Memory-based *same–different* conceptualization by pigeons

MICHAEL E. YOUNG, EDWARD A. WASSERMAN, and RHONDA M. DALRYMPLE
University of Iowa, Iowa City, Iowa

We first trained pigeons to peck one button (*same*) after the successive presentation of 16 identical pictures and to peck a second button (*different*) after the successive presentation of 16 non-identical pictures. Later, we tested the birds with other lists of *same* and *different* items composed of completely novel pictures. Accuracy to the testing lists reliably exceeded chance levels, thus demonstrating *same–different* conceptualization by pigeons under conditions that, for the first time, (1) eliminated purely perceptual mechanisms of discrimination learning and transfer and (2) required memory-based processing of the experimental stimuli.

The search for abstract conceptualization in nonhuman animals has been a protracted and arduous one that has spanned a full century of research in animal intelligence (Wasserman, 1993). Different methods of investigation and species of animals have been used in this search, with decidedly mixed results. With only a very few combinations of tasks and species has any positive evidence of abstract conceptual behavior been obtained; most other combinations of tasks and species have yielded negative evidence (for reviews see Delius, 1994; Herrnstein, 1990; Thompson, 1995).

There is now growing awareness that most laboratory investigations of conceptual behavior in nonhuman animals have probably underestimated the subjects' cognitive capacities (see Wright, 1992, and Zentall, 1993, for more on this issue). One problem has been the use of too few experimental tasks (e.g., matching to sample); another has been the use of too few and too simple stimuli. Any of these limiting factors might lead researchers to arrive at an inappropriately pessimistic view of animal cognition.

Wasserman, Hugart, and Kirkpatrick-Steger (1995) recently reported an experiment that used a still-uncommon behavioral task (*same–different* discrimination training) and a large number of highly complex visual items (Macintosh computer icons) in the hope of obtaining more encouraging data on the conceptual abilities of a familiar laboratory animal—the pigeon. Their results provided supportive evidence that the pigeon can learn a general *same–different* concept that transcends the particular stimuli with which it was originally trained. Specifically, pigeons were first taught to peck one button (*same*) when

they viewed an array of computer icons that comprised 16 copies of the same icon and to peck a second button (*different*) when they viewed an array that comprised 16 distinctly different icons. These *same* and *different* training displays were all created from one set of computer icons. The birds were later tested with novel *same* and novel *different* displays that were created from a second set of computer icons. Accuracy to the training stimuli averaged 83% correct, and accuracy to the testing stimuli averaged 71% correct, in each case reliably exceeding the chance score of 50%.

That experiment used arrays of *simultaneously* visible computer icons that occupied a sizable area of the computer monitor. Therefore, it is possible that the pigeons' discrimination behavior was not truly *conceptual*, but was the result of low-level visual processing of the *same* and *different* arrays (i.e., texture discrimination). Although the details of these data argued against this purely *perceptual* interpretation (see Cook, Cavoto, & Cavoto, 1995, for similar evidence and argument), we decided that this interpretive possibility might be explicitly tested by using lists of *successively* presented *same* and *different* icons. Only one item at a time would be visible, thereby eliminating texture discrimination as the basis for the pigeon's discrimination learning and generalization performance. Indeed, the successive presentation of individual icons would mean that any discrimination of *same* from *different* lists had to be based on the bird's remembering at least some of the 16 icons that had been presented prior to the choice test. Successful discrimination learning and transfer with lists of successively presented visual stimuli would then not only eliminate a highly plausible interpretation of our earlier results, but would also for the first time establish memory-based conceptualization in nonhuman animals.

METHOD

Subjects

The subjects were 4 feral pigeons kept at 85% of their free-feeding weights; the birds had served in unrelated studies involving different vi-

This research was supported by National Institute of Mental Health Grant MH 51562. Additional support was provided by the U.S. Department of Education Fund for the Improvement of Postsecondary Education and the Graduate College of the University of Iowa. Correspondence should be addressed to E. A. Wasserman, Department of Psychology, University of Iowa, Iowa City, IA 52242-1407 (e-mail: ed-wasserman@uiowa.edu).

