

Recognition by Humans and Pigeons of Novel Views of 3-D Objects and Their Photographs

Alinda Friedman, Marcia L. Spetch, and Anne Ferrey
University of Alberta

Humans and pigeons were trained to discriminate between 2 views of actual 3-D objects or their photographs. They were tested on novel views that were either within the closest rotational distance between the training views (interpolated) or outside of that range (extrapolated). When training views were 60° apart, pigeons, but not humans, recognized novel views of actual objects better than their pictures. Further, both species recognized interpolated views of both stimulus types better than extrapolated views, but a single distinctive geon enhanced recognition of novel views only for humans. When training views were 90° apart, pigeons recognized interpolated views better than extrapolated views with actual objects but not with photographs. Thus, pigeons may represent actual objects differently than their pictures.

The ability to recognize objects in the environment across different viewpoints is one of the most fundamental tasks the visual system must accomplish and is critical for the survival of both humans and animals. Yet much remains to be understood about the visual and cognitive mechanisms that allow organisms to accomplish this task. Not surprisingly, a great deal of research and theory has been devoted to understanding these processes in humans (e.g., Biederman & Gerhardstein, 1993; Bühlhoff & Edelman, 1992; Edelman, 1999; Tarr & Pinker, 1989) and, more recently, in animals (e.g., Logothetis, Pauls, Bühlhoff, & Poggio, 1994; Peissig, Wasserman, Young, & Biederman, 2002; Spetch & Friedman, 2003; Spetch, Friedman, & Reid, 2001; Wasserman et al., 1996).

For practical reasons, research on object recognition has relied almost exclusively on the use of two-dimensional (2-D) representations of three-dimensional (3-D) objects, in the form of drawings, photographs, or digitized images. Such pictorial displays allow rapid and automated presentations of the objects from exact preselected viewpoints. However, pictorial representations of objects—even photographs—do not provide all of the visual information that is available and that may influence object recognition

in the real world. Further, recognizing objects from pictures may involve interpretive or other processes in addition to those used in real-world object recognition, and such processes may vary with evolution, development, or experience (e.g., DeLoache & Burns, 1994; Deręgowski, 1989; Fagot, 2000; Nicholson & Seddon, 1977). Therefore, investigations of object recognition using pictorial stimuli may be tapping into processes that underlie the interpretation of pictures instead of, or in addition to, processes that underlie object recognition, and theories of object recognition based exclusively on research using pictorial stimuli may not provide a complete description of object recognition processes. This is an especially important consideration in comparative research on differences between species or differences across development. For example, when differences between humans and animals in recognition of depth-rotated objects are revealed (e.g., Spetch et al., 2001), it is important to determine whether these differences reflect mechanisms of picture perception, object recognition, or both.

The main goal of the present study was to explore differences between humans and pigeons in object recognition as a function of whether actual objects or their photographed images were used as stimuli. We used the same task and stimuli across both species to test the generality of findings we recently reported in two comparative studies (Spetch et al., 2001; Spetch & Friedman, 2003). First, we found that humans, but not pigeons, showed much less viewpoint dependence when the objects being discriminated each contained a single unique geon than when they contained no distinctive geon or were composed of multiple geons. Geons are 3-D shape primitives whose inherent characteristics make them easy to distinguish from each other from a variety of viewpoints (Biederman, 1987). Thus, our data indicated that humans' recognition performance reflected a qualitative distinction between one-geon and multi-geon objects, but pigeons' performance did not. The results for humans were consistent with previous findings (e.g., Tarr, Bühlhoff, Zabinski, & Blanz, 1997). The results for pigeons suggested that one or more aspects of their recognition processes, such as those used to decompose objects into their components, differed from the processes used by humans.

Alinda Friedman, Marcia L. Spetch, and Anne Ferrey, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to Alinda Friedman and Marcia L. Spetch. Anne Ferrey's honors thesis research generated part of the data, which were reported at the Conference on Comparative Cognition, Melbourne, Florida, March 2003. The 3-D printer used to generate the stimuli was funded by a Multimedia Advanced Computational Infrastructure grant to Robert Lederer, Industrial Design Division, University of Alberta.

Thanks go to Isaac Lank for help developing the apparatus to present real objects; Bernd Kohler for programming the apparatus and experiments and for help with the data analysis; Aaron McGaffey for stimulus generation; and Carleton Fougere, Emily Gray, Angela Nguyen, and Michelle Smith for collecting the data.

Correspondence concerning this article should be addressed to Alinda Friedman, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. E-mail: alinda@ualberta.ca

Our second novel finding was that when people acquired a discrimination between two views of the same three-geon object, presented as pictures, they recognized some of the views they had not seen before as accurately as they recognized the learned views. In contrast, for pigeons, there was a decline in performance for all of the novel views relative to the trained views. The behavioral patterns displayed by humans and pigeons were consistent with the predictions, respectively, of view combination (Bülthoff & Edelman, 1992; Edelman, 1999; Edelman & Bülthoff, 1992; Edelman, Bülthoff, & Bülthoff, 1999) and normalization (Tarr, 1995; Tarr & Pinker, 1989; Ullman, 1989) approaches to object recognition. In the present study, we examined whether the observed species differences would be maintained when stimuli were actual objects.

One of the principal behavioral means of distinguishing view combination from normalization mechanisms involves comparing performance on novel views that are “in between” two learned views with performance on novel views that are “outside” of the shortest distance between the two learned views. These conditions have been referred to as *interpolated* and *extrapolated*, respectively (Bülthoff & Edelman, 1992; Diwadkar & McNamara, 1997; Spetch et al., 2001; Spetch & Friedman, 2003). For example, Figure 1 is a schematic diagram of the top view of the stimulus situation we created. An object at the center of the circle in Figure 1 was learned from the arbitrarily labeled training locations of 0° and 60°. Thus, the novel interpolated test view is 30° from each training view. In contrast, the novel extrapolated test view is 30° from the closer training view and 90° from the farther one. The novel far test view is 60° from the closer training view and 120° from the farther one.

