Abstract concepts—rules that transcend training stimuli—have been argued to be unique to some species. Pigeons, a focus of much concept-learning research, were tested for learning a matching-to-sample abstract concept. Five pigeons were trained with three cartoon stimuli. Pigeons pecked a sample 10 times and then chose which of two simultaneously presented comparison stimuli matched the sample. After acquisition, abstract-concept learning was tested by presenting novel cartoons on 12 out of 96 trials for 4 consecutive sessions. A cycle of doubling the training set followed by retraining and novel-testing was repeated eight times, increasing the set size from 3 to 768 items. Transfer performance improved from chance (i.e., no abstract-concept learning) to a level equivalent to baseline performance (>80%) and was similar to an equivalent function for same/different abstract-concept learning. Analyses assessed the possibility that item-specific choice strategies accounted for acquisition and transfer performance. These analyses converged to rule out item-specific strategies at all but the smallest set-sizes (3–24 items). Ruling out these possibilities adds to the evidence that pigeons learned the relational abstract concept of matching-to-sample.

**Keywords:** matching-to-sample, same/different, abstract-concept learning, generalized matching, pigeon

Kent D. Bodily and Jeffrey S. Katz
Auburn University

Anthony A. Wright
University of Texas Health Science Center at Houston
sample with two stimuli (repeated on each trial) or with 152 stimuli (presented once per session and randomized across sessions). After reaching a training criterion of 75% correct, both groups were tested on novel, transfer stimuli. The two-stimuli group performed at chance on novel-transfer, whereas the 152-stimuli group performed as well on novel-transfer as on training stimuli, meeting the criteria for abstract-concept learning. These results suggested that a large training set (e.g., 152 images) could produce abstract-concept learning in pigeons.

Training set size was further explored using pigeons (Katz & Wright, 2006), capuchin monkeys (Wright et al., 2003) and rhesus monkeys (Katz, Wright, & Bachevalier, 2002) in a two-item same/different (S/D) task; a task somewhat different from the MTS task (for a review see Wright & Katz, 2006). In these S/D tasks, two comparison stimuli (travel slides) were presented, one above the other, and a white box was presented to the right side of the lower stimulus. If the two stimuli were the same, then contacting the bottom stimulus was reinforced. If the two stimuli were different, then contacting the white box was reinforced. Subjects were trained on eight pictures and then, upon reaching a performance criterion, tested on novel pictures. For all species, transfer performance was no different from chance. After testing, the training set-size was doubled in size to 16, 32, 64, and so forth. Novel-stimulus transfer tests were conducted after reacquisition at each set size (except 16). Percent correct on transfer tests increased for all species as the training set size increased until transfer-test performance was equal to baseline performance (i.e., full abstract-concept learning).

The purpose of the present study was to test the effect of systematically increasing the training set size in a MTS task with pigeons. The MTS task is similar to a psychophysical two-alternative forced-choice procedure, in which the correct comparison is presented on every trial, whereas the S/D task is similar to a YES/NO procedure (Katz, Wright, & Bodily, 2007). The MTS and S/D tasks may, therefore, be solved by different strategies (Pepperberg, 1987; Premack, 1978, 1983), making it important to test if set size produces the same effect on novel-transfer in MTS as in S/D.

**Method**

**Subjects**

Five male pigeons (*Columba livia*) served as subjects. Two pigeons, approximately two years old, previously served in MTS research involving three different sets of three training images and novel-transfer tests. At no time was the training set size expanded beyond three. The remaining three pigeons were approximately one year old and had no previous experimental training. All pigeons were maintained at 80% to 85% of their free-feeding body weights throughout the experiment. They were housed individually and had constant access to grit and water in their home cages. The colony room was maintained on a 12-hr light/dark cycle.

**Apparatus**

**Chamber.** The chamber was a wooden box (38-cm wide × 36.5-cm deep × 39.5-cm high). In the back panel, an axial fan, Dayton Electric (Model 4WT40), provided both ventilation and noise. A houselight (lamp #1829) centered in the ceiling provided internal lighting during intertrial intervals (ITI). The food hopper (custom design), operated by a computer-controlled relay interface (Keithley, ERA-01), was centered below the monitor (Eizo FlexScan T566; 17-inch flat screen CRT; 800 × 600 pixel resolution). A thin piece of glass mounted in a 25 × 17.5-cm viewing window separated and protected the monitor from the pigeon’s pecks. An infrared touch screen (Carroll Touch, UniTouch 17”), framed the viewing window and detected pecks at the monitor.