Figure 1. The 48 Macintosh icons that constituted Stimulus Sets A, B, and C.

sual stimuli in the same boxes used in the present experiment. We pre-trained the birds to peck anywhere on the central and peripheral areas of the touch screen (see Apparatus section below) on a fixed ratio (FR) 30 schedule; later, we taught them to peck directly at a single icon on an FR 3 schedule when the icon appeared in any of 25 randomly chosen locations in the central screen area. For this purpose, three icons of the

same general type as those used in the experiment were randomly chosen, but they were never shown thereafter.

Apparatus

The experiment used the same four chambers and Macintosh computers that were detailed by Wasserman et al. (1995). One wall of each chamber contained a large opening with a frame attached to the outside. The frame held a clear touch screen. An aluminum panel was placed in front of the touch screen to allow the pigeons access to circumscribed portions of a video monitor located behind the touch screen. There were five openings or buttons in the panel: a 7×7 cm square central "display" area in which the icon arrays appeared and four round "report" areas (1.9-cm diameter) located 2.3 cm from each of the four corners of the central opening. Only the lower two report areas were used; the lower left area could be lighted green and the lower right area could be lighted red. A food cup was centered on the rear wall level with the floor. A pellet dispenser delivered 45-mg food pellets through a vinyl tube into the cup. A houselight, mounted on the rear wall of the chamber, provided illumination during sessions.

Visual Stimuli and Experimental Design

Forty-eight black-and-white Macintosh icons were chosen as the total item pool; these icons were sorted into three sets of 16 icons each (Sets A, B, and C; Figure 1), and from each of these sets, 16-icon lists were constructed. Sets A and B had been used in Wasserman et al. (1995); Set C was chosen from the same collection from which Sets A and B had been selected. Special care was taken to ensure that the icons in each set were readily distinguishable from one another and that no set contained a disproportionate number of icons of a specific size, shape, or brightness. For a *same* list, a single icon from the appropriate set was randomly chosen and was used to make up a list of 16 identical icons; for a *different* list, all 16 of the icons in a set were used with no repetitions. The 16 *same* or 16 *different* icons were randomly distributed to 25 possible locations in the display area on the computer monitor so that the pigeon had to scan the entire display area and aim its pecks directly at the individually presented icons in a list. Therefore, on each trial, a random 16 of the 25 locations involved list icons; the other 9 were unused (see Figures 2A–2B for representative displays of *same* and *different* lists). During the first phase of the experiment, we trained 2 birds (26r and 29r) with *same* and *different* lists of icons from Set A, and we tested those birds with *same* and *different* lists of icons from Set B; we trained the other 2 birds (18b and 48r) with *same* and *different* lists of icons from Set B, and we tested those birds with *same* and *different* lists of icons from Set A. During the second phase of the experiment, we trained all 4 birds with *same* and *different* lists of icons from both Sets A and B, and we tested the birds with *same* and *different* lists of icons from Set C.

Procedure

Training 1. Training sessions consisted of 160 daily trials: 20 randomized blocks of eight trials each, four with *same* lists and four with *different* lists. For Birds 18b and 26r, pecks to the green button on *same* trials and to the red button on *different* trials were correct and were reinforced with one or two pellets of food (depending on each bird's 85% free-feeding weight); pecks to the red button on *same* trials and to the green button on *different* trials were incorrect and were punished by repetition of the trial until the correct response was made. For Birds 29r and 48r, the report button assignments were reversed.