According to normalization accounts of object recognition, speed and accuracy of recognition for both interpolated and extrapolated novel views is predicted to be equal, and inferior, to performance on the learned views (Tarr, 1995; Tarr & Pinker, 1989). This is because it is assumed that objects are represented as a number of “exemplar” views (e.g., the training views) and that novel views are recognized by transforming them until they are aligned with the nearest stored exemplar (e.g., Ullman, 1989). Thus, recognition performance is predicted to be a declining function of distance between a novel view and its nearest stored

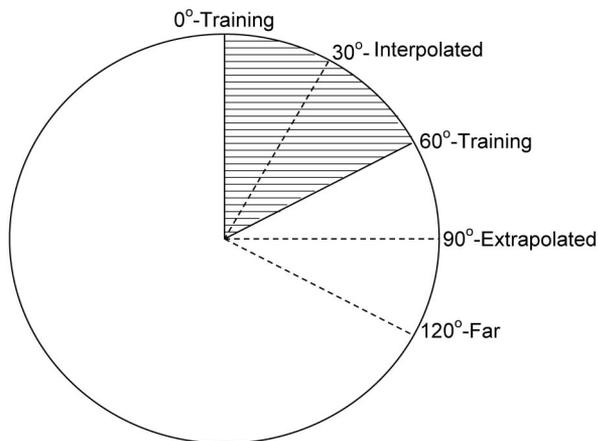


Figure 1. Diagram showing an example of the relationship between two training views and a novel interpolated, extrapolated, and far view.

exemplar, and the interpolated and extrapolated conditions are usually equated on this factor. Indeed, even when the interpolated view is equidistant to two different training views it is still assumed to be normalized to only one of them on each trial.¹

In contrast, view combination approaches permit some novel views to benefit from their similarity to more than one learned view. For example, Edelman (1999; see also Broomhead & Lowe, 1988; Poggio & Edelman, 1990) assumes that objects are represented as points in a multidimensional *shape space* spanned by similarities to a small number of reference objects, which act as prototypes. Recognizing known objects from novel viewpoints occurs by mathematically interpolating between two or more prototypes (Broomhead & Lowe, 1988) to compute what a novel view should look like in that region of the shape space. The predicted view is then matched to the novel view. Thus, when a novel view is relatively close in the shape space to two stored views, as in the interpolated case, the novel view should be similar to the predicted view and therefore relatively easy to recognize. In contrast, in the extrapolated case the more distant learned view will tend to reduce the similarity between the predicted and novel views. This is because similarity (and ease of recognition) is assumed to be an inverse function of the distance between the novel view and each of the stored views that are combined to make the predicted view (Edelman, 1999, p. 128; Shepard, 1968). Together, these assumptions predict that performance should be better on interpolated views than on extrapolated views.

The view combination mechanism just described is similar in many respects to the notion of generalization in the animal learning literature (e.g., Honig & Urcuioli, 1981). In particular, generalization is commonly believed to occur via the combination of excitatory and inhibitory activation gradients that have formed around representations of positive (S+) and negative (S-) stimulus values, respectively (Spence, 1937). When more than one stimulus is positive, as is the case when there are two training views, a positive gradient would be expected to form around the representation of each S+ view. Equally, a negative gradient would be expected to form around the representation of each S- view. Thus, when a novel stimulus view is presented, it causes activation and inhibition according to how similar it is to the S+ and S- views that are represented; the sum of that activation and inhibition determines the response. If the representations at the two training orientations overlap, then novel interpolated views should receive generalization from both training views, whereas novel extrapolated views might receive generalization from only one training view. Thus, the commonly held conceptualization of stimulus generalization makes the same predictions as the view combination approach for the difference in ease of recognition for inter-

¹ There is nothing about exemplar accounts in general that requires only a single exemplar to be retrieved. If the “nearest neighbor” assumption of a normalization approach is relaxed and novel views are allowed to be normalized to either of the training views, then the normalization approach can predict that responses to the interpolated view should be faster and more accurate than responses to the extrapolated view. However, in the context of object recognition models, the normalization approach loses much of its predictive uniqueness if the nearest neighbor assumption is dropped. Further, the view combination mechanism is more robust than normalization because it can predict recognition performance for views of entirely novel objects (Edelman, 1995).

polated versus extrapolated views.² In our previous studies, the pigeons' performance on the interpolated views did not reveal a benefit from their proximity to both of the training views, compared with their performance on extrapolated views that were close to only one training view. Thus, functionally, the training views were responded to as if they were distinctively different shapes. The implication of this was that the generalization gradients for the two training views must not have overlapped.

It should be noted that the difference predicted by the view combination approach regarding performance on interpolated and extrapolated novel views does not reduce to the difference between having test views that are equidistant or nonequidistant to the training views. For example, in Spetch and Friedman (2003), human participants were trained on views that were 90° apart and were tested on views that were either 30° or 45° "inside" or "outside" of the training views. The 30° interpolated views were 30° from one training view and 60° from the other one, whereas the 30° extrapolated views were 30° from the same training view as the interpolated stimulus but were 120° from the second training view. Thus, neither interpolated nor extrapolated test views were equidistant to the training views. However, participants were 23% more accurate and 133 ms faster to respond to the 30° interpolated views than to the 30° extrapolated views. Bülthoff and Edelman (1992) reported similar data over multiple interpolated and extrapolated views that were both equidistant and nonequidistant to the training views. Thus, the key factor influencing recognition performance appears to be the total distance between the novel views and the various training views; it is this distance that determines the similarity between the test view and the predictions based on the training views.

It was not surprising that the humans in our previous studies (Spetch et al., 2001; Spetch & Friedman, 2003) performed well on the interpolated views relative to the extrapolated views; others have observed the same sort of data (e.g., Bülthoff & Edelman, 1992; Bülthoff, Edelman, & Tarr, 1995). However, it is still unknown whether humans' performance would improve on extrapolated views when the stimuli are actual objects. There is some indication that it might. For example, Edelman and Bülthoff (1992) found that providing binocular depth cues to pictures of objects that looked like bent paperclips (i.e., they were made of three to four elongated cylinders connected at their ends and bent at odd angles) significantly lowered the error rate for recognition, though there was still a monotonic relation between reaction time (RT) and degree of rotation away from the learned views. This finding is of interest because of the implications it has for viewpoint-dependent versus invariant recognition. In particular, these results imply that humans' ability to use depth cues with actual 3-D objects might allow their recognition to be more viewpoint-invariant than their recognition of pictures, which places an important constraint on object recognition models and the methods required to test them.