**Stimuli.** Stimuli were color, computer-drawn, cartoon images (2.5-cm high × 3-cm wide at 28 pixels/cm) as used by Wright et al. (1988) and Wright (1997). Variations in cartoon images allowed for the creation of a large number of unique images, ensuring that every test trial consisted of novel images. Stimuli were arranged in the display such that the sample and comparisons formed a triangle (19.05 × 7.78 cm). The sample stimulus appeared centered horizontally at approximately 8-cm above the bottom of the monitor. Comparison stimuli appeared centered at 4-cm above the bottom with the left and right comparisons appearing 9-cm and 25 cm from the left side of the monitor, respectively.

**Experimental control.** Experimental events were controlled and recorded using custom software written in Visual Basic 6.0 on a microcomputer (Dell Dimensions 2100). A video card (ATI Xpert 98) controlled the graphics. A PCI card (Keithley KPCI-P10, Cleveland, OH) controlled the relay interface that operated the hopper, hopper-light, and houselight.

**Procedure**

**Pretraining.** Naive pigeons were trained by hand to eat from the food hopper. Responding was then autoshaped to the first three training stimuli (apple, duck, and grapes). Location (left or right) and stimulus presentation were counterbalanced within a session. Stimuli appeared on separate trials and randomly occurred 28 times in an 84-trial session. A stimulus was presented for 10 seconds, after which the food hopper was raised for 5 seconds. Pecking the stimulus extinguished it and immediately raised the food-hopper. A 50-s ITI separated trials. Once pigeons began consistently pecking the stimuli, a response dependent (FR1) procedure was implemented. This procedure was identical to the autoshaping procedure with the exception of response dependent food access (stimuli were presented until pecked) and a 15-s ITI. The response dependent procedure was used until rapid and consistent responding was maintained throughout a session. The two experienced pigeons started directly in training.

**Training.** Daily sessions consisted of 96 trials. Pigeons were trained to match-to-sample using three stimuli (apple, duck, and grapes). In two-choice MTS, a set size of three stimuli produces 12 unique stimulus combinations, or configurations. Each configuration was shown eight times in each daily session. A trial began with the presentation of the sample stimulus (e.g., duck). A response to the sample resulted in the immediate appearance of the two comparison images (e.g., apple and duck). The first response to the matching comparison image raised the food hopper, allowing access to mixed grain for 2 to 5 seconds, varying to maintain individual pigeon’s weight. The sample and foil (incorrect comparison) images were removed immediately after the matching response, whereas the correct comparison remained for 8 seconds. After an incorrect response, the foil was removed and the sample
and correct comparison remained for 4 seconds (responses during this time had no programmed consequence and were not analyzed). A darkened, 8 seconds timeout followed the removal of the remaining stimuli. A correction procedure (CP), in which incorrect trials were repeated until a correct response occurred, was employed beginning with the second session and continued until ≥85% correct was achieved on two consecutive sessions. Only the choice on the first presentation of trials was included in the analyses. A 15-s ITI followed every trial. The houselight was lit only during the ITI. The response requirement to the sample (FR) was systematically increased to 10 responses over seven sessions. Transfer testing began after subjects achieved ≥85% correct without the CP for one session.

Transfer testing. Twelve transfer trials were quasi-randomly inserted into a regular training session, replacing 12 training trials, one of each configuration. No transfer trials appeared within the first or last eight trials of the session and at least five baseline trials separated transfer trials. Each test trial was made up of two novel images that were never seen before and never used again in testing. Correct responses were reinforced and incorrect responses resulted in timeout, as with training trials (there was no CP during transfer sessions). Like training sessions, transfer sessions were counterbalanced for left/right correct. Four consecutive transfer sessions were conducted, providing 48 novel transfer trials (involving 96 novel images).

Set-size expansion. After transfer testing, the training set size was doubled and the original training regimen ensued. The set size was systematically expanded from 3 to 6, 12, 24, 48, 96, 192, 384, and 768 images. Set-size expansion sessions were 96 trials, counterbalanced for left/right-correct. Sample and comparison images were randomly assigned from the stimulus set. Criterion performance was confirmed by a three-way mixed ANOVA of Group (experienced, naïve) × Trial Type (baseline, transfer) × Set Size (3, 6, 12, 24, 48, 96, 192, 384, 768), yielding a significant interaction of Trial Type × Set Size, F(8, 24) = 10.12, p < .001. There was no effect of Group or its interaction with any factor. Baseline performance decreased across set size whereas transfer performance increased, as verified by significant linear components of separate trend analyses (baseline: F(1, 4) = 26.86, p = .007; transfer: F(1, 4) = 18.141, p = .013). Percent correct on novel-transfer trials at set-size 3 was not different from chance (50%), t(4) = 1.53, p = .201, whereas transfer percent correct at all other set-sizes was greater than chance, t(4) > 2.8, p < .05. At set-sizes 384 and 768, transfer percent correct was not significantly different from baseline percent correct, t(4) < 2.31, p > .08.

