



Learning an object from multiple views enhances its recognition in an orthogonal rotational axis in pigeons

Jessie J. Peissig^{a,*}, Edward A. Wasserman^a, Michael E. Young^b, Irving Biederman^c

^a Department of Psychology, University of Iowa, Iowa City, IA 52242, USA

^b Department of Psychology, Southern Illinois University at Carbondale, Carbondale, IL 62901, USA

^c Department of Psychology, The University of Southern California, Los Angeles, CA 90089, USA

Received 2 October 2001; received in revised form 3 April 2002

Abstract

In the natural environment, most objects are seen from several different viewpoints. We explored the nature of recognition after training with multiple views and compared it to recognition after training with only one view. Pigeons were taught with either five views or one view of each of four single-geon objects. Pigeons trained with five views responded more accurately to novel views of an object than did pigeons trained with only one view. This result held even when the novel views came from a rotational axis that was orthogonal to the training axis. These results do not accord with recognition processes involving mental rotation or direct interpolation. Pigeons trained with five views may have formed a view-invariant representation [Psychol. Rev. 94 (1987) 115; Vision Res. 39 (1999) 2885]; alternatively, they may have acquired a more detailed shape space of the objects in which to measure object similarity [Representation and recognition in vision, MIT Press, MA, 1999], or learned to attend to a broader range of features of each object [J. Exp. Anal. Behav. 54 (1990) 69].

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Pigeon (*Columba livia*); Object recognition; Rotation; Orientation

1. Introduction

Imagine that you learn to recognize the object depicted in Fig. 1a from this one view. If you are later shown different depictions of the object after it is rotated about the *y*-axis (by vertically piercing the object at its center and turning it around this axis; see Fig. 1c), then recognition might worsen as these rotations increasingly deviate from the training view. Such worsening of recognition performance has been found in pigeons, with distant views sometimes failing to produce above chance recognition (Peissig, Young, Wasserman, & Biederman, 1999, 2000; Wasserman et al., 1996).

One way to enhance the recognition of novel *y*-axis views is to learn the task with multiple views that are also *y*-axis rotations; now, interpolations from neigh-

boring views are possible. Such multiple-views training does indeed enhance generalization performance in pigeons (Peissig et al., 1999; Wasserman et al., 1996).

Several explanations can account for this improved performance. For example, training pigeons with multiple views may encourage them to ignore irrelevant details of the stimuli, such as the spatial location of specific line segments or patches of light and dark shading. After multiple-views training, the pigeons may learn to attend to features that are consistent across rotations in depth, such as the general shape of the objects. This type of improvement is consistent with theories of object recognition which propose that a general structural description of an object is stored in memory and used for normal recognition processes (Biederman, 1987, 2000; Hummel & Stankiewicz, 1996). Another possibility is that multiple-views training provides additional exemplars of the category, thus increasing the likelihood that a novel view will be similar to a known exemplar (cf. Bhatt & Wasserman, 1989) and providing a broader definition of the object class.

* Corresponding author. Present address: Department of Neuroscience, Brown University, Box 1953, Providence, RI 02912, USA.

E-mail address: jessie_peissig@brown.edu (J.J. Peissig).

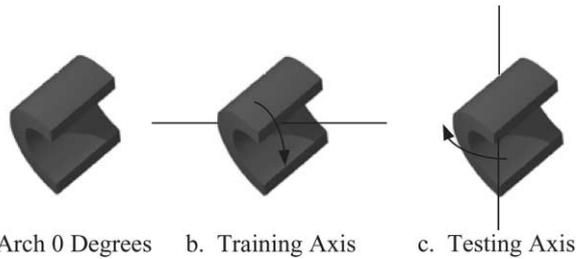


Fig. 1. (a) The arch shown at the training view (0°) for the one-view group; the five-views group was trained with this view and four others generated by rotating the objects at 72° intervals in the x -axis. (b) The direction of rotation used to generate views in the x -axis. (c) The direction of rotation used to generate views in the y -axis.

Improvement of this type is consistent with theories of recognition which propose that objects are represented in memory as individual views (Bülthoff & Edelman, 1992; Tarr & Pinker, 1989).

A second way to enhance the recognition of novel y -axis views is to learn the task with multiple views that are x -axis rotations (by horizontally piercing the object at its center and then turning it around this orthogonal axis; see Fig. 1b). The idea here is that multiple x -axis views might better create a representation of the object than a single view, even though no interpolations from nearby views are possible because of the orthogonality of the x -axis and y -axis rotations. The resulting representation may be a generalized three-dimensional model or a set of individual views that allows a broader specification of the object class.

Despite the obvious significance of studying visual recognition performance along orthogonal dimensions of rotation during training and testing, very few studies have explored this specific phenomenon. In one such experiment, Bülthoff and Edelman (1992) trained people with a movie of an object rotating slightly in one axis of rotation, and then tested them with novel views in the same axis and in an orthogonal axis of rotation. Although Bülthoff and Edelman tested generalization to views in an orthogonal axis, they did not explore the effects of single-view versus multiple-views training, certainly one of the most important variables in visual recognition performance. We did so in the present research project.

For our experimental subject, we chose the pigeon, an animal with no prior experience at recognizing two-dimensional pictorial representations of objects, but with excellent visual discrimination ability (Wasserman, 1991). To simplify the discrimination problem for the pigeon and to sharpen the theoretical interpretation for us, we used as discriminative stimuli two-dimensional computer renderings of four basic three-dimensional shapes (see Figs. 2 and 4) that would neither gain nor lose parts as they were rotated about the x - and y -axes

(cf. Wasserman et al., 1996; also see Biederman & Gerhardstein, 1993).

Pigeons that were given single-view training learned to peck four different buttons in response to four different shapes; each shape was seen at only one view. Pigeons that were given multiple-views training also learned to peck four different buttons in response to four different shapes; here, however, each shape was seen at five different rotations in depth about the x -axis. Then, all of the pigeons were tested with novel views of the four different shapes depicted at several rotations in depth about the y -axis. In addition, we explored whether the same pigeons that were previously trained with only a single view would exhibit increased stimulus generalization to novel views if they too were given additional training with multiple views. This result would provide convergent evidence that training with several views leads to improved recognition performance, even for pigeons that were originally trained with only one view.