Each trial began with illumination of the display area with a white field containing a black cross in the center. One peck anywhere in the display area turned on a single icon as a black-on-white picture, with the white background otherwise occupying the entire display area; that icon could randomly appear in any of the 25 possible locations in an invisible 5×5 grid. A single peck to the 1.6-cm-wide $\times$ 1.4-cm-high screen area containing the icon turned off that item and turned on the next item in the list in 1 of the 24 unused grid locations, randomly chosen. This process of selecting icon locations at random and without replacement and moving the response button to the corresponding location continued until the 16th icon was pecked; by that time, either 16 identical icons would have been seen on *same* trials (each of the 16 icons in a training set having an equal likelihood of being presented on any trial) or 16 nonidentical icons would have been seen on *different*

Figure 2A. Representative list of 16-icon displays shown on *same* trials.

ent trials (the serial position of each of the 16 icons in a training set having been randomly chosen). With the 16th peck, the last list icon reversed to a white-on-black picture (to signal the availability of the report buttons), and the lower two report areas were lighted green (left) and red (right). A correct choice response blackened the display area and the report areas and delivered food pellet reinforcement; an incorrect choice response blackened the display area and the report areas, turned off the houselight for 5 or 6 sec, and began a series of one or more correction trials that were not scored in data analyses. Inter-trial intervals averaged 8 sec (range, 6–10 sec). When a pigeon met or exceeded 80% correct responses for 5 consecutive days with a minimum of 30 days of training, it moved on to testing.

Testing 1. Eight testing sessions were given. Testing sessions began with 32 warm-up trials involving eight *same* training lists and eight *different* training lists shown in each of two blocks of 16 trials. Thereafter, eight randomized blocks of 18 trials were given consisting of eight *same* training lists, eight *different* training lists, one *same* testing list, and one *different* testing list. Thus, the last 144 trials of the testing sessions consisted of 128 training trials with lists of familiar icons and 16 testing trials with lists of novel icons.¹ Differential food reinforcement for left and right report responses was given on all training trials. During the first 4 days of testing, nondifferential food reinforcement for left and right report responses was given on testing trials; food was delivered regardless of the pigeon's choice responses. In order to de-

Figure 2B. Representative list of 16-icon displays shown on *different* trials.

terminate the generality of the results of nondifferential reinforcement testing, no reinforcement was given for left or right report responses on testing trials during the last 4 days of testing. Therefore, the designations “correct” and “incorrect” for choice responses on testing trials were for scoring purposes only. No correction trials were necessary on testing trials. Testing sessions were separated by at least 1 day of retraining, as in Training 1. If discriminative performance on a re-training day dropped below the 80% criterion, the pigeon was returned to retraining on the next day. Once performance met or exceeded the 80% criterion, testing resumed.

Training 2. These sessions were just like those in Training 1, except that training lists were composed of items from both Sets A and B. Daily sessions involved 160 trials: 10 randomized blocks of 16 trials each,

4 with *same* lists and 4 with *different* lists from Set A, and 4 with *same* lists and 4 with *different* lists from Set B. Training 2 continued until a pigeon met or exceeded 80% correct responses on each of the Set A lists and the Set B lists for 5 consecutive days.

Testing 2. The eight testing sessions in Testing 2 were similar to those in Testing 1. Testing sessions began with 32 warm-up trials involving four *same* training lists and four *different* training lists from Set A, and four *same* training lists and four *different* training lists from Set B, shown in each of two blocks of 16 trials. Thereafter, eight randomized blocks of 18 trials were given, consisting of four *same* training lists from Set A, four *different* training lists from Set A, four *same* training lists from Set B, four *different* training lists from Set B, one *same* testing list from Set C, and one *different* testing list from Set C.

Thus, the last 144 trials of the testing sessions consisted of 128 training trials with lists of familiar icons from Sets A and B and 16 testing trials with lists of novel icons from Set C. Testing sessions were separated by at least 1 day of retraining, as in Training 2. If discriminative performance on a retraining day dropped below the 80% criterion for either Set A lists or Set B lists, the pigeon was returned to retraining on the next day. Once performance met or exceeded the 80% criterion on both Set A lists and Set B lists, testing resumed.