Discussion

The present study is the first to show the range of the relationship between training set size and transfer in MTS. Percent correct on transfer trials in the present study increased from 55% at set-size 3 to 82.5% at set-size 768, a 27.5% increase. An important contribution of the present study is the trial-type analysis during the first-session of set-size expansion (top panel of Figure 1). Performance on each trial type provides insights into how pigeons solved the MTS task. That is, if pigeons learned MTS via configural rules, then performance on ST-FT trials should be high, and, because of the presence of untrained stimuli, performance on all the other trials should be low. If pigeons learned if-then rules, then performance on trials were those in which the sample and foil came from the new set of expansion stimuli. Sample Trained-Foil Untrained (ST-FU) trials were those in which the sample came from the pool of trained stimuli, but the foil came from the expansion pool. Sample Untrained-Foil Trained (SU-FT) trials were those in which the sample came from the expansion pool, but the foil came from the pool of trained stimuli.

The top panel of Figure 1 presents the mean percent correct for each of these four trial types during the first session of training at each set-size. A two-way repeated measures ANOVA of Trial Type (ST-FT, ST-FU, SU-FT, SU-FU) × Set Size (6, 12, 24, 48, 96, 192, 384, 768) revealed significant main effects of Trial Type, F(3, 12) = 53.5; p < .001, and Set Size, F(7, 28) = 6.06; p < .001, and a significant interaction F(21, 84) = 2.9; p < .001. The interaction was due to the consistently good performance across set size for ST-FT and ST-FU, whereas performance improved for SU-FU and SU-FT trial types, as confirmed by separate two-way repeated-measures ANOVAs of Trial Type (ST-FT, ST-FU) and (SU-FT, SU-FU) × Set Size. There were no main effects of Trial Type, or interaction, but there was a main effect of Set Size, F(7, 28), = 5.941, p < .001, for the ANOVA with the Trial Type (SU-FU, SU-FT).

Novel Transfer

The lines labeled MTS Baseline and Transfer in the bottom panel of Figure 1 represent mean baseline and novel-transfer performance for each training set-size. As the training set size increased, performance on novel-transfer increased, as confirmed by a three-way mixed ANOVA of Group (experienced, naïve) × Trial Type (baseline, transfer) × Set Size (3, 6, 12, 24, 48, 96, 192, 384, 768), yielding a significant interaction of Trial Type × Set Size, F(8, 24) = 10.12, p < .001. There was no effect of Group or its interaction with any factor. Baseline performance decreased across set size whereas transfer performance increased, as verified by significant linear components of separate trend analyses (baseline: F(1, 4) = 26.86, p = .007; transfer: F(1, 4) = 18.141, p = .013). Percent correct on novel-transfer trials at set-size 3 was not different from chance (50%), t(4) = 1.53, p = .201, whereas transfer percent correct at all other set-sizes was greater than chance, t(4) > 2.8, p < .05. At set-sizes 384 and 768, transfer percent correct was not significantly different from baseline percent correct, t(4) < 2.31, p > .08.
in which the sample had been trained (ST-FT and ST-FU) should be high and performance on the remaining trials should be low. If pigeons learned abstract concepts, then performance should be high on all trial types. In the present study, the consistently high performance on ST-FU relative to ST-FT trials suggests that pigeons did not learn configural rules, but instead learned if-then rules. As set-size increased, performance on trials in which the sample was untrained (SU-FT, SU-FU) increased from chance-level to a level equal to performance on sample-trained trials (ST-FT, ST-FU). This interaction suggests that learning strategies shifted from item-specific, if-then rules at small set sizes to abstract-concept learning at large set-sizes. These results contribute to the growing body of evidence that pigeons can learn abstract concepts.

