from one baboon to another.

At the beginning of training, the human participants were told that they had to manipulate the joystick to select one of the two choice arrays. The participants received no instruction on the matching rule that they had to apply; they themselves had to discover it.

Testing. Each baboon received three testing sessions of 128 trials each. The testing sessions contained 96 baseline training trials (48 same and 48 different) that were like those of the training sessions (except for the absence of correction trials, which were unnecessary to maintain excellent discriminative behavior) plus 32 testing trials (16 same and 16 different trials) in which the sample display was created from Set 3, and the comparisons were created either from Set 1 (for half of the testing trials) or from Set 2 (for the other half of the testing trials). All of these different types of trials were randomly presented in the testing sessions. Each baboon thus received 288 baseline training trials (144 same and 144 different) plus 96 testing trials (48 same and 48 different) involving the novel set of icons (i.e., Set 3). For both baboons and humans, differential reinforcement for "same" and "different" report responses was given on the baseline training trials. Nondifferential reinforcement was given on the testing trials; food was delivered (for baboons) and the word *vrai* was displayed (for humans) regardless of the choice response. All other aspects of the task were identical to those of the training phase.

Results

Baboons

The acquisition of relational matching-to-sample proceeded gradually. It took B03 4,992 trials to respond consistently in excess of 80% correct, and it took B08 7,104 trials to do so. Over the final four sessions (384 trials) of relational matching-to-sample training, B03 and B08 each averaged 84% correct.

The key findings that emerged during the three subsequent testing sessions are shown in the left column of Figure 2. Across both baboons and all three sessions of testing, discriminative performance remained high to the sample stimuli that were created from the two familiar sets of computer icons—a mean of 84% correct was recorded. Across both baboons and all three sessions of testing, discriminative performance was also high to the sample stimuli that were created from the novel set of computer icons—a mean of 70% correct was recorded. Each of these two accuracy scores differed reliably from the 50% value that was expected by chance ($z = 18.94$, $z = 8.40$, both $ps < .001$, respectively). Also reliably higher than chance were (a) the accuracy scores of each individual baboon to both the familiar and the novel sample stimuli across all three sessions of testing ($zs = 12.26$ and 10.72 for the familiar stimuli, $zs = 3.67$ and 4.29 for the novel stimuli for B03 and B08, respectively; $ps < .001$) as well as on each of the three individual sessions of testing (all zs were between 2.12 and 7.76, $ps < .05$) and (b) the accuracy scores to both the same and the

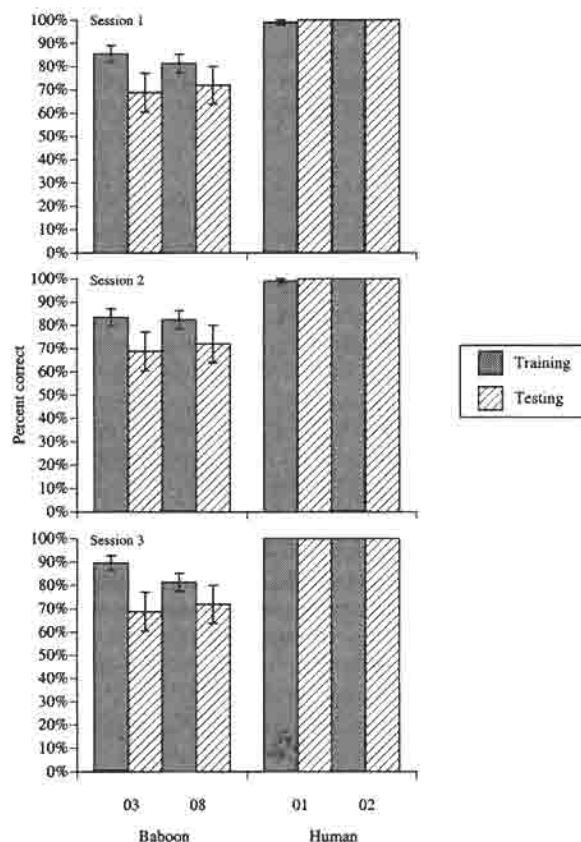


Figure 2. Mean discriminative performance on training and testing trials in the three testing sessions of Experiment 1. The left column shows the performance by the baboons (03 and 08). The right column shows the performance by the humans (01 and 02). The figure shows the means and standard errors of the percentage correct data.

different sample stimuli on both the familiar and the novel sample trials across both baboons and all three sessions of testing ($zs = 12.02$ and 10.96 for the familiar stimuli; $zs = 6.12$ and 1.84 for the novel stimuli for same and different samples, respectively; $ps < .05$).

Although choice accuracy was high to the sample stimuli that were created from the novel set of computer icons, accuracy to the sample stimuli that were created from the two familiar sets of computer icons was reliably higher—B03: $t(510) = 4.69$, $p < .01$; B08: $t(510) = 3.13$, $p < .01$. It should be remembered that choice responding was differentially reinforced on the training trials that involved the two familiar sets of sample icons, whereas choice responding was nondifferentially reinforced on the testing trials that involved the novel set of sample icons. Any effects of non-differential reinforcement should have acted against reliable transfer to the testing stimuli.

Humans

The acquisition of the matching-to-sample task proceeded very quickly. It took H01 and H02 fewer than 96 trials to respond consistently in excess of 80% correct. During the second session

(96 trials) of relational matching-to-sample training, H01 and H02 averaged 97% and 100% correct, respectively.

The key findings that emerged during the three subsequent testing sessions are shown in the right column of Figure 2. Across both people and all three sessions of testing, discriminative performance remained high to the sample stimuli that were created from the two familiar sets of computer icons—a mean of 99% correct was recorded. Across both people and all three sessions of testing, discriminative performance was also high to the sample stimuli that were created from the novel set of computer icons—a mean of 100% correct was recorded. Each of these accuracy scores differed reliably from the 50% value that was expected by chance ($z = 23.83$, $z = 13.67$, both $ps < .0001$). Also reliably higher than chance were (a) the choice scores of each individual person to both the familiar and the novel sample stimuli across all three sessions of testing ($zs = 16.73$ and 16.97 for the familiar stimuli; $zs = 9.75$ and 9.64 for the novel stimuli for H01 and H02, respectively; all $ps < .0001$) as well as on each of the three sessions of testing (all zs were between 5.48 and 9.80 , all $ps < .001$) and (b) the choice scores to both the same and the different sample stimuli on both the familiar and the novel sample trials across both people and all three sessions of testing ($zs = 16.73$ and 16.97 for the familiar stimuli; $zs = 9.75$ and 9.64 for the novel stimuli for same and different samples, respectively; all $ps < .0001$).