Though it is still controversial whether people need to learn how to interpret line drawings as representations of 3-D objects (DeLoache & Burns, 1994; Derogowski, 1989; Segall, Campbell, & Herskovits, 1966), it seems likely that there are differences between recognizing objects depicted as drawings or photographs and recognizing those same objects in the real world. Even with line drawings, Jolicoeur, Regehr, Smith, and Smith (1985) showed that discrimination performance was worse when unfamiliar ob-

jects were depicted as 3-D objects than when the same shapes were rendered as 2-D objects by rearranging their inner lines; they concluded that interpreting 3-D drawings required either qualitatively or quantitatively different processes. If objects engage different recognition processes than their pictures, it is of considerable interest to determine whether the difference between interpolated and extrapolated views typically found with pictorial stimuli will also occur when humans are viewing actual objects. Again, this is because of the constraints that finding such a stimulus type difference would place on object recognition theories.

The difference between pigeons and humans in their responses to the interpolated views in our previous studies (Spetch et al., 2001; Spetch & Friedman, 2003) suggested that the species differed in their recognition processes. The invariance shown by humans to the pose between the trained views combined with the large decrement to poses outside the training range suggested that they had benefited from having learned two views of the same object by being able to interpolate (generalize) between them. In contrast, finding that pigeons' performance on the interpolated view was inferior to their performance on the trained views implied that they learned how the objects looked at the two training views but did not combine that information to generate a prediction about the views in between; their recognition performance instead likely reflected generalization gradients that were narrowly tuned. We thus concluded that the pigeons had represented the training views as distinctively different shapes. However, if the pigeons did not interpret the images as representing 3-D objects, then the alternative possibility is that pigeons' performance reflected generalization gradients that were based on their perception of the 2-D similarity of the test views to each of the trained views. If so, their results might not reflect an inability to combine information from two views of the same 3-D object. There is some evidence that, for humans, the perceived degree of similarity in the shape space that is derived from 3-D objects is different from, and often greater than, that of their 2-D representations as pictures (Edelman, 1999, p. 187; Watson, Friedman, & McGaffey, 2000). If the same is true for pigeons, and if the pigeons responded to pictures of objects on the basis of 2-D image information rather

² In our previous studies, we made a distinction between interpolation (for humans) and generalization (for pigeons) that we no longer endorse. View combination is, in principle, a kind of generalization, and mathematically interpolating between *radial basis functions* is one method of implementing view combination (Bülthoff & Edelman, 1992; Edelman, 1999). A radial basis function is a kind of neural network in which the responses of neural units are Gaussians; thus, there is a gradual decay in the response of the units as the dissimilarity between the stimulus image and the learned views increases. In general, we believe the important similarities between the view combination approach and stimulus generalization outweigh any implementation differences. For example, the notion that similarity is a function of the inverse distance between a novel view and a set of stored prototypes (or S+ representations) is equally functional in both schemes. Similarly, differently tuned generalization gradients should act in most circumstances like the differently tuned radial basis functions (Edelman, 1999) that underlie viewpoint interpolation. Of importance, both view combination and generalization mechanisms provide a meaningful contrast to recognition by normalization to a nearest neighbor.

than an implied 3-D structure, then they may show better recognition of interpolated views when viewing actual objects.

In several previous studies examining the relation between picture and object perception in birds, the method of comparison involved transfer of training between actual objects and their photographic images. For example, Cabe (1976) trained pigeons to discriminate between two objects that had been painted white and found that the birds transferred the discrimination to photographs and silhouettes but not line drawings. Similarly, Watanabe (1993) showed that birds trained to categorize objects or their pictures as either food or nonfood showed transfer in both directions. However, in most studies of object–picture stimulus transfer, the objects are shown from a single point of view and the pictures display the objects in that same orientation. These kinds of tests do not directly address the issue of whether animals have abstracted the 3-D structure of the objects. For instance, positive transfer could be based on stimulus shape or size as opposed to recognition of the actual 3-D objects (Fagot, Martin-Malivel, & Depy, 1999).

Watanabe (1999) trained pigeons to discriminate between a pair of wooden blocks placed in their cages; thus, the birds could see the objects from a variety of views. They later learned to discriminate video images of the blocks more quickly than images of unfamiliar stimuli. Further, they responded more frequently to new video views of the familiar stimuli than to new video views of the unfamiliar objects. Watanabe (1999) concluded that the familiarization with the actual objects had enhanced viewpoint invariance of their 2-D images. Similarly, Lumsden (1977) found that generalization gradients obtained to an object presented at one of several orientations and to its photographs at the same orientations were virtually identical, although he tested only one bird. In general, however, in an extensive review of the literature on picture–object equivalence in animals, Fagot et al. (1999) concluded that the experimental evidence was often weak and inconsistent. More important, no study has systematically examined pigeons' ability to combine information from two views of the same object when it is presented as an actual object or as a pair of photographs.

In Experiment 1, we trained pigeons and humans to discriminate between pairs of either actual objects or their photographs and tested them on recognition of new views that were either in between and equidistant to both of the training views (i.e., interpolated views) or extrapolated outside of the rotational boundary of the training range. The new interpolated and extrapolated views were each rotated 30° from one of the training views, and which view was interpolated or extrapolated was counterbalanced across participants.

We used both one-geon and three-geon objects to further investigate the difference between species in the effect of distinctive object parts on recognition previously reported by Spetch et al. (2001). In this previous study, we found that humans recognized all of the new views of one-geon objects as accurately as the trained views, whereas pigeons displayed generalization decrements for novel views of the one-geon objects just as they did for the three-geon objects. But this species difference also might have been due to the pigeons' failure to interpret the pictures as reflecting a 3-D structure. If so, pigeons might perform similarly to humans with both one- and three-geon actual objects.

To summarize, we expected that the humans' performance would benefit from their ability to combine information from the two training views for both the actual objects and their pictures; of

interest is whether humans would also perform better on the extrapolated views of the actual objects than on those same views when presented as pictorial stimuli. We also expected that the pigeons would show generalization gradients for views that were interpolated and extrapolated away from the training views of both the one- and three-geon photographs. That is, they would treat the two 2-D training views as distinct shapes. However, it was an open question how pigeons would perform with the actual objects.