The results were clear: (1) single-view x -axis training supported reliable y -axis generalization, but recognition precipitously fell as the testing views rotated away from the training view and (2) y -axis generalization performance was much higher after multiple-views than after single-view x -axis training. These results suggest that multiple-views training better creates representations of objects that support generalized recognition than does single-view training.

2. Method

2.1. Subjects

The subjects were eight feral pigeons maintained at 85% of their free-feed weights by controlled daily feeding. The pigeons had earlier participated in unrelated studies.

2.2. Apparatus

The pigeons were trained in four specially constructed plywood chambers. One side of each chamber consisted of a large opening with an aluminum frame attached to the outside of the box. Inside the frame was a clear touch screen (Accutouch Model #002744-FTM-K1; Elographics, Oak Ridge, TN) that was coated with mylar for durability. A serial controller board (Model #E271-2210, Elographics) processed the subjects' pecks to the touch screen. A brushed aluminum panel was placed directly in front of the screen to allow subjects access to limited portions of the video monitor. There were five openings in the aluminum panel. The center opening was a 7×7 cm square opening in which the stimuli appeared. The remaining four openings were circular, 1.9 cm in diameter, and were located 2.3 cm

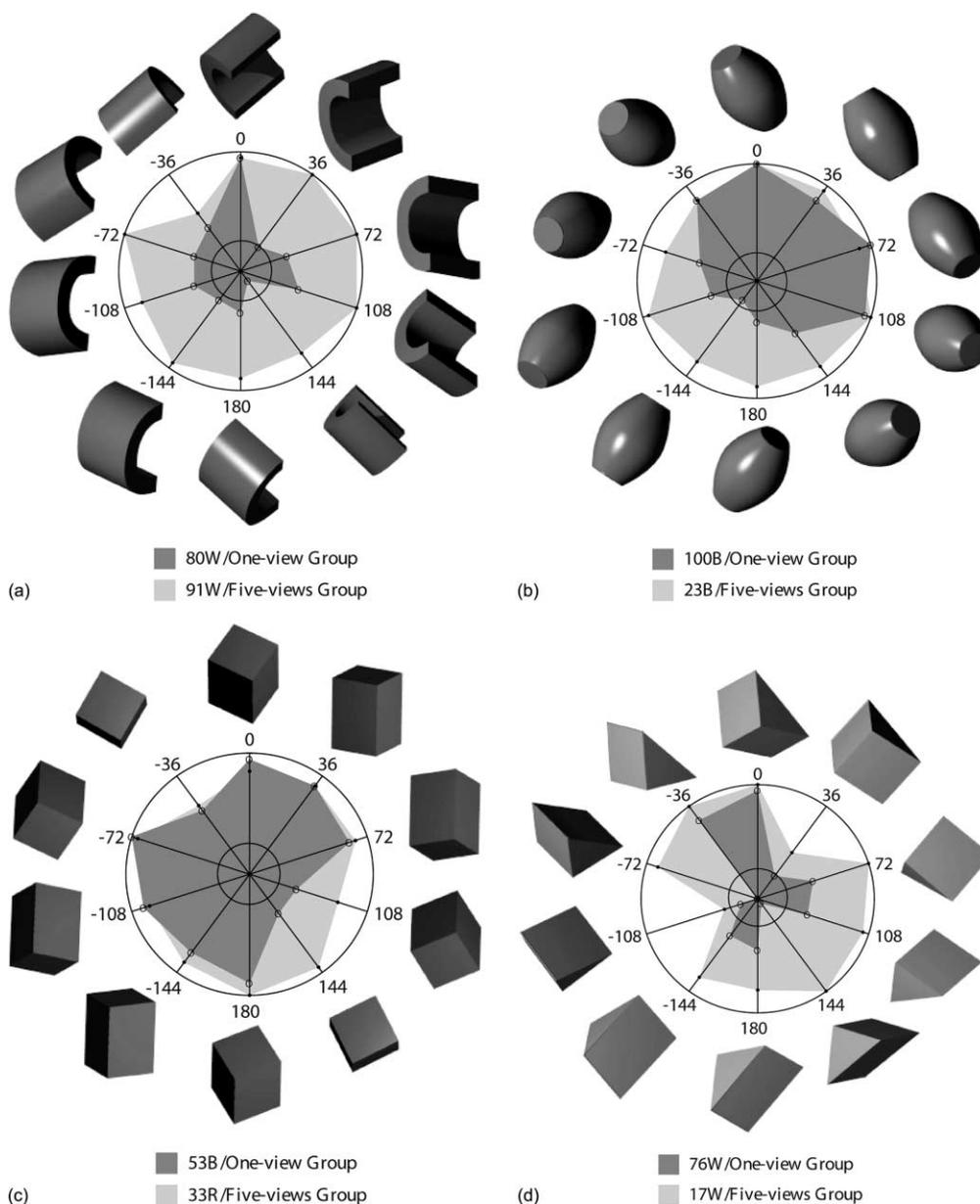


Fig. 2. The Phase 1 testing accuracy scores: (a) for pigeons 80W and 91W for the arch, (b) 100B and 23B for the barrel objects, (c) 53B and 33R for the brick, and (d) 76W and 17W for the wedge. The places where the dark gray (connected by open circles) intersects the spokes indicate the choice performance of the bird in the one-view group (it was trained at the 0° view). The places where the light gray (connected by filled circles) intersects the spokes indicate the choice performance of the five-views group (it was trained at the 0°, 72°, 144°, -144°, and -72° views). The solid inner circle indicates chance performance (25% correct).

from each corner of the center display opening. The four corner openings served as “report” buttons. In the rear of the chamber, a clear Plexiglas food cup was placed level with a wire mesh floor to prevent subjects from perching on the cup. Noyes 45-mg pigeon pellets were delivered through a vinyl tube into the cup using a rotary pellet dispenser (Model #ENV-203M; MED Associates, Lafayette, IN). During experimental sessions, constant illumination was provided by a houselight mounted on the upper rear wall of the chamber. A

digital I/O interface board (National Instruments Model #NB-DIO-24, Austin, TX) controlled the pellet dispenser and the houselight.

Control of the houselight, feeder, and stimulus presentation (via the I/O interface), and recording of pigeons’ responses (via the serial controller board) were accomplished by four Apple Macintosh 7100/66 Power PC computers. The pigeon’s monitor and an identical monitor located in an adjacent room were connected by a distribution amplifier (Model #MAC/2 DA2; Extron

Electronic, Santa Fe Springs, CA). The programs were developed in Hypercard 2.3.