Choice accuracy was equally high to the sample stimuli that were created from the novel set of computer icons and to the sample stimuli that were created from the two familiar sets of computer icons—H01: $t(381) = -0.81$, $p > .25$; H02: $t(379) = 0.00$, $p > .25$.

Discussion

Our primary objective in this first experiment was to see if a nonape, primate species, the baboon, could successfully solve a task that appears to require the discrimination of an abstract same–different relation. The results were clear: Our 2 baboons successfully acquired relational matching-to-sample to levels in excess of 80% correct. The task was not easy for them; it took B03 approximately 5,000 trials to learn, and it took B08 approximately 7,000 trials to do so. Earlier, the animals had learned to report 16-icon same versus different displays by making one choice response (e.g., “up”) for same displays and a different choice response (e.g., “down”) for different displays (Wasserman et al., 2001); it took B03 only 700 trials to learn, and it took B08 only 600 trials to do so. Relational matching-to-sample is obviously much more difficult for baboons to acquire than is same–different discrimination learning.

We were also interested in whether the baboons could generalize their relational matching-to-sample behavior to novel sample displays. Here, too, the results were clear. On the very first session of testing, the 2 baboons discriminated the displays of novel sample items at a mean level of 70% correct, whereas they discriminated the displays of familiar sample items at a mean level of 83% correct. Two subsequent sessions of testing yielded very similar levels of discriminative performance. Such robust generalization to the novel testing displays supports the hypothesis that these monkeys had indeed learned an abstract and general concept.

The lower choice accuracy that was observed to the novel sample (testing) displays than to the familiar sample (training)

displays suggests that the baboons may have processed and remembered some or all of the 48 individual icons with which they had become acquainted during training (for similar generalization results in pigeons' same–different discrimination learning, see Wasserman et al., 1995; Young & Wasserman, 1997; for similar generalization results in baboons' same–different discrimination learning, see Wasserman et al., 2001); memorization of the individual icons is unsurprising, given the extensive training that was necessary for the baboons to learn relational matching-to-sample in the first place.

Compared with our baboons, our human research participants (a) learned relational matching-to-sample far faster, (b) discriminated the same sample stimuli from the different sample stimuli at ever higher levels of accuracy, and (c) completely generalized their discrimination to sample displays of novel computer icons.

Experiment 2

When we set out to study same–different conceptualization in the baboon, we modeled our task along the same lines that we had used earlier in our investigations of the pigeon; we used large arrays of pictorial items because prior work in other laboratories had been rather unsuccessful in finding robust same–different conceptualization in the pigeon. How critical was it for obtaining the baboons' successful discrimination and generalization data in Experiment 1 that our sample displays involved 16 rather than 2 items?

To find out, in Experiment 2, we varied the number of items in both same sample displays and in different sample displays (we held the number of items in the choice displays constant at 16). If presenting fewer sample items serves to distill the relational matching-to-sample task to its essence, then we might expect even more accurate relational matching performance with fewer sample items than we obtained with more sample items. However, if there were something special about larger numbers of sample items, then accuracy might fall when the number of sample items is reduced for baboons as it did for pigeons (Young, Wasserman, & Garner, 1997). We again tested human participants so that we could have another comparative data set for our baboons.

Method

Subjects

The same 2 baboons¹ and the same 2 humans² served in this experiment as had served in Experiment 1.

Apparatus and Stimuli

The apparatus was the same as in Experiment 1. Only icons from Stimulus Set 1 and Stimulus Set 2 were used to create the sample and choice arrays.

¹ Experiment 2 actually followed Experiment 3; the chronological order has been reversed for expository purposes. Experiments 1, 2, and 3 were conducted without a break over a several week period.

² Experiment 2 actually followed Experiment 3; the chronological order has been reversed for expository purposes. All three experiments were completed in 2 consecutive days. The order of training and testing was the same for the humans as it was for the baboons.

Testing Procedure

Each baboon received eight testing sessions of 128 trials each, two sessions daily on 4 successive days. Testing sessions involved 96 baseline training trials (48 same trials and 48 different trials) with 16 sample icons that were like those of the baseline training sessions of Experiment 1 (also without correction trials), randomly intermixed with 32 testing trials during which the sample display contained either 2, 4, 8, or 12 icons (4 same trials and 4 different trials for each display size) and the choice displays contained 16 icons. Regardless of the number of sample icons shown on each trial, the location of each sample icon within the 4×4 matrix was randomly determined and varied from one trial to the next. Baseline and testing trials were randomly presented in the testing sessions. Each subject thus received 768 baseline training trials (384 same and 384 different) plus 256 testing trials (32 per test condition). Differential food reinforcement for "same" and "different" report responses was given on the baseline training trials; nondifferential food reinforcement was given on the testing trials. Analogous testing procedures were given to the human participants.

Results and Discussion

Baboons

As we decreased the number of sample icons from 16 to 12 to 8 to 4 to 2, the overall accuracy of relational matching performance fell to the chance score of 50% correct for both baboons: The corresponding percentage correct response scores were 86%, 73%, 80%, 63%, and 47% for B03, and 82%, 77%, 58%, 50%, and 52% for B08.

This decrease in overall choice accuracy might be attributed to the novelty of the testing displays that contained fewer than 16 items; as the number of sample items is progressively decreased, the novelty of the display is increased, thereby progressively lowering accuracy. Although there may be some truth to this generalization decrement interpretation, it cannot be the whole story. Accuracy did not fall symmetrically on both same sample trials and on different sample trials as it should have according to a generalization decrement account; instead, accuracy was more adversely affected on different sample trials than it was on same sample trials. The top of Figure 3 shows that both baboons correctly reported "same" on same sample trials regardless of the number of icons that were contained in the sample; the baboons correctly reported "different" on different sample trials with eight or more icons, but they incorrectly reported "same" on different sample trials with four (B08) or two (B03 and B08) icons.