Experiment 1

Method

We collected data from four conditions that were run independently (humans vs. pigeons; 3-D objects vs. photographs). The similarities among them are sufficient to permit us to describe the methods together, noting the differences between conditions as necessary.

Human participants. The participants were undergraduate volunteers from the University of Alberta Department of Psychology's pool; they received partial credit toward a course requirement for their participation. All had normal or corrected-to-normal vision.

Participants were trained to a criterion of 80% correct (computed independently over blocks of eight training trials), and their data were kept if their performance on each of the training poses during the test trials remained at 80% and their performance on the other poses was at least 33% correct. Thirty-seven participants (30 women; 7 men) saw actual objects and 5 (2 women; 3 men) failed to meet the criterion; 42 (35 women; 7 men) saw photographs and 10 failed to meet the criterion (7 women; 3 men). This left 32 participants in each stimulus type condition.

Pigeon subjects. Seven adult Silver King pigeons (*Columba livia*) completed testing in the picture condition, and 6 adult Silver King pigeons and 1 Racer pigeon completed testing in the actual object condition. None of the pigeons had previously been exposed to an object discrimination task. They were trained to a criterion of 80% correct for five consecutive sessions. For the picture condition, 4 birds received the three-geon stimuli first and 3 received the one-geon stimuli first. For the actual object condition, 4 adult Silver King or Racer pigeons received the three-geon condition first and 4 received the one-geon condition first. One of the pigeons in the former group failed to reach the criterion within 50 sessions and was dropped from the experiment. In addition, during initial sessions for 4 pigeons, a broken wire resulted in the house lights being off during the trials; the object chambers were visible only during the three brief warning flashes. When this was discovered, we ran these birds again with the normal procedure, and it is those data that are reported below.

The birds were housed in large individual cages under a 12-hr light–dark cycle (with light onset at 6:00 a.m.). All birds were maintained at approximately 85% of their free-feeding weights; they were fed mixed grain during experimental sessions supplemented with Kay Tee pigeon pellets after sessions. Water and grit were available ad libitum in the home cages.

Stimuli and design. The stimulus poses and design were identical for both species in each stimulus condition. Participants were assigned to one of eight presentation conditions. In four of them, participants received training and testing on the one- and then the three-geon stimuli, and in the other four conditions they received the opposite order of stimuli. Each of the four conditions in each order differed by whether Object A or Object B was designated as the S+ or S– and whether the training stimuli were shown in the 0° and 60° views or in the 60° and 120° views.

The 3-D models used in the object condition and photographed for the picture condition were created with a computer-assisted design program (Rhinoceros 1.1; McNeel & Associates, 2001). They were based on the one- and three-geon stimuli used by Spetch et al. (2001). There were two one-geon shapes and two three-geon shapes; for each pair of shapes, one of the objects was arbitrarily designated A, and the other, B (see Figure 2). The one-geon stimuli consisted of a three-part "bracelet" in which the two

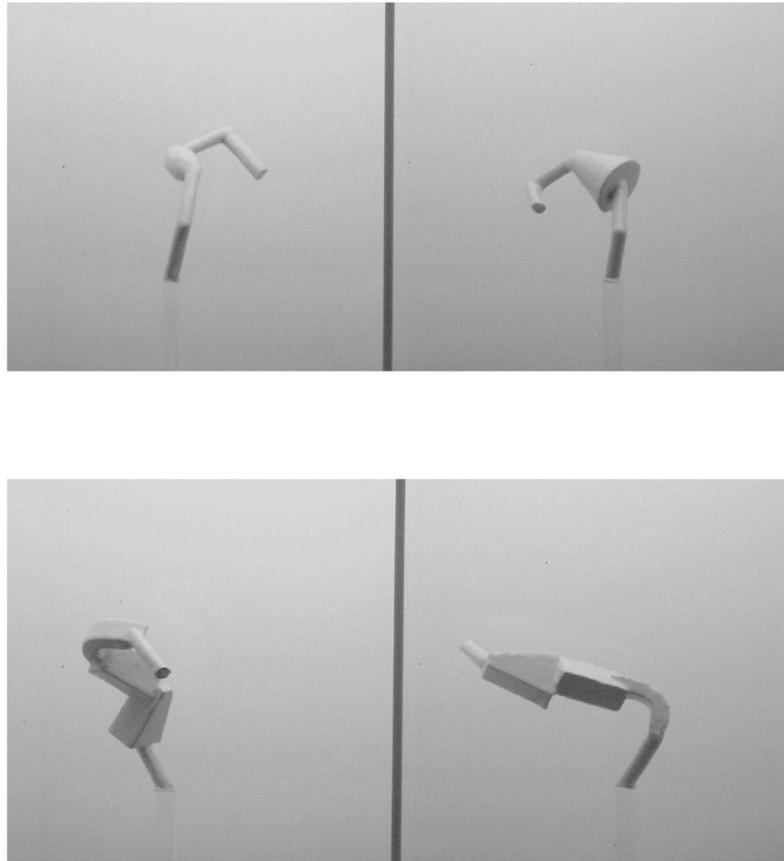


Figure 2. The one- and three-geon real objects used in Experiments 1–3. The figures are pictures taken from the front of the apparatus, through the Plexiglas screen. The top pair have one distinctive geon and are depicted as being 30° from their (arbitrary) start positions. The bottom pair have the same three geons each, but they are in a different order and spatial relation; these stimuli are depicted as being 60° from their start positions. From “An Automated Apparatus for Presenting Depth-Rotated Three-Dimensional Objects for Use in Human and Animal Object Recognition Research,” by A. Friedman, M. L. Spetch, and I. Lank, *Behavioral Research Methods, Instruments, & Computers*, 35, p. 347. Reprinted with permission.

ends for both the A and B figures were elongated cylinders and the A and B shapes were the distinctive center geons. The three-geon stimuli consisted of a five-part bracelet in which, again, the two ends for both the A and B objects were identical elongated cylinders, connected to the same three distinct geons in their centers, but the distinctive geons were in a different order and situated at different angles to each other.

For each A and B pair, we produced one example of the A object and two examples of the B object. During discrimination trials, the A object was placed in the middle chamber of the apparatus and the B objects were placed in chambers to the left and the right of it (see Figure 3). Only two of the three chambers (the left and center or the right and center) were visible to the participants on each trial. This arrangement permitted either the A or the B object to be the S+ or S– and to appear on either the left or the right side of the visible part of the display.