Figure 1. Top panel: Mean percent correct on the first session of training at each set-size expansion for Sample Trained-Foil Trained (ST-FT), Sample Trained-Foil Untrained (ST-FU), Sample Untrained-Foil Trained (SU-FT), and Sample Untrained-Foil Untrained (SU-FU) trial types. Bottom panel: Mean baseline and novel-transfer percent correct across set size for pigeons in Matching-to-Sample (MTS, circles) and Same/Different (S/D, squares) tasks. Symbol fills distinguish between baseline (filled) and novel-transfer (open) percent correct. Error bars represent SEMs. S/D data reproduced from Katz & Wright, 2006 with permission.
**MTS and S/D Compared**

The effect of set-size on learning strategy has been previously demonstrated in pigeon (and monkey) abstract-concept learning in the S/D task (Katz & Wright, 2006). If MTS and S/D tasks produce qualitatively different learning strategies, as has been asserted (Pepperberg, 1987; Premack, 1978, 1983), then differences should be obvious by comparing baseline and transfer results from systematically expanded training-set sizes. As shown in the bottom panel of Figure 1, novel-transfer and baseline performance in MTS and S/D was similar except for some differences at the smaller set sizes (3–24 in MTS, 8–32 in S/D). The smaller initial set-size in MTS (3 images; 12 configurations) relative to S/D (8 images; 64 configurations) may have produced the higher baseline accuracy on MTS. If MTS and S/D training had begun with an equal number of training configurations, baseline and transfer performance may have been equivalent across tasks. Despite this and some other differences between procedures (e.g., the MTS group received eight set size expansions whereas the S/D group received six), both produced an increasing novel-transfer function across set size to the level where baseline and transfer performance were equivalent and similar across procedures (a qualitative similarity). Overall, these results suggest that set size similarly affects novel-transfer performance in both the MTS and S/D tasks.

**The Roles of Experience and Set Size**

A potential limitation of the present study is that the effect of experience in the MTS task was intertwined with set-size expansion. Taken alone, one might argue that the increase in performance across set-size expansion may be because of the extended practice in the MTS task. Katz and Wright (2006) addressed the issue of experience in the S/D task by fixing the training set at the smallest set-size (i.e., 8) for a training-control group while presenting novel-stimulus transfer tests at the same training intervals as a set-size expansion experimental group. The training-control group had the same training experience with the same transfer pairs at equivalent points in training but with a fixed set size. Pigeons in the training-control group failed all novel-transfer tests, thereby ruling out training experience as a possible explanation for the high level of transfer in the experimental group. We have no reason to doubt that the same result would occur in MTS.

**Novel-Transfer Acquisition**

The gradual improvement in novel-transfer performance across set-size expansions has raised the possibility that pigeons do not learn an abstract concept per se, but might generalize responding from training exemplars to similar-looking novel stimuli. As set size increases, more features (e.g., hues, shapes, etc.) are unavoidably shared between novel-test and baseline-training stimuli, possibly allowing pigeons to perform more accurately on novel-test trials (Mackintosh, 2000). These shared features may have allowed pigeons to fully generalize choice responses to novel-test stimuli in the MTS task.

Wright & Katz (in press) developed a model to test the generalization hypothesis of novel-stimulus transfer. In the S/D task, generalization would have to occur across item pairs, that is, by configurational learning. In the MTS task, however, generalization can occur by configural or if-then rule learning. As the analysis of first-session trial-type performance (top panel of Figure 1) ruled out configural learning, only the if-then prediction is presented.1

The generalization model assumes that any above-chance novel-transfer performance is because of generalization. The first parameter is the number of new if-then rules to be learned at each set size ($N_{pairs}$). Assuming that a new rule has to be learned for each stimulus, this number is obtained by subtracting the number of trained stimuli from the total number of stimuli in an expansion set:

$$N_{pairs} = SET_{new} - SET_{old}$$

For example, if the newly expanded set ($SET_{new}$) is 12, then the trained set ($SET_{old}$) is 6 and $N_{pairs} = 6$.

The second parameter is the number of trials to learn each new if-then rule ($TTA_{pair}$). This number is determined by dividing the number of trials to reach criterion at the initial training set size by the number of training items:

$$TTA_{pair} = \frac{Trials \ to \ 3-Item \ Acquisition}{Number \ of \ Items}$$

The initial training set size is used because no generalization occurred at this set size, as evinced by chance performance on novel-transfer. If the number of trials to acquisition at set-size 3 was 480, then $TTA_{pair} = 160$.

The third parameter is the proportion of new if-then rules that will not be accounted for by generalization and will need to be learned ($P_{pairs}$). Assuming that above-chance performance on novel-transfer is because of generalization, transfer performance predicts how many new if-then rules can be accommodated by generalization. The proportion of new if-then rules to be learned ($P_{pairs}$) for each expansion is determined by dividing the difference between the previous baseline and transfer percentages by the difference between baseline and chance percentages:

$$P_{pairs} = \frac{(Baseline\% - Transfer\%)}{(Baseline\% - Chance\%)}$$

For example, if the baseline and transfer percent correct were 90% and 65%, respectively, then the numerator would be (90 – 65), the denominator would be (90 – 50) and $P_{pairs} = 0.625$.