2.3. Stimuli

The training stimuli (“geons”) consisted of two-dimensional representations of four three-dimensional objects: an arch, a barrel, a brick, and a wedge that varied from one another by a variety of nonaccidental properties (Biederman, 1987). The stimuli ranged in size from 2 to 4 cm in width and from 2 to 4 cm in height; they were rendered in Raydream Designer 4 at 300 dpi resolution. The objects were rotated in depth by 36° intervals plus or minus up to 6° within the x -axis to yield 10 views (see Fig. 2); rotational flexibility of 6° was necessary in order to prevent the objects from being shown at accidental views (Biederman, 1987). The geons were also rotated in depth by 36° intervals within the y -axis of rotation, yielding 10 views (see Fig. 4). One view, the 0° rotation, served as the starting point for rotation in both the x - and y -axes, as it fell within both rotational axes.

2.4. Procedure

At the beginning of a trial, the central display area was illuminated with a black cross centered on a white background. A single peck anywhere within that display area turned on a single-geon stimulus in the center. The pigeons were required to peck the center screen a fixed number of times (“observing responses”); the fixed ratio varied for different birds, ranging from 15 to 30 pecks per trial. After the final peck, the four corner report buttons were illuminated. After a correct choice, the stimulus was removed from the display area, the report buttons were darkened, and a food-pellet reinforcer was delivered; after an incorrect choice, the stimulus flashed off for 1 s, the report buttons were darkened, the houselight was turned off for 4–6 s (mean of 5 s), and one or more correction trials began (these were not scored for analysis). Intertrial intervals ranged from 6 to 10 s (mean of 8 s).

The experiment consisted of three phases: training, Phase 1 testing, and Phase 2 testing. There were two groups of pigeons: the one-view group and the five-views group, each comprising four birds. During training, the one-view group was taught to peck one of the four corner report buttons for the 0° view of each of the four single-geon objects. The 0° view was within both the x -axis and the y -axis because it was the starting point for each series of rotations (see Figs. 2 and 4). This procedure yielded one view of each of four objects for a total of four training stimuli in the one-view group. The five-views group was also taught to peck one of the four corner report buttons for the 0° view; pigeons in the five-views group were given four additional views of

each object rendered in the x -axis at 72° intervals from the 0° view. This procedure yielded five views of each of the four objects for a total of 20 training stimuli in the five-views group. Button assignments were counterbalanced within both groups using a Latin-square design.

In training, we used a randomized block design. For the one-view group, each block consisted of five presentations of a single view of the four objects; thus, there were a total of 20 trials in a single block. For the five-views group, each block consisted of one presentation of each of the five views for the four objects for a total of 20 trials in a single block. For both groups, there were 10 blocks in each daily session for a total of 200 trials per day. Each bird was required to meet a criterion of at least 85% correct overall; in addition, each pigeon was required to obtain a score of at least 80% correct for each of the four individual geons. This 85/80 criterion had to be met in at least one training session before a pigeon was allowed to proceed to testing. The accuracies of the five-views group for the individual geons were calculated by averaging performance across all five training views of each geon.

Phase 1 testing involved testing trials with novel rotations generated within the x -axis of rotation. Phase 2 testing involved testing trials with novel rotations within the y -axis of rotation (orthogonal to the axis of rotation for the training views in the five-views group).

In both phases of testing, trials involving testing stimuli were nondifferentially reinforced; the pigeons were given food reinforcement for any choice that they made on testing trials, so that repeated testing with the same stimuli would not teach the birds the correct response (if pecks were differentially reinforced) or cause the birds to cease responding to the testing stimuli (if pecks were not reinforced). This nondifferential reinforcement technique has been used many times in our laboratory with excellent results (e.g., Wasserman et al., 1996).

During Phase 1 testing, the five-views group was tested with five novel rotations rendered in the x -axis at approximately 36° intervals between the neighboring trained views. In addition to the five novel views of each geon, the training views were given as nondifferentially reinforced testing trials (0° , 72° , 144° , -144° , and -72° views) for a total of 10 testing views. The one-view group was tested with nine novel rotations of each geon rendered in the x -axis at 36° intervals from the 0° training view; these rotations were the same as those used for training and testing in the five-views group (36° , 72° , 108° , 144° , 180° , -144° , -108° , -72° , and -36° views). In addition, the 0° training view of each geon was given as a nondifferentially reinforced testing view. Including the training views as testing trials guaranteed that both groups would receive the same number of

testing trials within a session and yielded equivalent numbers of nondifferentially reinforced trials for each rotation. This process yielded a total of 40 testing views (10 views \times 4 geons) for each group.

Testing sessions began with 40 warm-up trials (20 different stimuli each shown twice for the five-views group and four different stimuli each shown 10 times for the one-view group). The remainder of the testing session followed a randomized block design; each block consisted of 25 trials. Twenty of the trials were the original training views (20 different stimuli each shown one time for the five-views group and four different stimuli each shown five times for the one-view group); the remaining five trials were testing stimuli. The five testing stimuli were chosen randomly, without replacement, from the total pool of 40 testing stimuli. There were eight blocks in each daily session for a total of 240 trials (40 differentially reinforced warm-up trials, 160 differentially reinforced training trials, and 40 nondifferentially reinforced testing trials). If a pigeon failed to meet the 85/80 criterion for the training trials by the end of a session, then it was returned to training for later sessions until it met the 85/80 criterion. Each pigeon continued until it was tested 20 times with each of the testing stimuli (20 sessions).

During Phase 2 testing, both the five views and the one-view groups were tested with nine novel rotations rendered at 36° intervals from the 0° training rotation within the y -axis. The axis of rotation for these novel views was orthogonal to the axis of rotation for the training views of the five-views group. In addition to the novel rotations, the 0° training view of each object was included as a nondifferentially reinforced testing trial. This process yielded equivalent numbers of nondifferentially reinforced trials for each view. There were a total of 40 testing views (10 views \times 4 geons).

Phase 2 testing sessions proceeded identically to those in Phase 1 testing. Each pigeon continued until it was tested 20 times with each of the testing stimuli (20 sessions).