To confirm the differential effect of the number of sample icons on relational matching accuracy, we performed individual subject full factorial analyses of variance (ANOVAs) of accuracy as a function of array type (same vs. different) and sample icon number (2, 4, 8, 12, or 16). The analysis of B03's behavior revealed significant main effects of number, $F(4, 1014) = 20.16, p < .0001$; and type, $F(1, 1014) = 13.38, p < .001$; and a significant Number $\times$ Type interaction, $F(4, 1014) = 3.98, p < .001$. The analysis of B08's behavior also revealed significant main effects of number, $F(4, 1014) = 19.83, p < .0001$; and type, $F(1, 1014) = 7.25, p < .01$; and a significant Number $\times$ Type interaction, $F(4, 1014) = 14.27, p < .0001$. Follow-up tests of the simple effects for both baboons revealed significant differences across icon numbers on different sample trials but not on same sample trials ($p < .0001$ for different samples, $p > .05$ for same samples).

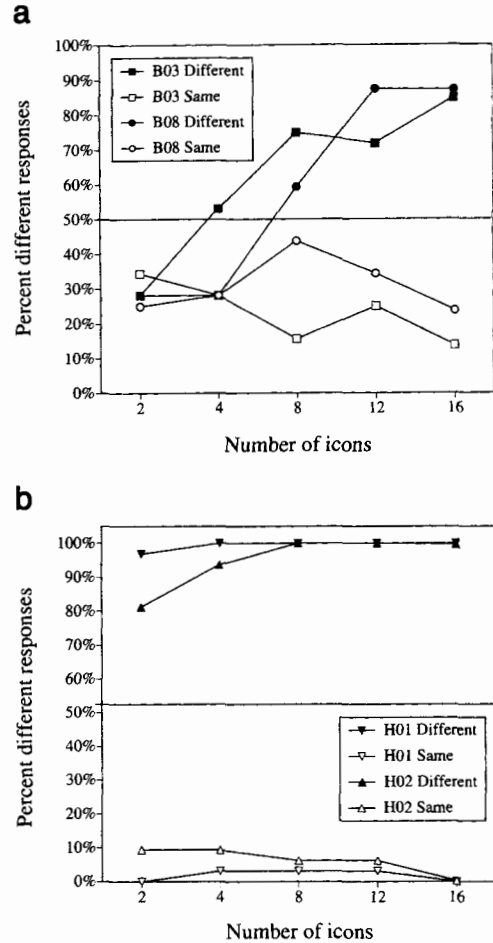


Figure 3. Mean discriminative performance in Experiment 2 plotted as a function of the number of icons in the sample display. The top graph shows the performance by the baboons (B03 and B08). The bottom graph shows the performance by the humans (H01 and H02). The figure shows the means of the percentage correct data.

This peculiar pattern of performance is precisely the one that we had observed earlier with pigeons that were trained on a same-different discrimination task (Young, Wasserman, & Garner, 1997); importantly, this pattern is understandable if both pigeons and baboons discriminate the amount of variability that is depicted in the visual displays. The variability of the displays can be derived by computing their entropy (Shannon & Weaver, 1949).³

The 16-item same and different samples respectively represent the minimal (entropy = 0) and the maximal (entropy = 4) amount of variability that can be shown in displays of 16 items. Reducing the number of items in same sample displays should not change the amount of pictured variability. There is still none (entropy = 0).

³ The formula for entropy is

$$H(A) = - \sum_{a \in A} p_a \log_2 p_a$$

where $H(A)$ is the entropy of categorical variable A , a is a category of A , and p_a is the proportion of observed values within that category.

However, reducing the number of items in different sample trials should reduce the amount of pictured variability; in fact, with only two pictured items, the variability of different sample displays (entropy = 1) is numerically closer to 16-item same samples (entropy = 0) than it is to 16-item different samples (entropy = 4). Such variability discrimination closely conforms to the relational matching data that are shown in Figure 3.

Humans

As the number of icons decreased from 16 to 12 to 8 to 4 to 2, the overall accuracy of relational matching performance fell to a slight degree for both people (see bottom of Figure 3): The corresponding percentage correct response scores were 100%, 98%, 98%, 98%, and 98% for H01, and 100%, 97%, 97%, 92%, and 86% for H02.

To explore the effect of the number of icons on choice accuracy, we performed individual participant full factorial ANOVAs of accuracy as a function of array type (same vs. different) and number (2, 4, 8, 12, or 16). The analysis of H01's responding revealed main effects of number, $F(4, 1014) = 3.07, p < .05$; and type, $F(1, 1014) = 4.00, p < .05$; and a significant Number $\times$ Type interaction, $F(4, 1014) = 3.83, p < .01$. The analysis of H02's responding also revealed a main effect of number, $F(4, 1014) = 19.73, p < .0001$; and a significant Number $\times$ Type interaction, $F(4, 1014) = 3.92, p < .01$, but the main effect of type was not significant ($F < 1$). Follow-up tests of the simple effects for both people found significant effects of icon number on both same and different sample trials (all $ps < .05$).

These statistical results notwithstanding, the effect of the number of sample icons on humans' relational matching-to-sample was numerically small; more importantly, the human participants failed to evidence the striking asymmetry in responding on same and different trials that was shown by our baboons.

Experiment 3

The results of Experiment 2 implicated sample stimulus variability or entropy in the control of our human and nonhuman participants' relational matching-to-sample behavior. Especially for the baboons, the precise parametric effects of varying the number of sample items are difficult to explain by any other account of which we are aware.

Another parametric look at stimulus control by pictorial variability or entropy can be obtained by systematically manipulating the mixture of same items and different items in a 16-item sample, thereby eliminating the involvement in the obtained results of a stimulus generalization decrement due to a change in sample item number (Young & Wasserman, 1997). In Experiment 3, we did just that. In fact, we manipulated the combination of same icons and different icons within the 16-icon sample stimuli in two entirely different ways to distinguish stimulus control by entropy from stimulus control by other features of the sample displays, such as the number of different kinds of items in those displays. Baboons and humans were again studied under similar experimental circumstances.

Method

Subjects

The same 2 baboons and the same 2 humans served in this experiment as had served in Experiments 1 and 2.

Apparatus and Stimuli

The apparatus was the same as in Experiments 1 and 2. Only icons from Stimulus Set 1 and Stimulus Set 2 were used to create the sample and choice stimuli.