RESULTS

Training 1

Mean discriminative performance began near 50% correct and first exceeded the 80% level on Day 27 of Training 1. Analysis of variance revealed that the rise in performance was reliable [$F(1,3) = 316.90, p < .01$]. Overall accuracy on *same* and *different* trials did not differ [$F(1,3) < 1$], nor did this factor interact with days of training [$F(1,3) < 1$].² Birds 18b, 26r, 29r, and 48r took 41, 30, 31, and 36 days, respectively, to reach the 5-day 80% criterion on their training lists. Over those last 5 days of training, the mean percentages of correct choice responses for Birds 18b, 26r, 29r, and 48r, respectively, were 89%, 85%, 83%, and 83%.

Testing 1

During the eight testing sessions, the discrimination of *same* from *different* lists of familiar training icons remained high, averaging 90% during the four nondifferential reinforcement tests and 93% during the four extinction tests. The discrimination of *same* from *different* lists of novel testing icons was lower, averaging 57% during the nondifferential reinforcement tests and 58% during the extinction tests. Across both types of testing sessions, lists of familiar training icons were better discriminated than were lists of novel testing icons [$F(1,3) = 42.5, p < .01$]. Still, during both nondifferential reinforcement tests and extinction tests, discrimination performance for lists of familiar training icons and for lists of novel testing icons each differed reliably from the chance score of 50% (binomial $ps < .05$). Lists of novel *same* icons and lists of novel *different* icons each differed reliably from the chance score of 50%; although performance to novel *same* lists was reliably above chance (81%), performance to novel *different* lists was reliably below chance (34%). As to the individual pigeons, binomial tests indicated that all 4 birds showed reliable discrimination of lists of familiar training icons during both the nondifferential reinforcement tests and the extinction tests; 2 birds (18b and 29r) did so for lists of novel testing icons during the extinction tests, and 1 bird (29r) did so for lists of novel testing icons during the reinforcement tests.

Overall mean discrimination scores to lists of familiar training icons and to lists of novel testing icons are shown in the top of Figure 3, along with the corresponding binomial statistics.

Training 2

Birds 18b, 26r, 29r, and 48r took 6, 6, 13, and 20 days of Training 2, respectively, to reach the joint 5-day 80%

criterion on both their first and second training set lists. Over those last 5 days of training, the mean percentages of correct choice responses to first training set lists for Birds 18b, 26r, 29r, and 48r, respectively, were 98%, 95%, 93%, and 87%; the mean percentages of correct choice responses to second training set lists for Birds 18b, 26r, 29r, and 48r, respectively, were 85%, 91%, 90%, and 83%.

Testing 2

During the eight testing sessions, the discrimination of *same* from *different* lists of the first set of training icons remained high, averaging 97% during the four nondifferential reinforcement tests and 94% during the four extinction tests; the discrimination of *same* from *different* lists of the second set of training icons also remained high, averaging 91% during the four nondifferential reinforcement tests and 95% during the four extinction tests. The discrimination of *same* from *different* lists of novel icons was lower, averaging 72% during the nondifferential reinforcement tests and 74% during the extinction tests. Across both types of testing sessions, lists of the first and second sets of training icons were better discriminated than were lists of novel testing icons [$F(2,6) = 17.8, p < .01$]. Nevertheless, during both nondifferential reinforcement tests and extinction tests, discrimination performance for lists of the first set of training icons, for lists of the second set of training icons, and for lists of the novel testing icons each differed reliably from the chance score of 50%. Lists of novel *same* icons and lists of novel *different* icons each differed reliably from the chance score of 50%; performance to novel *same* lists was reliably above chance (71%), as was performance to novel *different* lists (75%). As to the individual pigeons, binomial tests indicated that all 4 birds showed reliable discrimination of lists of the first and the second sets of training icons during both the nondifferential reinforcement tests and the extinction tests. All 4 birds showed reliable discrimination of lists of the novel testing icons during the nondifferential reinforcement tests; 3 of the 4 birds showed reliable discrimination of lists of the novel testing icons during the extinction tests (the sole exception was Bird 48r).

Overall mean discrimination scores to lists of first training set icons, to lists of second training set icons, and to lists of novel testing icons are shown in the bottom of Figure 3, along with the corresponding binomial statistics.