Together the three parameters predict the number of trials needed to reach criterion at each set-size expansion ($TTA_{predicted}$):

$$TTA_{predicted} = TTA_{pair} \times N_{pairs} \times P_{pairs}$$

Figure 2 presents the obtained and predicted number of trials to reach criterion across set size. The predicted if-then function diverged from the obtained results, as confirmed by a two-way repeated measures ANOVA of Condition (observed, predicted) × Set Size (6, 12, 24, 48, 96, 192, 384, 768) on trials-to-acquisition

---

1 The configurational prediction failed to fit the observed results, as confirmed by a two-way repeated measures ANOVA of Condition (observed, predicted) × Set Size (6, 12, 24, 48, 96, 192, 384, 768) on trials-to-acquisition which yielded a main effect of Condition, $F(1, 4) = 11.1, p = .029$, and Condition × Set Size interaction, $F(7, 28) = 7.2, p = .001$. All manipulations of $TTA_{pair}$ failed to fit the configurational prediction to the obtained results.
which yielded a main effect of Condition, $F(1, 4) = 12.2, p = .025$, and Condition $\times$ Set Size interaction, $F(7, 28) = 5.8, p = .001$. According to the model, generalization via if-then rule learning cannot account for the present results.

The model can be adjusted to explore the extent and limits to which the predictions can fit the data. See Wright & Katz (in press) for manipulations of learning-rate stability, generalization during learning, and learning rate changes based on familiarity with the stimuli. As one example, to test the possibility that the obtained results might be explained by increased acquisition rates across set-size expansions (i.e., learning to learn), the learning rate was gradually improved across set-size expansions such that learning was twice as fast (50% decrease in $TTA_{pair}$) at the point where baseline and transfer accuracy were equivalent (set-size 384). This modified prediction for if-then generalization (line $TTA_{pair} 50\%$ in Figure 2) also failed to fit the obtained results, as confirmed by a two-way repeated measures ANOVA of Condition (observed, predicted) $\times$ Set Size (6, 12, 24, 48, 96, 192, 384, 768), which yielded a main effect of Condition, $F(1, 4) = 11.3, p = .028$, and Condition $\times$ Set Size interaction, $F(7, 28) = 5.16, p = .001$.

The if-then prediction can be made to fit the obtained results if the learning rate is improved by 37% at each set-size expansion such that learning was 16 times faster at the point where baseline and transfer accuracy were equivalent (set-size 384). A two-way repeated measures ANOVA of Condition (observed, predicted) $\times$ Set Size (6, 12, 24, 48, 96, 192, 384, 768) on trials to acquisition yielded no main effects of Condition, $F(1, 4) = 1.87, p = .243$, or Set-Size, $F(7, 28) = 2.3, p = .054$, and no Condition $\times$ Set Size interaction $F(7, 28) = 1.15, p = .36$. However, such a precipitous change in learning rate contradicts the existing literature that has demonstrated highly stable learning rates of pigeons sorting arbitrarily selected images into two categories (Cook, Levison, Gillett, & Blaisdell, 2005). In the absence of evidence for such extreme changes in learning rate, it is implausible to suggest that learning-to-learn could account for the obtained results.

**Conclusion**

Both the first-session trial-type analysis and the novel-transfer tests suggest that pigeons did not learn the abstract concept at small set sizes. Consistent with many claims in the literature (Cumming, Berryman, & Cohen, 1965; Farthing & Opuda, 1974; Santi, 1978, 1982; for an exception see Wright, 1997), responding must have been tied to specific features of the training stimuli, although it is unclear what those features might be. As set size increased, improved performance on both novel-transfer tests and untrained stimuli, and the implausibility of generalization, combine to suggest that behavioral control by specific features was relinquished in favor of relational learning, that is, abstract-concept learning.

In summary, the present study demonstrates that pigeons can match novel stimuli as accurately as training stimuli in the MTS task. The present results also emphasize the importance of multiple-exemplar training by providing the first MTS set-size function for any species. Configural generalization is ruled out. If-then learning is suggested at small set-sizes, but ruled out at
large set sizes. The present study also provides the best MTS to S/D comparison to date, demonstrating that novel-transfer functions are similar for pigeons in each task. Taken together, these findings suggest that the role of set size is similar in both MTS and S/D tasks, and adds to the growing body of evidence that pigeons can learn abstract concepts.

References


Cook, R. G. (2002). Same-different concept formation in pigeons. In M. Bekoff, C. Allen & G. M. Burghardt (Eds.), The cognitive animal (pp. 229–237). Cambridge, MA: MIT.


Received April 30, 2007
Revision received September 21, 2007
Accepted September 25, 2007