To assess the effect of sample icon mixture on performance, two different kinds of sample mixtures were given during the testing trials. The first series (Mixture Series 1) involved an additive change in the preponderance of different icons and same icons in a 16-icon sample display: $m(\text{Different}) + n(\text{Same}) = 16$. Five different kinds of sample mixtures were given in this series: 2D/14S, 4D/12S, 8D/8S, 12D/4S, and 14D/2S. These five types of sample mixtures respectively involved 3, 5, 9, 13, and 15 different kinds of icons. Thus, entropy and the number of kinds of icons covaried in this mixture series.

The second series (Mixture Series 2) also arranged mixtures of same and different sample icons, except that exactly four kinds of icons were used in order to create each sample display. This series contained four kinds of sample mixtures: 13-1-1-1 (i.e., 13 repetitions of the first icon, 1 presentation each of the other three), 10-3-2-1, 7-5-3-1, and 4-4-4-4. Here, entropy and the number of kinds of icons did not covary in this mixture series.

Testing Procedure

Testing with Mixture Series 2 preceded testing with Mixture Series 1. Eight testing sessions were conducted with both Mixture Series 1 and Mixture Series 2. The testing sessions consisted of 96 baseline trials randomly intermixed with 4 trials per kind of sample mixture. There were thus 96 baseline trials plus 20 test trials (total = 116) for the test involving Mixture Series 1, and 96 baseline trials plus 16 test trials (total = 112) for the test involving Mixture Series 2. Differential food reinforcement for "same" and "different" report responses was given on baseline training trials; nondifferential food reinforcement was given on mixture testing trials. Analogous testing procedures were given to the human participants.

Results and Discussion

Baboons

The baboons' relational matching-to-sample performance was strongly affected by the mixtures of same and different sample items that we presented.

Mixture Series 1 testing sessions included same, 2D/14S (2 different and 14 same icons), 4D/12S, 8D/8S, 12D/4S, 14D/2S, and different trials; the corresponding percentage of different report responses was 15%, 13%, 16%, 50%, 72%, 81%, and 84% for B03, and 19%, 31%, 66%, 72%, 81%, 75%, and 81% for B08.

Mixture Series 2 testing sessions included same, 13-1-1-1 (13 repetitions of one icon type, and 1 of each of three other icon types), 10-3-2-1, 7-5-3-1, 4-4-4-4, and different trials; the corresponding percentage of different report responses was 5%, 19%, 28%, 25%, 41%, and 86% for B03, and 18%, 25%, 52%, 61%, 68%, and 75% for B08.

These choice scores are plotted in Figure 4 as a function of entropy. This plot enables us to place both methods of manipulating sample display variability along a single quantitative scale.

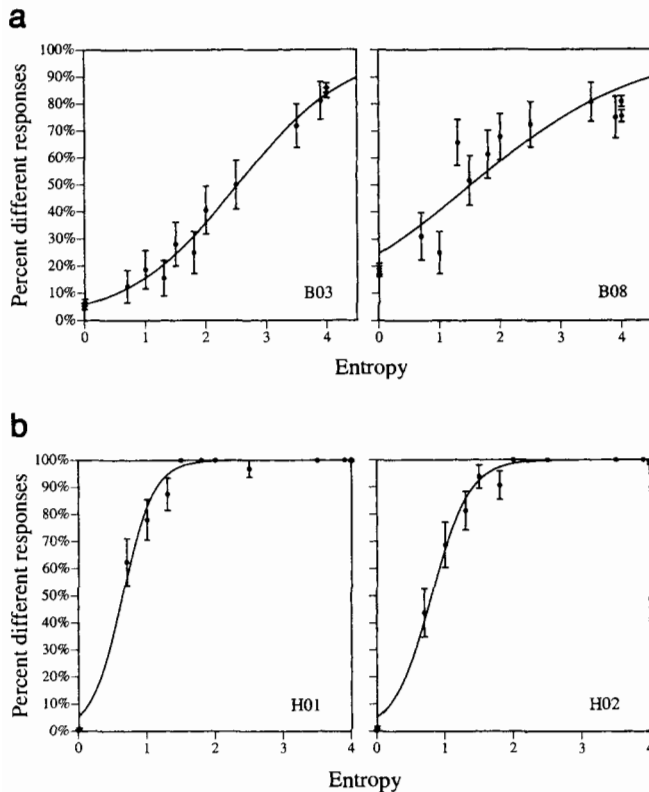


Figure 4. Mean discriminative performance in Experiment 3 plotted as a function of sample display entropy. The top graph shows the performance by the baboons (B03 and B08). The bottom graph shows the performance by the humans (H01 and H02). The figure shows the means and standard errors of the percentage correct data and includes a sigmoid curve fit that was produced using nonlinear regression.

Each of the baboons showed strong control by entropy, with the percentage of different responses systematically rising with increases in entropy.

We performed nonlinear regression analyses that involved fitting a two-parameter sigmoid to the data (for methodological details, see Young & Wasserman, 2001). The *gradient* of the sigmoid corresponded to the degree of categorical distinction that was evidenced; gradient is a measure of the slope of the middle of the sigmoid. High gradients indicate that the subject made a sharp distinction along the different levels of entropy (e.g., between entropies greater than 2 and those less than 2), whereas low gradients indicate that the subject evidenced a more gradual distinction along the different levels of entropy. The *boundary* of the sigmoid corresponded to the psychological midpoint of the entropy scale (i.e., the point at which the subject is predicted to make a different report response 50% of the time). Both gradient and boundary were determined by fitting the following equation to the standardized data:

$$\frac{1}{1 + e^{-\text{gradient} \times (\text{entropy} - \text{boundary})}} \quad (1)$$

These regression analyses revealed that, for B03, entropy accounted for 99% of the variance; for B08, entropy accounted for

84% of the variance. Each of these fits was statistically significant ($p < .0001$). Furthermore, entropy fit the data better than did other possible mathematical properties of the sample displays: the number of icon types, the frequency of the most prevalent icon, the frequency of the least prevalent icon, and the mean frequency of all of the icon types that were presented in a sample display (contact the authors for analysis details). This superiority of fit is especially notable because these other properties of the sample displays were sometimes very highly correlated with one another and with entropy. (The gradient and boundary results follow in the *Parameter Comparisons* section.)