DISCUSSION

Our results clearly show that pigeons can generalize the *same* and *different* report responses learned with *same* and *different* lists of familiar icons to other *same* and *different* lists of novel icons. The obtained mean of 73% correct responding to lists of novel testing icons (in Testing 2) after training with lists of *same* and *different* icons from both Sets A and B (in Training 2) was not only well above the chance level of 50% correct, but it also exceeded the obtained mean of 57% correct responding to lists of novel testing icons (in Testing 1) after training with lists of *same* and *different* icons from just Set A or Set B (in Training 1). This finding closely corresponds with other research on natural and abstract conceptualization by both humans and nonhumans showing that larger sets of training stimuli promote

Figure 3. Mean percentage of correct choice responses during the entire 8 days (4 days of nondifferential reinforcement followed by 4 days of experimental extinction) of Testing 1 and Testing 2. Accuracy scores are shown separately for lists of icons in the training and testing sets. In Testing 1, training icons came from either Set A or Set B, and testing icons came from the remaining stimulus set. In Testing 2, training icons came from Set A and Set B, and testing icons came from Set C. Significant binomial *ps* are denoted as follows: **p* < .05; ***p* < .01.

stronger generalization performance (see Wasserman, 1993, for a review).

Just how does the present experiment, which used successive stimulus presentation, compare with and extend Wasserman et al.'s (1995) earlier experiment on *same-different* conceptualization using simultaneous stimulus presentation? Extended training in each experiment yielded average testing performance of 71% correct with simultaneous arrays of icons and 73% correct with successive lists of icons. Clearly, highly similar generalization performance was produced by both methods of multiple stimulus presentation. Earlier testing in each experiment yielded less accurate, but nevertheless statistically significant, generalization performance: 64% correct with simultaneous arrays and 57% correct with successive lists. At these earlier testing times,

each experiment yielded an asymmetry in generalization to novel *same* and *different* displays. With simultaneous stimulus presentation, accuracy on *different* trials (82% correct) exceeded accuracy on *same* trials (46% correct), whereas with successive stimulus presentation, accuracy on *same* trials (81% correct) exceeded accuracy on *different* trials (34% correct). These different patterns of stimulus generalization suggest that different cognitive processes may have contributed to discriminative responding in these two experimental situations. Efforts to elucidate those processes are now under way in our laboratory.

Our past and present research convincingly shows that pigeons are indeed capable of *same-different* conceptualization and that this form of cognition does not require the simultaneous display of multiple visual stimuli. The earlier project (Wasserman et al., 1995), which used

simultaneous displays of 16 *same* or 16 *different* stimuli, yielded strong evidence of discrimination learning and generalization to novel testing displays. Low-level texture discrimination could not completely explain those data (nor those of Cook et al., 1995) because the training displays were better discriminated than were the testing displays; this discrimination differential meant that the birds did see the specific items in a display, not just homogeneous or heterogeneous textures. Critically, low-level texture discrimination was impossible in the present project because only one stimulus at a time was shown to the pigeons.

The present project also contributes an unprecedented finding—namely memory-based conceptualization by a nonhuman animal. Unlike all previously reported results of which we are aware, *same-different* conceptualization in this experiment had to be based on the bird's remembering 1 or more of the 16 icons that had been presented in a list. Precisely how many icons were remembered and the role of serial position in that memory is now being studied in our laboratory. Early evidence on the number of items in a successive list of icons suggests that, as in the case of a simultaneous array of icons (Young, Wasserman, & Garner, in press), discriminative performance improves as the number of displayed items increases. This result testifies to the importance of using a large number of stimuli in the teaching of complex conceptual behavior. Indeed, at no point in any of our studies have successive lists or simultaneous arrays of fewer than eight icons supported reliable discrimination between *same* and *different* trials, a result that may be explained by the complexity or the variability of the displayed visual information (Young & Wasserman, 1997).