For each of the four stimulus models, we arbitrarily designated a particular pose as the 0° pose. The models were rotated clockwise in 30° increments to create five views for each model: 0° , 30° , 60° , 90° , and 120° . The A and B shapes were always presented in the same pose on each training and test trial (e.g., they might both appear at 60°). The task on any given trial was to respond to the S+ irrespective of its pose.

For the groups trained with the 0° and 60° poses, the 30° , 90° , and 120° poses were the interpolated, extrapolated, and far test conditions, respec-

tively (see Figure 1). For the groups trained with the 60° and 120° poses, the interpolated, extrapolated, and far test views were 90° , 30° , and 0° . Thus, across the two groups, the 30° and 90° views were counterbalanced perfectly in terms of being either the interpolated or extrapolated test pose. As a consequence, any peculiarities about the poses themselves, such as whether they were foreshortened, and so forth, should have canceled out across groups.

For the picture conditions, we used a digital camera mounted on a tripod to photograph what the display looked like in each possible combination of pose, side, and S+ assignment. These images were shown on a computer monitor for both training and test trials. The stimuli were displayed on the monitor with a visual angle matched to that seen by the people who viewed the actual objects. In particular, the vertical visual angle of the displays was approximately 10° , and the horizontal angle was approximately 17° . For both the actual objects and the pictures, the centers of the two objects were approximately 14 cm apart.

Apparatus. The apparatus used to display the actual objects was a large operant animal chest with inside dimensions of 20 in. wide \times 15 in. deep \times 15 in. high, designed to contain a pigeon subject (see Figure 3). Behind the box was a small movable table with the three object chambers. Between the movable table and the box were a sliding door cover and a computer touch screen mounted on a piece of Plexiglas. When the sliding cover was



Figure 3. The front view of the apparatus used to present the object stimuli in Experiments 1–3. Both humans and pigeons saw the same stimulus tableaux; the door to the box was closed for the pigeon subjects and remained open for the human participants. From “An Automated Apparatus for Presenting Depth-Rotated Three-Dimensional Objects for Use in Human and Animal Object Recognition Research,” by A. Friedman, M. L. Spetch, and I. Lank, *Behavioral Research Methods, Instruments, & Computers*, 35, p. 345. Reprinted with permission.

opened, two of the three objects on the movable table were visible through the Plexiglas viewing area. The sliding cover was in the closed position when human participants entered the room and when the pigeons were placed in the box; it prevented them from seeing the starting position of the objects.

For the humans, the door to the operant box was opened during training and test trials. Pigeons were inside the box with the door closed. Participants were tested in a completely darkened room, and the lights within the object chambers were automatically turned off during all table movement and while the objects were rotating, to eliminate visual cues. Timing for each trial coincided with turning on the lights within the chambers. The time spent in darkness between trials was insufficient to allow dark adaptation (i.e., < 10 s).

To prevent humans or pigeons from using auditory cues to anticipate or interpret table movement, the table moved on all trials, whether or not the S+ was programmed to stay on the same side or to shift sides. On “stay”

trials the table moved halfway, paused, and then moved back to the original side. On “shift” trials the table moved halfway, paused, and then continued to move. Thus, both shift and stay trials sounded alike and took approximately the same time between trials. Finally, to prevent auditory cues from signaling the degree of rotation, the objects were rotated on every trial by the shortest amount required to achieve the pose for the current trial, and then by an additional 360°. Further details of the apparatus are described in Friedman, Spetch, and Lank (2003).

For the picture conditions with humans, the operant box and object chambers were replaced by a computer monitor, which displayed the stimuli. For the pigeons, the computer monitor replaced the object chambers at the back of the operant box.

Human participants placed their heads in a chin rest with a forehead stop and wore a pair of earphones to receive the feedback. They responded using a serial response box made by Psychology Software Tools (Pittsburgh, PA). They pressed the button on the far right of the box to select the right-side object as the S+ and the button on the left side to select the left-side object.

Pigeon subjects responded to both types of stimuli by pecking to either the left or the right side of the Plexiglas screen. The screen was surrounded by a Carroll Touch infrared touch frame (Elo Touch Systems, Inc., Menlo Park, CA), which recorded the *x* and *y* coordinates of the pigeons’ pecks. Solenoid-type bird feeders on both sides of the operant box were used to reinforce correct responses. Lamps located within each feeder illuminated feeder presentations, and photocells measured head entries into the hoppers to limit feeding duration to 2 s per food presentation.

General procedure. We presented the training trials in randomized groups of eight (two with the S+ on the left for each of the two training views and two with the S+ on the right for each view). This meant that by the end of training there was an equal amount of experience with each of the training poses on each side of the display (real or pictorial).

On test trials, there were 10 possible stimulus tableaux. Each of the five possible poses for each stimulus could be shown on either the left or the right side, for a total of 10 possible types of trials. Human participants received three blocks of 10 trials, with trial type randomized within blocks. Pigeons received 10 sessions of testing, each with 30 test trials interspersed among 32 training trials.

Procedure for humans. When a participant entered the laboratory, he or she was asked to sit in front of the pigeon box or the computer monitor, depending on the condition. They were instructed that their task was to “discriminate and remember objects,” and that they would see two objects on each trial and were to figure out which was the correct one.

During training trials, participants received feedback over earphones; a prerecorded female voice said “one point” if they were correct, “two points” if they were correct and responded quickly, and “wrong” if they made an incorrect choice. They received the 2-point feedback if they made a correct response in less than 1 s; they were told that 2 points were awarded for speedy responses but were unaware of the exact value that triggered the reward. Participants were told that once they had learned which was the correct object, they should respond as quickly and as accurately as possible. They were also told that partway through the session the feedback message would stop, but they would still earn points. Finally, they were told that halfway through the session, there would be a new set of trials with new objects and that they would again receive feedback for a while, followed by no-feedback trials. They would be informed of their point total at the end of the experiment. After the participant was given the opportunity to ask any questions, the experimenter started the experiment, turned off the room light, and left the room before the first trial began.

For the object condition, on each trial, after the table had moved and the objects had rotated, there was a warning beep for 1 s, followed 500 ms later by the illumination of the objects. When the participant responded, the chamber lights were turned off, and on training trials, the participant immediately received the feedback message over the headphones and the next trial began immediately. On test trials, there was no feedback.