Humans

The humans' relational matching-to-sample performance was also strongly affected by the mixtures of same and different sample items that we presented. Mixture Series 1 testing sessions included same, 2D/14S, 4D/12S, 8D/8S, 12D/4S, 14D/2S, and different trials; the corresponding percentage of different report responses was 0%, 62%, 87%, 97%, 100%, and 100% for H01, and 1%, 44%, 81%, 100%, 100%, and 99% for H02. Mixture Series 2 testing sessions included same, 13-1-1-1, 10-3-2-1, 7-5-3-1, 4-4-4-4, and different trials; the corresponding percentage of different report responses was 0%, 78%, 100%, 100%, 100%, and 100% for H01, and 0%, 69%, 94%, 91%, 100%, and 99% for H02.

These choice scores are plotted in Figure 4 as a function of entropy. Each of the human participants showed strong control by entropy, with the percentage of different responses systematically rising with increases in entropy. Nonlinear regression analyses revealed that, for both participants, entropy accounted for 99% of the variance. Each of these fits was statistically significant ($p < .0001$). Furthermore, entropy fit the data better than did other possible mathematical properties of the sample displays: the number of icon types, the frequency of the most prevalent icon, the frequency of the least prevalent icon, and the mean frequency of all of the icon types that were presented in a sample display (contact the authors for analysis details).

Parameter Comparisons

The gradients of the fitted sigmoids were 1.10, 0.73, 4.40, and 3.60 for B03, B08, H01, and H02, respectively. The higher gradients for our human participants revealed a more categorical (abrupt) distinction between low and high levels of entropy than was revealed by our baboon subjects.

The boundaries of the fitted sigmoids were 2.52, 1.52, 0.64, and 0.80 for B03, B08, H01, and H02, respectively. Thus, the psychological midpoint of the entropy scale was much lower for humans (less than 1) than it was for baboons; indeed, at an entropy level of 1, both baboons reported "same" about 80% of the time, whereas both humans reported "different" about 70% of the time.

Experiment 4

As a final issue in this series of experiments, we explored whether we could eventually train the baboons to respond discriminatively to same displays and different displays that contained fewer than 16 sample items—particularly fewer than 8 sample items (see the top of Figure 3)—because they did not spontane-

ously do so in Experiment 2. We investigated this issue in the final two experiments of this study because the results of varying the mixture of same icons and different icons might have been dramatically altered had we conducted Experiment 3 at the end of the series.

We began Experiment 4 by training the baboons with 16-icon sample displays. We then progressively decreased the number of sample icons when the baboons were responding at a high level of discrimination accuracy on both same and different sample trials. To help the baboons in this stepwise acquisition process, we sometimes included training sessions that retained the identical number of icons as had been given in the prior step.

Method

Subjects

The same 2 baboons served here as had served in Experiments 1, 2, and 3. Approximately 2 months passed between Experiments 3 and 4, during which time the baboons were on holiday from the experiments.

Apparatus and Stimuli

The apparatus was the same as in Experiments 1, 2, and 3. Only icons from Stimulus Set 1 and Stimulus Set 2 were used to create the sample and choice stimuli.

Training Procedure

Training consisted of multiple 96-trial sessions, each containing 48 trials in which the sample array involved the "same" relation plus 48 trials in which the sample array involved the "different" relation. Correction trials were again included because we expected the task of teaching the baboons to match to sample with small numbers of icons to be quite difficult. The

baboons received from two to eight sessions per day with training taking place 5 days per week.

Some training sessions entailed one number of sample items. Other training sessions entailed random mixtures of two different numbers of sample items: the target training number (N) plus the training number that had been used in the prior step (either $N + 1$ or $N + 2$). We progressively reduced the number of sample items from session to session, as the accuracy of the baboons' discrimination behavior permitted. These reductions occasionally required us to revert to greater numbers of sample items because of a deterioration in the baboons' discrimination behavior when the number of sample items was decreased—especially below eight items. During training, the choice stimuli always comprised 16 icons, regardless of the number of icons in the sample stimulus.

The vicissitudes of discrimination training preclude a complete description of the precise sequence of training for each of the 2 baboons. However, Table 1 shows the general order of training sessions for both of the subjects; the entries in this table cover a total of 245 sessions for B03, and 132 sessions for B08. Additional training yielded no sign of further improvements in the baboons' discrimination accuracy.

Results and Discussion

Reducing the number of icons in the sample stimulus arrays had a small effect on relational matching-to-sample accuracy down to 10 icons. However, with further reductions in the number of icons in the sample stimulus arrays, it often took the baboons longer to simultaneously attain a discrimination criterion of at least 79% correct (19/24 trials, $z = 2.25$, $p < .05$; 38/48 trials, $z = 5.51$, $p < .001$) on both same trials and different trials in a single session.

The total numbers of sessions that were necessary for each of the baboons to attain this dual discrimination criterion for the various types of training sessions are shown in Table 1, along with the overall percentage of correct choice responses across both

Table 1
Percentage of Correct Choices to N -Icon Samples on Criterion Sessions for Both Subjects

Icons (N)	Type of session ^a	Baboon 03		Baboon 08	
		Sessions	% correct	Sessions	% correct
14	14/16	1	93.75	1	79.17
12	12/14	4	89.58	2	83.33
10	10/12	1	85.42	3	85.42
8	8/10	4	81.25	16	83.33
7	7/8	5	85.42	5	83.33
6	6/7	2	91.67	2	91.67
	6	—	—	12	82.29
5	5/6	17	85.42	2	83.33
	5	1	85.42	—	—
4	4/5	—	—	1	79.17
	4	39	81.25	8	79.17
3	3/4	13	85.42	2	77.08
	3	11	81.25		
2	2/3	—	—		
	2	14	71.88		

Note. This table presents data for the number of sessions to at least 79% correct on both same trials and different trials with different numbers (N) of icons in the sample displays. Dashes indicate that the dual 79% criterion was not attained to N -icon samples in that training step. Italicized scores indicate that, in the final step of training, a more relaxed dual criterion of at least 70% correct on both same and different trials was attained to N -icon samples.