In summary, it is becoming increasingly evident that the pigeon can exhibit highly advanced forms of conceptual behavior. That conceptual behavior involves relatively concrete, basic-level categories (Bhatt, Wasserman, Reynolds, & Knauss, 1988) and higher order (Wasserman, DeVolder, & Coppage, 1992) categories as well as more abstract stimulus relations (Cook et al., 1995; Edwards, Jagielo, & Zentall, 1983; Santiago & Wright, 1984; Wasserman et al., 1995; Young & Wasserman, 1997; Young et al., in press), thus argues against Morgan's (1894/1896) early assertion that nonhuman animals are incapable of conceptualization and Premack's (1983) more recent contention that, among nonhuman animals, only language-trained chimpanzees can show *same-different* conceptualization.

REFERENCES

- BHATT, R. S., WASSERMAN, E. A., REYNOLDS, W. F., JR., & KNAUSS, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 219-234.
- COOK, R. G., CAVOTO, K. K., & CAVOTO, B. R. (1995). Same-different texture discrimination and concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 253-260.
- 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.
- EDWARDS, C. A., JAGIELO, J. A., & ZENTALL, T. R. (1983). "Same/different" symbol use by pigeons. *Animal Learning & Behavior*, **11**, 349-355.
- HERRNSTEIN, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition*, **37**, 133-166.
- MORGAN, C. L. (1894/1896). *An introduction to comparative psychology*. London: Walter Scott.
- PREMACK, D. (1983). The codes of beast and man. *Behavioral & Brain Sciences*, **6**, 125-167.
- SANTIAGO, H. C., & WRIGHT, A. A. (1984). Pigeon memory: Same/different concept learning, serial probe recognition acquisition, and probe delay effects on the serial-position function. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 498-512.
- 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.
- WASSERMAN, E. A. (1993). Comparative cognition: Beginning the second century of the study of animal intelligence. *Psychological Bulletin*, **113**, 211-228.
- WASSERMAN, E. A., DEVOLDER, C. L., & COPPAGE, D. J. (1992). Non-similarity-based conceptualization in pigeons via secondary or mediated generalization. *Psychological Science*, **3**, 374-379.
- WASSERMAN, E. A., HUGART, J. A., & KIRKPATRICK-STEGE, K. (1995). Pigeons show same-different conceptualization after training with complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 248-252.
- WRIGHT, A. A. (1992). Testing the cognitive capacities of animals. In I. Gormezano & E. A. Wasserman (Eds.), *Learning and memory: The behavioral and biological substrates* (pp. 45-60). Hillsdale, NJ: Erlbaum.
- 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., & GARNER, K. L. (in press). Effects of number of items on the pigeon's discrimination of same from different visual displays. *Journal of Experimental Psychology: Animal Behavior Processes*.
- ZENTALL, T. R. (1993). Animal cognition: An approach to the study of animal behavior. In T. R. Zentall (Ed.), *Animal cognition: A tribute to Donald A. Riley* (pp. 3-15). Hillsdale, NJ: Erlbaum.

NOTES

1. Novel icons were presented on only 16 of the 176 trials in a testing session (testing sessions always followed at least one retraining session that involved no novel icons). *Different* list trials necessarily involved all 16 of the icons in the testing set; *same* list trials necessarily involved just one icon, which was randomly chosen from the 16 icons in the testing set. Because only eight *same* list testing trials were given in a testing session, it took at least two testing sessions (and usually many more) before each of the 16 icons in the testing set had been shown on *same* list testing trials. The rarity of testing icon presentation, the fact that testing trials involved either nondifferential reinforcement or experimental extinction, and the failure of testing performance to change discernibly during the course of repeated testing sessions all argue persuasively for the sensitivity, reliability, and purity of our testing methods.

2. Additional statistical analyses of responding to lists of training icons during later periods of training and testing failed to disclose any differences in performance between lists of *same* icons and lists of *different* icons. Therefore, no further mention is made of this difference to training icons. Differential responding to lists of novel testing icons on *same* and *different* trials was examined and is separately reported in Testing 1 and Testing 2.

(Manuscript received December 17, 1996;
revision accepted for publication April 7, 1997.)