Once the pigeons in the one-view group completed Phase 2 testing they were given additional training to make them equivalent to the five-views group. During this additional training, the one-view group was differentially reinforced for responses to the four additional views of each object (72°, 144°, -144°, and -72° views) as well as with the previously trained 0° view. These training sessions proceeded in an identical manner to those given to the five-views group. Once they met the 85/80 criterion, the pigeons in the one-view group were retested in both Phases 1 and 2 testing using the same procedure that was used to test the five-views group. Each pigeon in the one-view group remained in testing until it was retested 20 times with each of the testing stimuli (20 sessions) in each phase of testing.

3. Results

The pigeons in the one-view group took a mean of 30 days to complete the training phase; the pigeons in the five-views group took a mean of 28 days to complete the training phase (two pigeons were replaced in the five-views group because they failed to learn). During Phase 1 testing, the one-view group required a mean of 13 additional days of training and the five-views group required a mean of 21 additional days of training. During Phase 2 testing, the one-view group required a mean of 10 additional days of training and the five-views group required a mean of 23 days of additional training. The five-views discrimination thus seemed to be a bit more difficult to sustain during testing than did the one-view discrimination.

In Phase 1 testing, the novel views were rendered within the same axis of rotation as the training views of the five-views group. Fig. 2 comprises four polar plots showing the accuracy of a representative pigeon within each group (one view and five views) for each of the four objects. This figure gives a snapshot of a single bird's performance to an individual geon for Phase 1 testing. These specific birds were chosen because their data were relatively representative of the performance of all the birds within each group. Although the pigeons exhibited individual variability in response accuracies to the novel views, the general pattern of results is evident even in single birds from each group. Fig. 3 is a polar plot showing the overall accuracy for Phase 1 testing of the one-view and five-views groups. The percentage of correct choices for each view was averaged across geons. In this figure, the dashed circle indicates discrimination performance that is significantly above chance (one-tailed binomial, $p < 0.05$). The five-views group responded at much higher levels of accuracy to the novel views than did the one-view group. The one-view group exhibited a large decrement in accuracy from the training view to the novel testing views; these pigeons had a mean of 94% correct for the 0° training view and 51% correct for the novel views during Phase 1 testing, a difference of 43% correct. The five-views group exhibited only a small decrement from the training to the testing views; these pigeons had a mean of 90% correct for the training views and 82% correct for the novel testing views, a difference of only 8% correct. The pigeons in both groups responded above chance to all novel views (binomial; $p < 0.05$).

To compare the one-view and five-views groups in Phase 1 testing, we analyzed the data from the views that were novel for both groups: the 36°, 108°, 180°, -108°, and -36° views. The data were assessed using a repeated-measures, full-factorial analysis of variance (ANOVA) with view (36°, 108°, 180°, -108°, and -36°) and geon (arch, barrel, brick, and wedge) as within-subject variables and group (one-view and five-views) as

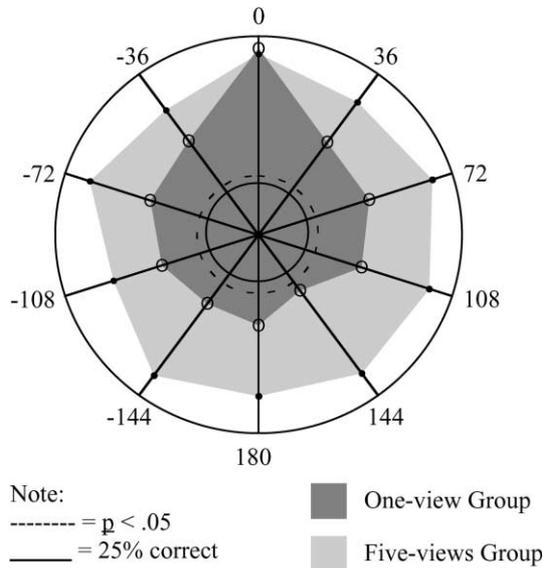


Fig. 3. The results of Phase 1 testing within the trained axis of rotation (x -axis). Plotted on the polar spokes is the mean choice accuracy (range: 0–100%) at each view. The places where the dark gray (connected by open circles) intersects the spokes indicate the choice performance of the one-view group (the training view was at 0°). The places where the light gray (connected by filled circles) intersects the spokes indicate the choice performance of the five-views group (the training views were at 0° , 72° , 144° , -144° , and -72°). The solid inner circle is chance performance (25% correct) and the dashed outer circle indicates the criterion for above-chance statistical significance.

a between-subjects variable. The main effect of view was not significant [$F(4, 24) = 2.57$, $p > 0.05$], indicating that there were no significant differences among the views that were tested. The main effect of geon was not significant [$F(3, 18) = 0.51$, $p > 0.05$]. Thus, there were no general performance differences among the individual geons. The main effect of group was significant [$F(1, 6) = 28.99$, $p < 0.01$], indicating that pigeons in the five-views group responded at a significantly higher level of accuracy to the novel views (82% correct) than did pigeons in the one-view group (53% correct). There was significant interaction between view and geon [$F(12, 72) = 7.19$, $p < 0.001$]. As is evident from the individual bird data in Fig. 2, the pigeons vary in the degree to which they generalize to novel views of each geon. The geon \times group interaction was not significant [$F(3, 18) = 0.08$, $p > 0.05$], indicating that the five-views group exhibited higher levels of performance than the one-view group to the novel views of all the geons. The view \times group interaction was not significant [$F(4, 24) = 2.66$, $p > 0.05$], suggesting that the two groups showed similar patterns of responding to the five different views. The three-way interaction of view \times geon \times group was significant [$F(12, 72) = 3.69$, $p < 0.0001$]. This interaction indicates that there were differences in how the two groups generalized to the specific views of individual geons.

In Phase 2 testing, the novel views were rendered in the y -axis of rotation, which was orthogonal to the axis of rotation that was used to generate the training views of the five-views group. Fig. 4 comprises four polar plots showing the accuracy of a representative pigeon within each group (one-view and five-views) for each of the four objects. This figure gives a snapshot of a single bird's performance on an individual object for Phase 2 testing. These specific birds were chosen because their data were relatively representative of the performance of all the birds within each group. Once again, the general pattern of results is evident even in single birds from each group. Fig. 5 is a polar plot showing the mean accuracy of the four pigeons in each group averaged across the four objects during Phase 2 testing. The five-views group exhibited higher levels of accuracy to the novel views than did the one-view group. The one-view group exhibited a large decrement in accuracy from the training view to the novel testing views; the pigeons had a mean of 94% correct for the 0° training view and 43% correct for the novel views during Phase 2 testing, a difference of 51% correct. The five-views group also exhibited a decrement from the training to the testing views, but the decrement was not as large as that exhibited by the one-view group; the pigeons in the five-views group had a mean of 90% correct for the training views and 64% correct for the novel testing views, a difference of only 26% correct. The pigeons in both groups responded above chance to all novel views (binomial; $p < 0.05$).