^a Some training sessions involved samples of N icons only, whereas other training sessions involved samples of N mixed with $N + 1$ or N mixed with $N + 2$ icons.

same trials and different trials in these criterion sessions. B03 met the 79% discrimination criterion down to three-icon sample trials, and B04 met the 79% discrimination criterion down to four-icon sample trials. With a slightly more relaxed dual discrimination criterion of 70% correct on both same and different trials, B03 reliably discriminated two-icon sample trials, and B08 reliably discriminated three-icon sample trials; the overall levels of choice accuracy in these sessions were 71.88% correct ($z = 4.18$, $p < .001$) for B03 (two sample icons), and 77.08% correct ($z = 3.61$, $p < .001$) for B08 (three sample icons).

Clearly, the baboons were capable of reliably discriminating same samples from different samples that contained as few as 2 (B03) or 3 (B08) sample icons. The stepwise training that was required for them to do so is noteworthy and is decidedly different from the behavior of our human participants in Experiment 2, who quite effectively and spontaneously transferred discriminative responding to all of the various sample display numbers after 16-icon training.

Experiment 5

In the final experiment in this study, we explored the baboons' discrimination of same from different sample arrays that contained fewer than 16 items when each of the choice arrays contained the identical number of icons as the sample array. Although the behavior of our human participants was not adversely affected by reducing the number of sample icons, perhaps part of the problem that our baboons had in matching the relation between relations with fewer than 16 sample icons in Experiments 2 and 4 was because the choice arrays in those experiments did not depict the identical number of icons as the sample array. In the extreme case, the baboon was shown a 2-icon sample array and was forced to choose one of two 16-icon choice arrays; such a gross numerical disparity might be a source of the lower level of accuracy that was seen with sample arrays containing fewer than 16 icons.

Method

Subjects

The same 2 baboons served here as had served in Experiments 1, 2, 3, and 4.

Apparatus and Stimuli

The apparatus was the same as in Experiments 1, 2, 3, and 4. Only icons from Stimulus Set 1 and Stimulus Set 2 were used to create the sample and choice stimuli.

Training Procedure

Training in Experiment 5 immediately followed Experiment 4 and consisted of fifty-eight 80-trial sessions, each containing 40 trials in which the sample array involved the "same" relation plus 40 trials in which the sample array involved the "different" relation. Correction trials were again included. The baboons received from two to four sessions per day with training taking place 5 days per week.

Sample and choice displays always involved the identical number of icons: 2, 4, 8, 12, or 16; each value presented a total of 16 times in a session (8 times involving same sample trials and 8 times involving different sample trials). Differential food reinforcement was in effect on all of the

training trials regardless of the number of sample (and choice) icons that were shown.

Results and Discussion

Very little change was seen in the baboons' relational matching-to-sample performance throughout the 58 sessions of training, so we collapsed performance across those sessions for the purpose of analysis. As in Experiment 2, relational matching-to-sample accuracy was a direct and reliable function of the number of icons in the sample (and the choice) displays. Averaged across all 58 sessions, the overall mean accuracy of B03 on 2-, 4-, 8-, 12-, and 16-icon trials was 46%, 54%, 70%, 85%, and 88% correct, respectively; the overall mean accuracy of B08 on 2-, 4-, 8-, 12-, and 16-icon trials was 48%, 54%, 70%, 80%, and 86% correct, respectively. For each baboon, overall accuracy was reliably above chance on the 4-, 8-, 12-, and 16-icon trials (for B03, $z_s = 2.36$, 12.21, 21.21, and 23.37, respectively; for B08, $z_s = 2.49$, 12.08, 18.32, and 21.86, respectively; $p_s < .01$).

The effect of icon number was confirmed by individual subject full factorial ANOVAs of accuracy as a function of array type (same vs. different) and number (2, 4, 8, 12, or 16). The analysis of B03's responding revealed only a significant main effect of number, $F(4, 4630) = 168.29$, $p < .0001$, with no other significant effects ($p_s > .10$). The analysis of B08's responding revealed a significant main effect of number, $F(4, 4630) = 125.75$, $p < .0001$, no significant main effect of type ($F < 1$, $p > .25$), but a significant Number $\times$ Type interaction, $F(4, 4630) = 3.55$, $p < .01$. For B08, the deleterious effect of reducing the number of icons in the sample and choice arrays was slightly larger for the different arrays (falling from 88% for 16-icon arrays to 45% for 2-icon arrays) than it was for the same arrays (falling from 84% for 16-icon arrays to 51% for 2-icon arrays).

Unlike Experiment 2, relational matching-to-sample accuracy was generally similar on same and different trials, reliably differing as a function of the number of sample (and choice) icons only in the relatively weak interaction observed in B08. It is likely that the extensive differential reinforcement training that the baboons had received in Experiment 4 with samples containing fewer than 16 icons was responsible for the similar performance that was shown on same sample trials and on different sample trials.

These results thus show that the number of icons in the sample array probably exerts its effect on relational matching-to-sample performance quite independently of the number of icons in the choice arrays. Furthermore, with extensive discrimination training, baboons can show reliable relational matching performance with four sample (and choice) icons, but not with two sample (and choice) icons. Nevertheless, a larger number of sample icons dramatically benefits the baboons' discriminative performance, whether the numbers of sample and choice icons are identical to one another.

General Discussion

There is no evidence that monkeys can perceive, let alone judge, relations-between-relations. This analogical conceptual capacity is found only in chimpanzees and humans. (Thompson & Oden, 2000, p. 363)

The results of the present series of experiments suggest that we reconsider Thompson and Oden's (2000) recent appraisal of the

species generality of abstract relational conceptualization. Perhaps baboons too can judge the relation between relations. However, might their successful relational matching behavior have been based solely on a perceptual attribute of the displays?

Perceptual Versus Relational Matching

Our matching-to-sample task appears to require the discrimination of abstract same–different relations; there was no perceptual identity between the sample stimuli and the choice stimuli, only relational identity. As did humans who were tested in a highly similar experimental context, baboons successfully solved this relational matching-to-sample task.

A critical question is the extent to which successful matching performance in our task reflects the baboons' ability to match abstract same–different relations rather than some other attribute(s) of the displays (see Blough, in press, for a discussion of the effects of various perceptual attributes on perceived similarity). For example, inspection of Figure 1 suggests that the baboons might have solved the matching task by attending to the spatial regularity of the displays rather than to the depicted same–different relations; the different arrays were perceptually irregular, whereas the same arrays were perceptually regular. Could the baboons have used this "perceptual" cue as the basis for their choice behavior?