For the picture condition, there was a 1-s warning beep, followed immediately by the presentation of the stimulus display. When the participant responded, the display disappeared, and on training trials, the participant immediately received the feedback message over the headphones. On test trials, there was no feedback. The next trial began after a 4-s delay.

Procedure for pigeons. The procedure was identical for the picture and object tasks. Each pigeon was initially given one or more pretraining sessions to establish reliable eating from the food hopper and pecking at the screen. To establish pecking, a yellow square was intermittently presented on either side of the screen and paired with presentation of a feeder. A peck at the square produced immediate food reinforcement. The bird was moved to the training procedure once it reliably pecked at the yellow square on either side of the screen.

Training sessions lasted for 1 hr each and consisted of repeating blocks of the eight training trials. Training was identical to that with the human participants except for the following: (a) the warning stimulus at the start of each trial consisted of three brief flashes of the overhead house light rather than an auditory beep, (b) choices of the left or right object were made by pecking within an area (5 cm high × 4 cm wide) centered on the object on the left or right side of the touch screen, (c) correct choices were followed by 2 s access to the feeder and incorrect choices were followed by termination of the trial (screen display removed) without a feeder presentation, and (d) there was a 10-s delay before the next trial began.

Upon reaching the 80% accuracy criterion for five consecutive sessions, each bird received five test sessions, followed by additional training sessions (a minimum of five) until the criterion was again reached, and then five more test sessions. Test sessions consisted of a mixture of 32 training trials with reinforcement and 30 test trials without reinforcement. All of the test trials ended without food reinforcement regardless of whether the correct or incorrect object was pecked.

Results

The percentage correct and correct RT data were computed separately for each species, geon condition (one or three), and pose (training, interpolated, extrapolated, and far). For pigeons, we excluded trials in which RTs were over 10 s long (these were 3.5% and 4.3% of the data for the objects and their pictures, respectively). The two measures were analyzed using the same designs, and a $p < .05$ criterion was used for significance. We report partial η^2 as the measure of effect size.

Figure 4 shows the correct RTs from all conditions, and Figure 5 shows the percentage correct data. Overall, humans were faster and more accurate with the one-geon stimuli than with the three-geon stimuli for both photographs and objects. In addition, for the most part they were equally fast and accurate on all of the poses of the one-geon objects. For the three-geon objects, humans' recognition of extrapolated poses took longer and was less accurate than for interpolated poses, which in turn took about the same time to respond to as the training poses. Surprisingly, except for the absence of a main effect of stimulus type, the pigeons' data for both one- and three-geon objects looked remarkably similar to the humans' data for the three-geon objects. However, the interaction between stimulus type and the interpolated–extrapolated difference was larger for pigeons than for humans. These generalizations are corroborated with the analyses below.

We first conducted mixed-design analyses of variance (ANOVAs) in which the between-subjects factors were species and stimulus type (pictures or objects) and the within-subject factors were number of geons (one or three) and pose (training, interpolated, extrapolated, and far). Next, we computed two planned-contrast scores for each subject (interpolated minus train-

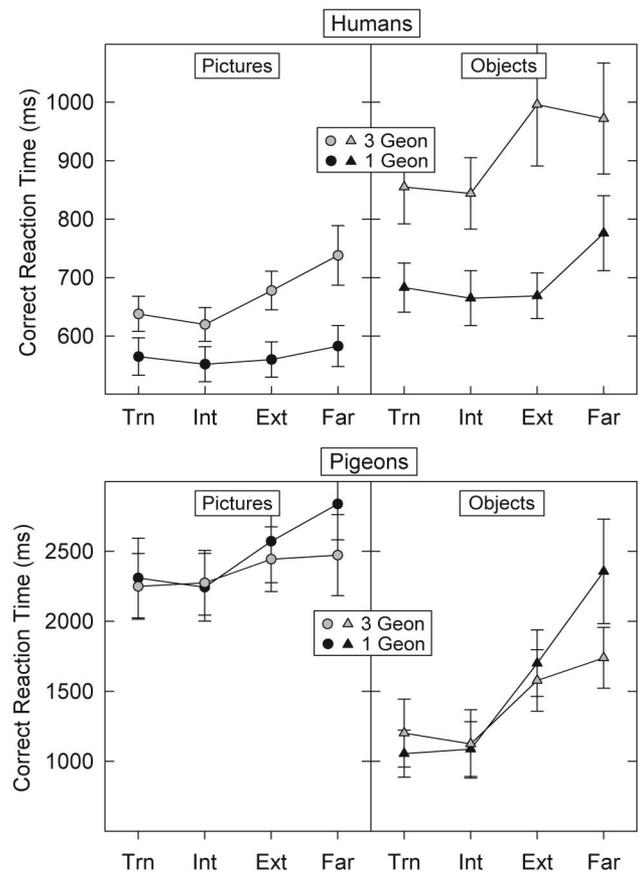


Figure 4. Mean correct reaction times in Experiment 1 as a function of the species, stimulus type, number of geons, and pose condition. Error bars are standard errors of the means, computed separately for each condition. Trn = training views; Int = interpolated view; Ext = extrapolated view; Far = far view.

ing and extrapolated minus interpolated). We analyzed the two contrast scores together in a mixed-design ANOVA with contrast type, number of geons, stimulus type, and species as factors. This analysis thus allowed us to compare the magnitude of the difference between the two contrast scores and all of the other factors (see Masson & Loftus, 2003). We then analyzed each contrast score separately in a species by stimulus type ANOVA. The contrast weights in the separate analyses were the same across both the one- and the three-geon conditions, thus effectively collapsing across conditions. The analysis of the individual contrasts permitted us to examine directly whether there were differences between the species in their responses to the training versus interpolated views, or to the interpolated versus extrapolated views, and whether these differences varied as a function of stimulus type.