To compare the one-view and five-views group in Phase 2 testing, we analyzed the data from the novel views: the 36° , 72° , 108° , 144° , 180° , -144° , -108° , -72° , and -36° rotations. The data were assessed using a repeated-measures, full-factorial ANOVA with view (36° , 72° , 108° , 144° , 180° , -144° , -108° , -72° , and -36°) and geon (arch, barrel, brick and wedge) as within-subject variables and group (one-view and five-views) as a between-subjects variable. The main effect of view was significant [$F(8, 48) = 3.74$, $p < 0.01$], suggesting that the birds did not respond equally to all nine views. The main effect of geon was not significant [$F(3, 18) = 1.79$, $p > 0.05$]. Thus, there were no general performance differences among the individual objects. The main effect of group was significant [$F(1, 6) = 18.81$, $p < 0.01$], indicating that pigeons in the five-views group responded at a significantly higher level of accuracy (64% correct) than did pigeons in the one-view group (43% correct). There was significant interaction between view and geon [$F(24, 144) = 4.66$, $p < 0.001$]. As is evident from the individual bird data in Fig. 4, the pigeons vary in the degree to which they generalize to novel views of each geon. The geon \times group interaction was not significant [$F(3, 18) = 1.03$, $p > 0.05$], indicating that the five-views group exhibited higher levels of performance than the one-view group to the novel views of

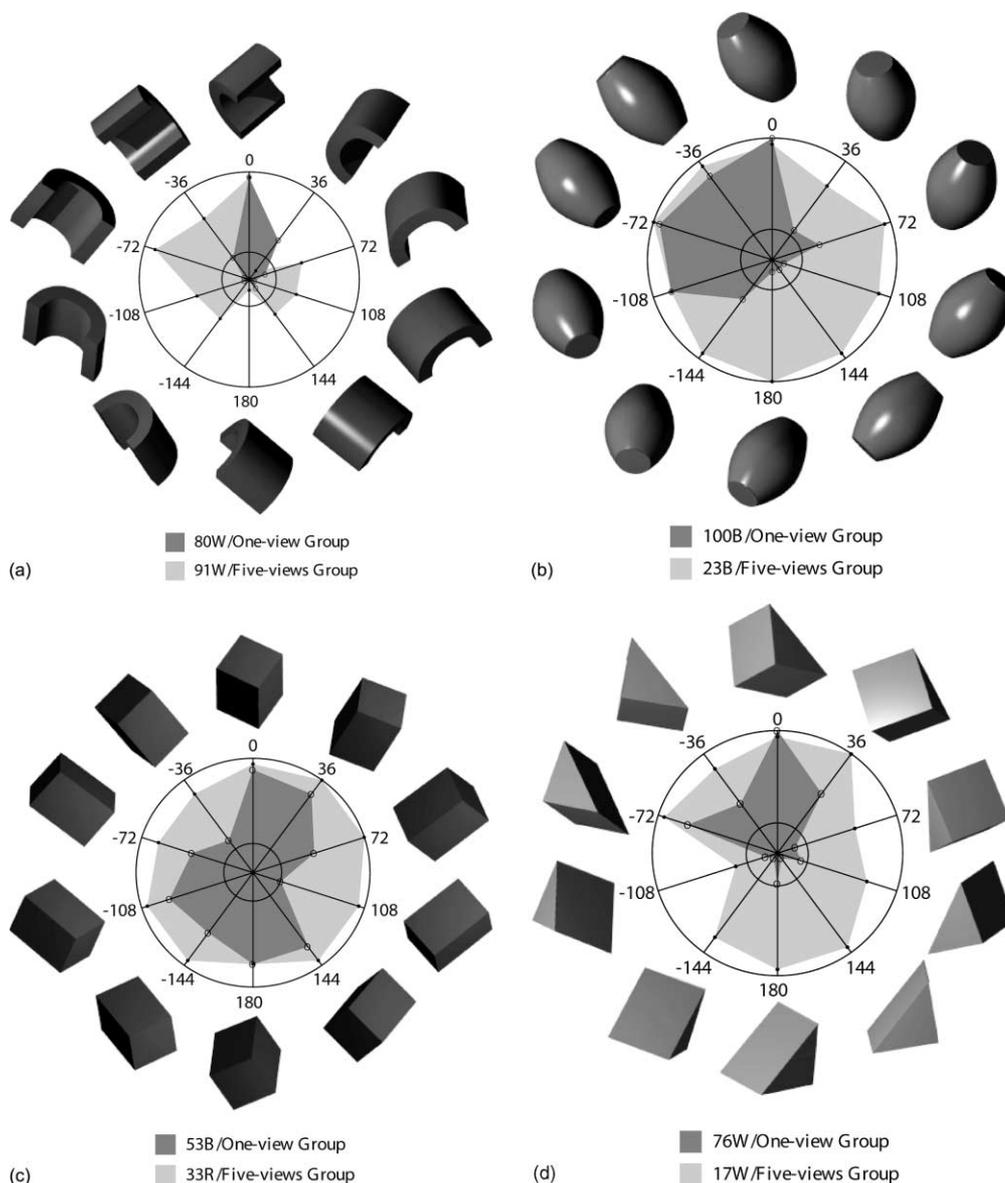


Fig. 4. The Phase 2 testing accuracy scores for (a) pigeons 80W and 91W for the arch, (b) 100B and 23B for the barrel objects, (c) 53B and 33R for the brick, and (d) 76W and 17W for the wedge. The places where the dark gray (connected by open circles) intersects the spokes indicate the choice performance of the bird in the one-view group (it was trained at the 0° view). The places where the light gray (connected by filled circles) intersects the spokes indicate the choice performance of the five-views group (it was also trained at only the 0° view in this axis). The solid inner circle is chance performance (25% correct).

all the objects. The view \times group interaction was significant [$F(8, 48) = 3.43, p < 0.01$]. This result indicates that the pattern of responding to the novel views was different in each group, perhaps due to more progressive decrements from rotation in the one-view group than in the five-views group; the five-views group responded similarly to all of the novel views in the y -axis, rather than showing systematically decreased accuracies when views were rotated farther from the training view, as in the one-view group. The three-way interaction of view \times geon \times group was also significant [$F(24, 144) = 1.72, p < 0.05$]. This interaction indicates that there were

differences in how the two groups generalized to the specific views of individual objects.