A natural test of the spatial regularity hypothesis is to disrupt the regularities, for example, by irregularly spacing the items. Although this manipulation was not used in the present relational version of the same–different task, it was used in an earlier study in which these same baboons were trained to make one response on presentation of a single same display and to make a second response on presentation of a single different display (Wasserman et al., 2001). In Experiment 4 of that study, the 6 baboons that were trained on regularly arranged same displays and different displays were tested on spatially disarranged versions of those displays; the baboons' already high discrimination accuracy on the regular trained displays ($M = 93\%$ correct) was not at all decremented by the introduction of considerable spatial irregularity in the testing displays ($M = 94\%$ correct). Furthermore, in a similar study involving pigeons, disruptions in the perceptual regularity of same displays and different displays also had absolutely no adverse effect on the pigeons' discrimination accuracy (Young & Wasserman, in press).

Additional evidence contrary to a purely perceptual account of our data is found in Experiment 4 of the present study; there, the baboons achieved reliably above-chance relational matching performance when the sample contained only 2 (for B03) or 3 icons (for B08) and the choice stimuli contained 16 icons. Large perceptual differences between the sample array and the choice arrays due to differences in the number of icons greatly reduce the likelihood that the baboons were responding only on the basis of perceptual similarity.

In sum, although perceptual similarity may in some way affect relational matching-to-sample performance, the current experiments suggest that the baboons had also processed the abstract same–different relations illustrated by the sample and choice arrays (for more on the distinction between perceptual and conceptual processes, see Goldstone & Barsalou, 1998).

Still unknown is the role that prior experience might have played in the acquisition of relational matching-to-sample behavior. Our baboons had earlier been trained on a variety of matching-to-sample problems in which physical identity was the basis for correct discriminative responding; more recently still, the baboons had also been taught to make one of two different responses in order to report the presence of same versus different stimulus displays (Wasserman et al., 2001). We suspect that the earlier mastery of both of these kinds of problems helped the baboons learn the present relational matching-to-sample task, but we cannot be sure without conducting additional research with baboons lacking both of these prior experiences.

We are certain that our baboons did not spontaneously exhibit relational matching-to-sample behavior after mastering the other two kinds of problems; in fact, it took the baboons thousands of trials to learn the relation between relations that was required by the present matching-to-sample problem (cf. Thompson et al., 1997). Although discriminating the relation between relations may not be an intellectual forte of baboons, it is nevertheless within their ken.

Baboons Versus Humans

Is relational matching-to-sample similarly performed by humans and baboons? Other data that we collected in these experiments help us answer this question. However, the answer is not a simple yes–no matter.

Baboons and humans both learned the relational matching-to-sample task, in which physical identity between the sample and testing stimuli was eliminated as a controlling stimulus. However, humans learned far faster than baboons, and their final level of discriminative performance was much higher.

Baboons and humans both transferred the relational matching-to-sample discrimination to novel sample stimuli, thus testifying to the abstractness of the behavior in question. However, only the baboons showed a generalization decrement to the novel testing samples; the humans responded with equivalent accuracy to the training and testing samples. It is not known, however, whether the baboons' differential transfer to novel sample stimuli might have been produced by the extensive overtraining that they received on the relational matching-to-sample task (approximately 5,000 and 7,000 trials for B03 and B08, respectively); in contrast, the humans needed fewer than 100 trials to reach discrimination learning criterion, which may have produced less attention to and memory of the specific items that constituted the sample displays.

Baboons and humans both showed a worsening of relational matching-to-sample performance when the number of sample icons was reduced from 16 to 12 to 8 to 4 to 2. However, only the baboons exhibited a strong tendency to report "same" to different sample displays that contained 2 or 4 icons; the humans' discrimination was much more mildly and symmetrically affected by reductions in the number of sample icons.

Baboons and humans both were strongly controlled by the entropy of sample displays that contained various mixtures of same icons and different icons. Conversely, humans more abruptly and completely reported "different" to mixture sample arrays with an entropy of 1 or more.

This evidence suggests two main conclusions. First, baboons and humans both discriminate the entropy in the sample and choice

stimuli. Second, baboons and humans both use the computed entropy of the sample stimuli to guide their choice behavior. The prime difference between baboons and humans appears to be where they set the entropy cutoff between "same" and "different" sample stimulus arrays. In general, humans set that cutoff near 0, whereas baboons set that cutoff near the midpoint of the range of entropy scores, here approximately 2.

Using the sample entropy scores from Experiment 3 and calculating the sample entropy scores from the different sample displays in Experiment 2 (entropy scores for 2 different items = 1.0, 4 different items = 2.0, 8 different items = 3.0, 12 different items = 3.6, and 16 different items = 4.0; entropy scores from the same sample displays were always 0.0), we looked at that sample entropy value in each experiment in which the participants went from first reporting "same" to equivalently reporting "same" and "different" to sample stimuli with entropies from 0.0 to 4.0. Those points were 0.7 for H01 and 1.0 for H02 in both experiments; those points were 2.5 and 2.0 for B03, and 3.0 and 1.3 for B08 in Experiments 2 and 3, respectively.

Just why baboons and humans should adopt such disparate entropy cutoff points is not yet known. One interesting possibility is that extensive extra-experimental language experience leads humans to reserve the verbal response "same" for a much narrower class of stimuli than the verbal response "different" (for more on the perceptual basis of same and different stimulus classes, see Cook & Wixted, 1997). Thus, the word *same* denotes a much more restrictive stimulus set than the word *different*. Because the humans, but not the baboons, could use these verbal responses as mediational devices in the present experimental procedures, the humans' response profiles could differ markedly from those of the baboons.

This notion might be tested in one of two general ways. First, we could try to reduce the involvement of humans' verbal behavior in our experimental procedures. We could test preverbal children. Alternatively, we could overload adults' verbal processing capabilities by requiring them to perform a demanding verbal task while they are concurrently engaged in relational matching-to-sample behavior. In either case, humans might now respond more like baboons.

Second, we could try to simulate the human's categorical discrimination behavior—albeit without any verbal accompaniment—in baboons. This approach would reserve one response for only same displays and a second response for displays containing one or more different items. After such training, baboons might now respond more like humans.