RTs. In the omnibus ANOVA, there were main effects of species, stimulus type, and pose, but these were all mitigated by higher order interactions (see Table 1 for all statistics). The species by number of geons interaction was driven primarily by the fact that the difference in RTs to the one- versus three-geon conditions was larger for the humans (161 ms) than for the pigeons (135 ms). The species by geon effect also interacted with pose, and the

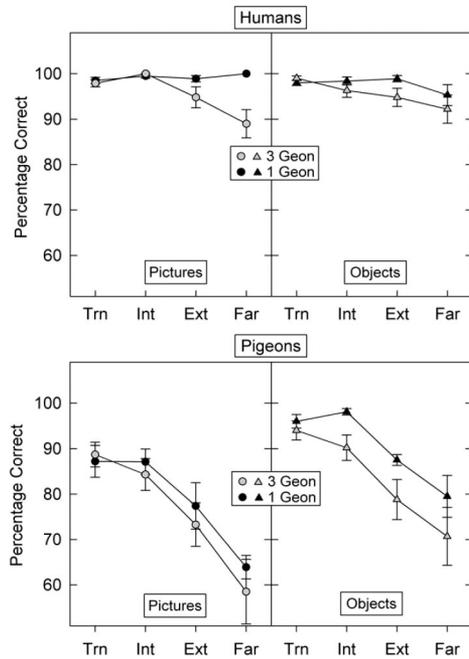


Figure 5. Mean percentage correct in Experiment 1 as a function of the species, stimulus type, number of geons, and pose condition. Error bars are standard errors of the means, computed separately for each condition. Trn = training views; Int = interpolated view; Ext = extrapolated view; Far = far view.

linear component of that interaction was significant, $F(1, 74) = 21.41, MSE = 54,879, \eta_p^2 = .224$. This indicates that the slope of the RT function across poses differed as a function of both the species and the number of geons.

The species by stimulus type interaction showed that whereas humans responded more quickly to the pictures than to the objects (617 vs. 808 ms, respectively), pigeons had the opposite pattern (2,426 vs. 1,481 ms). Most important, the species by stimulus type effect interacted with pose, and the linear component of that triple interaction was also significant, $F(1, 74) = 8.64, MSE = 99,871, \eta_p^2 = .105$. The latter indicates that the slope of the RT function across the different poses changed as a function of both species and stimulus type.

In the ANOVA that used both of the contrast scores, the important new information was in the interactions between contrast type (training vs. interpolated and interpolated vs. extrapolated) and species, $F(1, 74) = 20.45, MSE = 65,559, \eta_p^2 = .216$; contrast type and stimulus type, $F(1, 74) = 4.75, MSE = 65,559, \eta_p^2 = .060$; and the triple interaction between contrast type, species, and number of geons, which approached significance, $F(1, 74) = 2.79, MSE = 63,907, \eta_p^2 = .036, p < .10$. For humans, the pattern of responding between the training and interpolated trials was the same for pictures and objects in both geon conditions. However, the interpolated versus extrapolated difference was larger for the three-geon than for the one-geon condition, and largest overall for the three-geon actual objects. By contrast, for the pigeons, there was no difference between the geon conditions for either the objects or their pictures. Further, the difference between the training and interpolated conditions was negligible, and the difference between the interpolated and extrapolated conditions was larger for objects than for their pictures.

In the ANOVA using only the contrast between training and interpolated trials, none of the effects was significant, $F(1, 74) < 1$ for species, stimulus type, and the species by stimulus type interaction. However, all of the effects in the ANOVA using the contrast between interpolated and extrapolated trials were significant: $F(1, 74) = 31.67, MSE = 163,340, \eta_p^2 = .300$, for the main effect of species; $F(1, 74) = 7.62, MSE = 163,340, \eta_p^2 = .093$, for

Table 1
Results of Omnibus Analyses of Variance for Experiment 1

Source	df	Correct RT			Percentage correct		
		F	MSE	η_p^2	F	MSE	η_p^2
Between							
Species (Sp)	1, 74	234.69*	602,927	.760	160.23*	125.15	.684
Stimulus (St)	1, 74	21.68*	602,927	.227	13.53*	125.15	.155
Sp × St	1, 74	49.17*	602,927	.399	18.41*	125.15	.199
Within							
Geon (G)	1, 74	<1	264,886	.001	14.43*	94.90	.163
Sp × G	1, 74	7.59*	264,886	.093	<1	94.90	.011
St × G	1, 74	<1	264,886	.003	<1	94.90	.005
Sp × St × G	1, 74	<1	264,886	.004	2.07	94.90	.027
View (V)	3, 222	53.19*	57,266	.418	43.95*	86.13	.373
Sp × V	3, 222	30.96*	57,266	.295	20.86*	86.13	.220
St × V	3, 222	8.22*	57,266	.100	<1	86.13	.009
Sp × St × V	3, 222	5.68*	57,266	.071	<1	86.13	.011
G × V	3, 222	7.83*	38,887	.096	3.37*	63.15	.044
Sp × G × V	3, 222	11.32*	38,887	.133	<1	63.15	.005
St × G × V	3, 222	1.85	38,887	.024	<1	63.15	.008
Sp × St × G × V	3, 222	1.09	38,887	.015	<1	63.15	.004

Note. RT = reaction time.
* $p < .05$.

the main effect of stimulus type; and $F(1, 74) = 4.06$, $MSE = 163,340$, $\eta_p^2 = .052$, for the species by stimulus type interaction. In general, the difference between the interpolated and extrapolated RTs was larger for both species when they were responding to objects as opposed to pictures, but the differences as a function of stimulus type were exaggerated more for the birds than for the humans (the mean difference between interpolated and extrapolated trials for pigeons responding to objects vs. pictures, respectively, was 533 ms vs. 249 ms, and for humans the means for these conditions were 78 ms vs. 33 ms). However, when we repeated this analysis using just the contrast scores from the three-geon stimuli, which was the condition in which the two species were the most similar in their responses, the two main effects remained significant but the interaction did not, $F(1, 74) = 1.15$, $MSE = 90,370$, $p = .288$.

Because the pigeons and humans had such widely disparate RTs, we were concerned that some of the findings above might have been attributable to scale or range differences between species. Accordingly, we computed relative latency scores, as we have done previously (Spetch et al., 2001): For each individual, we computed the average correct RT as a function of each geon and pose condition as a percentage of the average of whichever condition had yielded that individual's maximum RT. We then analyzed these scores as well as the contrasts using the same designs as we had with the mean RTs. Except for the absence of a main effect of species (because everything was normalized), the significant effects were the same as reported above.