After completing Phases 1 and 2 testing, the pigeons in the one-view group were given additional training with the same five views in the x -axis used in training the five-views group. This additional training allowed us to compare a within-subjects manipulation of the number of training views to the between-subjects manipulation that we had already implemented. The pigeons in the one-view group required a mean of 18 days additional training with all five views to reach criterion. This amount of training time was shorter than was required

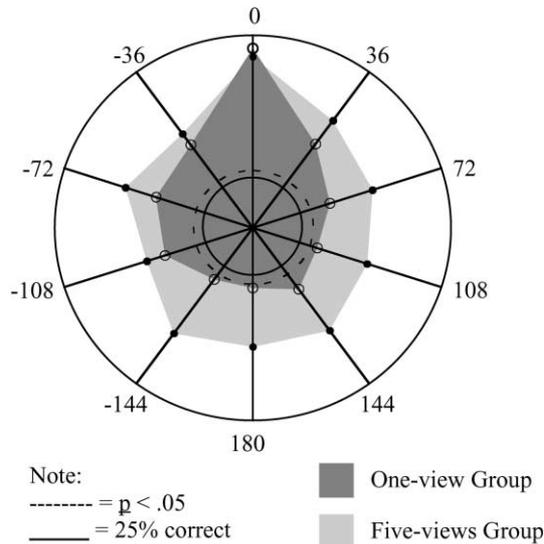


Fig. 5. The results of Phase 2 testing within the orthogonal axis of rotation (y -axis). Plotted on the polar spokes is the mean choice accuracy (range: 0–100%) at each view. The places where the dark gray (connected by open circles) intersects the spokes indicate the choice performance of the one-view group (the training view was at 0°). The places where the light gray (connected by filled circles) intersects the spokes indicate the choice performance of the five-views group (the training view was only at 0° for this axis). The solid inner circle is chance performance (25% correct) and the dashed outer circle indicates the criterion for above-chance statistical significance.

for the five-views group (28 days) and it was also shorter than the period of training required by these same birds to meet criterion for the one-view discrimination (30 days). Thus, after additional training with all five views, the one-view group learned the discrimination more quickly than the five-views group, indicating that previous one-view training provided some benefits when learning additional object views.

Prior to training with the additional views, the pigeons in the one-view group exhibited a large decrement in accuracy from the training view to the novel testing views in Phase 1 testing; these pigeons had a mean of 94% correct to the 0° training view and 51% correct to the novel views, a difference of 43% correct. After the training with the additional views, however, these pigeons exhibited only a small decrement from the training to the testing views in Phase 1 testing; they showed a mean of 93% correct to the training views and 86% correct to the novel testing views, a difference of only 7% correct.

The pigeons in the one-view group exhibited similar performance changes in Phase 2 testing after training with the additional views. Prior to training these pigeons with five views, the pigeons exhibited a large decrement in accuracy from the training view to the novel testing views in Phase 2 testing; they showed a mean of 94% correct to the 0° training view and 43% correct to the

novel views during Phase 2 testing, a difference of 51%. After additional training, however, the pigeons exhibited a much smaller decrement from the training to the testing views; they showed a mean of 96% correct to the training views and 69% correct to the novel testing views, a difference of only 27% correct.

4. Discussion

In the present experiment, we explored the pigeon's ability to form a view-invariant object representation after discrimination training with a single view versus multiple views. In previous studies in which pigeons were trained with only a single object view, the birds exhibited a sharp decrease in accuracy to novel views (Peissig et al., 1999, 2000). From these studies alone, it was unclear whether the pigeons were capable of forming a robust representation that would allow them to respond at a high level of accuracy in spite of rotational transformations. It is clear from the present study that pigeons are indeed able to form a representation that can be used more generally in recognition tasks.

The results of this experiment clearly showed that, when pigeons were trained with multiple views within one axis of rotation, this training led to improved stimulus generalization to novel views. In Phase 1 testing, the pigeons were tested with novel views that were generated within the same axis of rotation as that used to generate the training views of the five-views group (x -axis). The results indicated that the pigeons trained with multiple views exhibited a significantly higher level of recognition performance to these novel views than did the pigeons trained with a single view. Although increased stimulus generalization in the five-views group may be due to the use of a more view-invariant object representation, the pigeons in this group could also have used individual view representations to perform this task. In the five-views group, each testing view was approximately 36° from two of the five training views. In the one-view group, the smallest rotational distance for a novel view was 36° from the one trained view of each object, but the rotational distance was greatest at 180° from the trained view. Therefore, the difference in accuracy for the one-view and five-views groups could be attributed to a normalization process, such as mental rotation, that produced rotation costs that were imperceptible at only 36° of rotation. At the greater rotational distances from the training views experienced by the one-view group, however, the rotation costs could have been more readily detected. The difference between groups could also be attributed to a similarity-based process in which closer views are perceived as being more similar than distant views.

In Phase 2 testing, the amount of rotation from the training view and the number of trained views were no longer confounded in the two training groups. By testing along an axis that was orthogonal to the training axis of the five-views group, we could—without qualification—assess the effect of training with multiple views. Even though both groups of pigeons experienced only one training view within the testing axis, the pigeons trained with five views exhibited accuracies that were significantly higher than those exhibited by the pigeons trained with one view. Therefore, the pigeons could not have used direct interpolation between training views to assist with the recognition process in the orthogonal axis. In addition, the amount of mental rotation required for novel views in the orthogonal axis should have been equivalent in both groups; for both the one-view and five-views groups, only the 0° training view was within the orthogonal testing axis. These results suggest that pigeons do not use mental rotation or direct interpolation in this recognition task; the results instead imply that training with multiple views of an object produces a representation that better supports generalized recognition through other mechanisms.