These and other researchable ideas suggest that much more can and should be learned about relational matching-to-sample in nonhuman animals before further debating the phylogenetic limits of this cognitive skill. To be sure, the stakes are high. What is at issue is no arcane point, but the very distinction between the minds of human beings and nonhuman animals. So, to quote French (1995), the ability to perceive sameness in all of its subtlety "sets human cognition apart from any other on our planet" (Introduction, p. xvi).

Conceptual Relational Matching and Analogy

Thompson and Oden (2000) have proposed that the conceptual matching-to-sample task is the essence of analogy, in which all of

the logical arguments are visually provided for the research participant. If this is the case, then the evolutionary origins of analogical thinking truly require comparative study. It is highly unlikely that analogical thinking could have emerged full blown without any phylogenetic antecedents.

According to many theorists (Gentner & Markman, 1997; Holyoak & Thagard, 1997; Sternberg, 1977), the very essence of human reasoning and intelligence—both basic and advanced—is grounded on analogical thinking.

[Analogy] is central in the study of learning and discovery. Analogies permit transfer across different concepts, situations, or domains and are used to explain new topics. Once learned, they can serve as mental models for understanding a new domain. . . . Analogies are often used in problem solving and inductive reasoning because they can capture significant parallels across different situations. Beyond these mundane uses, analogy is a key mechanism in creativity and scientific discovery. (Gentner, 1999, p. 17)

Analogical thinking and its possible precursors (like same-different discrimination learning) may very well be found in non-human animals—if only we assiduously look for them. The search may be long and hard, as Darwin (1871/1920) appreciated well over a century ago when he contemplated the formidable task of determining the limits of animal mentation. However, the developmental study of analogical thinking is no less daunting, and, the toils of that experimental undertaking are beginning to bear fruit (e.g., Gentner & Markman, 1997; Goswami, 1989; Halford, 1992; Tyrrell, Stauffer, & Snowman, 1991). The study of relational matching-to-sample may prove to be an important step toward the identification of the evolutionary and ontogenetic origins of analogical thinking.

References

- Blough, D. S. (in press). The perception of similarity. In R. G. Cook (Ed.), *Avian visual cognition*. Retrieved from <http://www.pigeon.psy.tufts.edu/avc>
- Cook, R. G., & Wixted, J. T. (1997). Same-different texture discrimination in pigeons: Testing competing models of discrimination and stimulus integration. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 401–416.
- Darwin, C. (1920). *The descent of man, and selection in relation to sex* (2nd ed.). New York: Appleton. (Original work published 1871)
- Delius, J. D. (1994). Comparative cognition of identity. In P. Bertelson, P. Eelen, & G. d'Ydewalle (Eds.), *International perspectives on psychological science* (Vol. 1, pp. 25–40). Hillsdale, NJ: Erlbaum.
- Fagot, J., & Deruelle, C. (1997). Processing of global and local visual information and hemispheric specialization in humans (*Homo sapiens*) and baboons (*Papio papio*). *Journal of Experimental Psychology: Human Perception and Performance*, 23, 429–442.
- Fagot, J., Kruschke, J. K., Dépy, D., & Vauclair, J. (1998). Associative learning in humans (*Homo sapiens*) and baboons (*Papio papio*): Species differences in learned attention to features. *Animal Cognition*, 1, 123–133.
- French, R. M. (1995). *The subtlety of sameness*. Cambridge, MA: MIT Press.
- Gentner, D. (1999). Analogy. In R. A. Wilson & F. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 17–20). Cambridge, MA: MIT Press.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45–56.

- Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65, 231–262.
- Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development*, 4, 251–268.
- Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development*, 35, 193–217.
- Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *American Psychologist*, 52, 35–44.
- Premack, D. (1983). The codes of man and beast. *The Behavioral and Brain Sciences*, 6, 125–137.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning*. Hillsdale, NJ: Erlbaum.
- Thompson, R. K. R. (1995). Natural and relational concepts in animals. In H. L. Roitblat & J.-A. Meyer (Eds.), *Comparative approaches to cognitive science* (pp. 175–224). Cambridge, MA: MIT Press.
- Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, 24, 363–396.
- Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naïve chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 31–43.
- Tyrrell, D. J., Stauffer, L. B., & Snowman, L. G. (1991). Perception of abstract identity/difference relationships by infants. *Infant Behavior & Development*, 14, 125–129.
- Wasserman, E. A. (1993). Comparative cognition: Beginning the second century of the study of animal intelligence. *Psychological Bulletin*, 113, 211–228.
- Wasserman, E. A. (1997). The science of animal cognition: Past, present, and future. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 123–135.
- Wasserman, E. A., Fagot, J., & Young, M. E. (2001). Same–different conceptualization by baboons (*Papio papio*): The role of entropy. *Journal of Comparative Psychology*, 115, 42–52.
- Wasserman, E. A., Hugart, J. A., & Kirkpatrick-Steger, K. (1995). Pigeons show same–different conceptualization after training with complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 21, 248–252.
- Wasserman, E. A., Young, M. E., & Nolan, B. C. (2000). Display variability and spatial organization as contributors to the pigeon's discrimination of complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 26, 133–143.
- Wilde, J., Vauclair, J., & Fagot, J. (1994). Eye movements in baboons performing a matching-to-sample task presented in a divided-field format. *Behavioural Brain Research*, 63, 61–70.
- Young, M. E., & Wasserman, E. A. (1997). Entropy detection by pigeons: Response to mixed visual displays after same–different discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 157–170.
- Young, M. E., & Wasserman, E. A. (2001). Entropy and variability discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 278–293.
- Young, M. E., & Wasserman, E. A. (in press). Evidence for a conceptual account of same–different discrimination learning in the pigeon. *Psychonomic Bulletin and Review*.
- Young, M. E., Wasserman, E. A., & Dalrymple, R. M. (1997). Memory-based same–different conceptualization by pigeons. *Psychonomic Bulletin and Review*, 4, 552–558.
- Young, M. E., Wasserman, E. A., & Garner, K. L. (1997). Effects of number of items on the pigeon's discrimination of same from different visual displays. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 491–501.

Received August 11, 2000

Revision received April 3, 2001

Accepted April 3, 2001 ■

Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at journals@apa.org and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands. (Your donation is of course tax deductible.)