In addition to this between-subjects test of object recognition, the pigeons in the one-view group were subsequently trained with multiple views in the x -axis and tested a second time with novel views in both the x - and the y -axes as a within-subjects test of training with multiple views. The additional training given to the one-views group resulted in large increases in generalization accuracy to novel views in both the x -axis and y -axis of rotation. Our results are thus robust and can be obtained in both within- and between-subject designs.

4.1. *Mental rotation and interpolation*

The results of this experiment suggest that pigeons are not using mental rotation or interpolation in our recognition task. Mental rotation has been proposed by several viewer-based theories of recognition as a process that is used to align the retinal image of an object with its stored representation. To illustrate this process, consider viewer-based theories which propose that organisms store in memory individual views of objects (Bülthoff & Edelman, 1992; Tarr & Bülthoff, 1995; Ullman, 1998). During the recognition process, retinal images of viewed objects must be transformed so they can be compared with the stored views. Tarr and Pinker (1989) suggested that mental rotation may be the mechanism that permits the visual system to align the retinal image of an object with its stored representations in memory; thus, the retinal image is rotated in the “mind’s eye” for alignment with the closest stored view. This suggestion was based on numerous studies demonstrating linear time increases when people were instructed to mentally rotate one stimulus to see

if it matched another (Cooper, 1975; Jolicoeur, 1985). These time increases in mental rotation experiments were similar to the increases in recognition time reported in experiments in which people were trained with one view of an object and tested with rotated views of the object.

Proponents of viewer-based theories have more recently suggested normalization mechanisms other than mental rotation for object recognition in humans (Tarr & Bülthoff, 1998; Tarr, Bülthoff, Zabinski, & Bland, 1997). The data supporting alternative normalization mechanisms in humans, however, do not eliminate mental rotation as the normalization mechanism used by pigeons to recognize rotated objects. In a study exploring mental rotation in pigeons, Hollard and Delius (1982) presented pigeons and people with a match-to-sample task using two-dimensional objects. Both groups of participants were required to choose either the identical object or its mirror reflection. During the test phase, the comparison and sample stimuli were shown at different planar rotations. Hollard and Delius reported results that were consistent with the use of mental rotation in humans, but not for pigeons. Consequently, their results argue against the pigeon’s use of mental rotation during recognition.

Hamm, Matheson, and Honig (1997) suggested that Hollard and Delius’s results (1982) indicated that the particular task they used was insensitive to measurements of mental rotation in pigeons. To address this issue, Hamm et al. (1997) used a go/no-go paradigm to test whether pigeons were sensitive to the planar rotation of objects in a recognition task. They found that pigeons produced behavioral patterns that were nearly identical to those reported in experiments studying mental rotation in people. Hamm et al.’s results, along with the results of others (Neiwirth & Rilling, 1987), suggest that pigeons may be capable of mental rotation and use this mechanism in visual recognition tasks.

In our experiments, the 0° view would be the closest training stimulus to all of the testing stimuli in the y -axis for pigeons trained with one or five views in the x -axis. Therefore, if the pigeons were using mental rotation in our recognition task, then there should be no difference between the performance of the one-view and five-views groups for novel views in the y -axis. The pigeons trained with five views, however, exhibited significantly higher accuracies to the novel y -axis rotations than did the pigeons trained with one view. Clearly, our results suggest that pigeons are not using mental rotation to recognize objects in this particular task. These data are consistent with arguments that mental rotation does not provide a parsimonious account of recognition data, and is thus unlikely to be the transformation process that is used for object recognition (Perrett, Oram, & Ashbridge, 1998), if indeed object recognition requires a transformation process at all.

A strict interpolation model can also be eliminated as a possible account of recognition in pigeons. In the interpolation model, stored views are transformed to match the retinal image by interpolating between the nearest stored views in the same axis of rotation (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Poggio & Edelman, 1990). This model predicts that recognition will be more accurate for novel views that fall between the stored views and less accurate for views that fall outside the stored views in the rotational axis. In our experiments, the one-view and five-views groups were trained with only one view in the y -axis of rotation. Thus, neither group could perform this recognition task by interpolating between stored views in the y -axis. Accordingly, both groups would be forced to use an alternative recognition process and should perform equivalently—but, they did not do so.

4.2. *Other theoretical accounts*

Despite the elimination of two prominent explanations of the pigeons' recognition behavior, our data can still be accounted for by more than a single model of object recognition. Geon theory, for example, proposes that organisms represent objects as collections of the individual object parts (or geons) and their spatial relations (Biederman, 1987, 2000; Biederman & Gerhardstein, 1993). In geon theory, the orientation of a geon should be taken into account as a feature of that geon and changes in this feature may cause decrements in performance (Hummel & Biederman, 1992). Although performance at all testing rotations was above chance for both groups, the pigeons trained with a single view may have allocated more attention to the orientation of the geon due to extensive training with each object at a single view. Extensive training at a single view may also train pigeons to attend to features of the objects that do not generalize across views (e.g., shading differences), producing the significant recognition costs observed in the one-view group. Training with more than one view may teach the pigeons to ignore these view-specific features. In addition, multiple-view training exposes the pigeons to varying orientations of individual geons, thereby reducing rotation costs. The pigeons may then learn to attend to the characteristics that all views of an object share (the nonaccidental properties, see Biederman, 1987).

It is also possible that the pigeon's behavior reflects the operation of an intricate similarity-matching mechanism within a shape space (Edelman, 1999). Edelman proposed that representations of object views are mapped into a shape space. This shape space is defined as a multi-dimensional feature map in which the points correspond to different views of objects. The similarity of views of two (identical or different) objects is judged

by the relative distance of the corresponding points in this multi-dimensional map. Thus, two object views that are next to one another in shape space would be judged to be more similar to one another than two object views that are distant from one another in shape space.

In Edelman's recognition model, an encounter with a novel view of an object locates the new view in the current shape space. A novel view of an object should be more similar to stored views of the same object than to stored views of completely different objects. Therefore, training with a single view provides only a single point of reference for the novel view, whereas training with five views provides five points of reference, making for a greater likelihood that the novel views from the orthogonal axis would be similar (i.e., be close in shape space) to one of the five trained views (see Edelman (1999) for more details about this theory of recognition).

Finally, the multiple feature theory of categorization can be applied to our data (Fersen & Lea, 1990; Huber & Lenz, 1996; Jitsumori & Yoshihara, 1997). This theory is similar to Edelman's theory of recognition (1999), although it was originally proposed to account for the pigeon's ability to learn polymorphous categories of stimuli, both natural and artificial. In this theory, no single feature is necessary or sufficient to determine category membership, but if a stimulus contains enough features associated with a specific category of objects, then it will be classified as belonging to that category. For example, trees are associated with features such as a roughly-textured main trunk that meets the ground, branches protruding from the trunk, green leaves on the branches, and a shape that is larger at the top than it is at the bottom. Although not every tree will exhibit all of these features, objects that have a subset of these features are more likely to be classified as trees than objects that have none of these features.

Accordingly, in our study, the pigeons may have learned to attend to specific features of each of the objects; the more features the new view shares with a trained view, the more likely the pigeon will be to choose the response associated with that trained view. Thus, the pigeons trained with five views should have had a larger pool of features associated with that object and should have exhibited better generalization to new views of the same object.

Consequently, Biederman's geon theory (1987), Edelman's similarity matching theory (1999), and the multiple feature theory (Fersen & Lea, 1990) can account for our data. Testing which of these theories most accurately describes the process of object recognition in the pigeon will not be easy, although geon theory does propose a specific form of representation, one that is lacking in Fersen and Lea's and Edelman's accounts. In the current study, the theories make nearly identical behavioral predictions. Although there is some disagreement (Edelman, 1999), the conceptual differences

between these theories may be more apparent than real. These theoretical perspectives are often caricatured in order to draw distinctions, but the theories continue to evolve in response to the available empirical data.

Our investigative approach is to gather data that will more fully inform and constrain these evolving theories and their suitability as accounts of avian object recognition. Although the precise recognition mechanism is unknown, our data clearly show that exposure to multiple views of an object is necessary for a pigeon to show robust recognition of the object at novel depth rotations. More importantly, a relatively small number (five) of depth-oriented views proved sufficient to produce generalized recognition of views occurring in an axis that is orthogonal to the presented rotations of an object. We can certainly surmise that a pigeon in its natural environment will rapidly encounter enough views of an object to support its recognition of nearly any and all novel views of that object.

The results of our study indicated that pigeons are capable of flexibly using object representations in a forced-choice recognition task. If the reason why exposure to different views was required for strong generalization was that the pigeons had to learn which cues were reliable (e.g., shape, not shading), then an expectation of geon theory would be that the pigeons would show more rapid generalization to a new set of geons from a single exposure. The flexibility exhibited by the pigeons should also be apparent in tests of recognition other than rotation, such as recognizing occluded objects. Future studies will explore the robustness of the pigeon's visual recognition flexibility under a variety of conditions and within a number of paradigms, and determine with greater certainty the precise nature of the pigeon's object representation.

Acknowledgements

This research was supported by a National Institute of Mental Health Grant, MH 47313. We would like to thank Christina Antes and Michael Prior for their assistance collecting data.

References

- Bhatt, R. S., & Wasserman, E. A. (1989). Secondary generalization and categorization in pigeons. *Journal of The Experimental Analysis of Behavior*, 52, 213–224.
- Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94, 115–147.
- Biederman, I. (2000). Recognizing depth-rotated objects: a review of recent research and theory. *Spatial Vision*, 13, 241–253.
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1162–1182.
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Science USA*, 89, 60–64.
- Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, 7, 20–43.
- Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.
- Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32, 2385–2400.
- Fersen, L. V., & Lea, S. E. G. (1990). Category discrimination by pigeons using five polymorphous features. *Journal of the Experimental Analysis of Behavior*, 54, 69–84.
- Hamm, J., Matheson, W. R., & Honig, W. K. (1997). Mental rotation in pigeons (*Columba livia*)? *Journal of Comparative Psychology*, 111, 76–81.
- Hollard, V. D., & Delius, J. D. (1982). Rotational invariance in visual pattern recognition by pigeons and humans. *Science*, 218, 804–806.
- Huber, L., & Lenz, R. (1996). Categorization of prototypical stimulus classes by pigeons. *The Quarterly Journal of Experimental Psychology*, 49B, 111–133.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480–517.
- Hummel, J. E., & Stankiewicz, B. J. (1996). Categorical relations in shape perception. *Spatial Vision*, 10, 201–236.
- Jitsumori, M., & Yoshihara, M. (1997). Categorical discrimination of human facial expressions by pigeons: a test of linear feature model. *The Quarterly Journal of Experimental Psychology*, 50B, 253–268.
- Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory and Cognition*, 13, 289–303.
- Neiworth, J. J., & Rilling, M. E. (1987). A method for studying imagery in animals. *Journal of Experimental Psychology: Animal Behavior Processes*, 13, 203–214.
- Peissig, J. J., Young, M. E., Wasserman, E. A., & Biederman, I. (1999). The pigeon's perception of depth-rotated shapes. *Cahiers de Psychologie (Current Psychology of Cognition)*, 18, 657–690 (invited paper).
- Peissig, J. J., Young, M. E., Wasserman, E. A., & Biederman, I. (2000). Seeing things from a different angle: the pigeon's recognition of single geons rotated in depth. *Journal of Experimental Psychology: Animal Behavior Processes*, 26, 115–132.
- Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. In M. J. Tarr, & H. H. Bülthoff (Eds.), *Object recognition in man, monkey, and machine* (pp. 111–145). Cambridge, MA: MIT Press.
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343, 263–266.
- Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1494–1505.
- Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. In M. J. Tarr, & H. H. Bülthoff (Eds.), *Object recognition in man, monkey, and machine* (pp. 1–20). Cambridge, MA: MIT Press.
- Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint. *Psychological Science*, 8, 282–289.
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21, 233–282.
- Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition*, 67, 21–44.
- Wasserman, E. A. (1991). The pecking pigeon: a model of complex visual processing? A review of quantitative analyses of behavior

(Volume XIII): behavioral approaches to pattern recognition and concept formation. In M. L. Commons, R. J. Herrnstein, S. M. Kosslyn, & D. M. Mumford (Eds.), *Contemporary psychology* (36) (pp. 605–606).

Wasserman, E. A., Gagliardi, J. L., Cook, B. R., Kirkpatrick-Steger, K., Astley, S. L., & Biederman, I. (1996). The pigeon's recognition of drawings of depth-rotated objects. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 205–221.