Percentage correct. The accuracy data mirrored the RT data in most important respects (see Table 1 for all statistics from the omnibus ANOVA), and there did not appear to be any indication of a speed-accuracy trade-off for either species. The humans were nearly at ceiling in all conditions; however, performance can be uniformly accurate across different poses in a discrimination or categorization task (as opposed to an identification task) when the differences between the shapes are relatively easy to discern (Biederman, 1987; Biederman & Gerhardstein, 1993; Edelman, 1999) or are overlearned (e.g., Tarr & Pinker, 1989). In such cases, differences in the ease of recognition across conditions are manifested primarily in RT.

In the ANOVA that examined both contrast scores as a function of species, stimulus type, and number of geons, the difference between training versus interpolated trials and interpolated versus extrapolated trials interacted with species, $F(1, 74) = 13.44$, $MSE = 105.76$, $\eta_p^2 = .154$, and with stimulus type, $F(1, 74) = 5.68$, $MSE = 105.76$, $\eta_p^2 = .071$. Essentially, for the humans, there were very few differences between conditions, but the largest difference in accuracy (5.2%) was between interpolated and extrapolated trials in the three-geon picture condition. For the pigeons, the drop in accuracy between interpolated and extrapolated trials for the one- and three-geon stimuli was approximately the same for the pictures (9.7% and 11.0%, respectively) and the objects (10.6% and 11.4%, respectively).

Of importance, as with the RTs, none of the effects in the analysis of the contrast between training and interpolated trials was significant: $F(1, 74) = 2.16$, $p = .15$, for species; $F(1, 74) < 1$ for stimulus type; and $F(1, 74) = 2.73$, $p = .10$, for the species by stimulus type interaction. However, in the analysis of the contrast between interpolated and extrapolated trials, the difference in contrast scores between species was significant, $F(1, 74) = 16.26$,

$MSE = 228.78$, $\eta_p^2 = .180$. Essentially, there was no difference in accuracy between interpolated and extrapolated trials for the humans (1.7%), but there was a relatively large difference for the pigeons (10.7%).

In sum, for both species, performance on novel views of pictures and actual objects that were equidistant to the two training views was better than performance on novel views that were at unequal distances to the training views. Both humans and pigeons also had bigger differences between the interpolated and extrapolated novel views of the actual objects than they did between those conditions for the pictures; for humans, this effect was limited to the three-geon objects, but for pigeons it was true for both one- and three-geon shapes. Nevertheless, the pigeons were clearly faster and more accurate in their responses to the actual objects than they were to their images. Indeed, they responded faster to the extrapolated views of actual objects than to the training views of their images. In contrast, the humans actually responded more slowly to the objects than to their images, and there was no evidence that the actual objects conferred any advantage to extrapolated novel views, relative to their images.

Discussion

That pigeons performed faster and more accurately to virtually all views of actual objects than to photographs of the very same views is a new finding. That pigeons responded faster and more accurately to novel interpolated views than they did to novel extrapolated views for both objects and their images is also new, as is the fact that the interpolated-extrapolated difference was much larger for responses to the actual objects than to their images for both species.

The fact that the birds in the present study responded more quickly and accurately to pictures of novel views that were interpolated between the two training views than they did to the novel extrapolated views is a finding that differs from our previous results with pigeons using picture stimuli (Spetch & Friedman, 2003). However, in the present experiment the training poses were 60° apart, and the interpolated and extrapolated views were rotated by 30° from the nearest training view. In our previous research with pictures (Spetch & Friedman, 2003), the training views were 90° apart, and the novel test views were 45° from the nearest training view. In that study, recognition speed and accuracy for the birds was manifested by a graceful generalization gradient away from the training views for both interpolated and extrapolated poses.

The cause of the difference between our previous and present findings for picture stimuli is difficult to pinpoint because of the studies' differences in distance between the training views. For example, in the present case it is possible that the 30° poses were sufficiently "close" to the training poses that it was irrelevant whether the pigeons represented the training poses as separate objects or as two views of the same object: The excitatory activation (or tuning of the radial basis functions; Edelman, 1999) would have overlapped with that of the novel interpolated view in either case. But the overlap in representations can still occur for different reasons with actual objects and pictures. That is, it is possible that the pigeons did represent the two training views in the picture condition as different shapes, rather than as two views of the same shape. If so, in principle, we should have observed poorer perfor-

mance to the interpolated views relative to the training views. But even different shapes can have a similarity function; indeed, the interpolation implementation of view combination can predict how well novel shapes, not just novel views of the same shape, should be recognized (Edelman, 1995, 1999). Thus, it is possible that we would obtain performance differences for pigeons as a function of stimulus type if the training stimuli were farther apart.

To address this issue, we used a new set of birds in Experiment 2 to replicate the picture and object conditions of Experiment 1, but we used only the three-geon stimuli and training views that were 90° apart. The novel interpolated view was 45° from both training views, and the extrapolated view was 45° from the nearest training view and 135° from the farther one. These were the training and test rotations we used with previous picture stimuli when we obtained no evidence that pigeons showed a speed or accuracy benefit in their responses to interpolated stimuli, relative to extrapolated stimuli (Spetch & Friedman, 2003). Therefore, we expected that for the picture stimuli, the speed and accuracy of responding to the novel interpolated and extrapolated views would be roughly equivalent and would be poorer than the RTs and accuracy for the training trials. However, if pigeons process actual objects and their images differently, then the data for the objects in Experiment 2 might replicate those of Experiment 1 and show a difference between interpolated and extrapolated views.

Experiment 2

Method

Stimuli. We used the same objects as in Experiment 1 but selected a different start position, and the different poses were made in 45° increments from there. We used only the three-geon objects. The design was similar to that of Experiment 1, except that we tested only pigeons and the poses for the two counterbalanced groups were 0°, 45°, 90°, 135°, and 180°. One group within each stimulus type condition was trained on the 0° and 90° poses, so that the 45° pose was interpolated, the 135° pose was extrapolated, and 180° was the far pose. The second group was trained on the 90° and 180° poses, so that the 135° pose was interpolated, the 45° pose was extrapolated, and 0° was the far pose. We used a digital camera mounted on a tripod to photograph the display in each possible combination of pose, side, and S+ assignment.

Apparatus and procedure. The apparatus and procedure for both stimulus conditions were the same as those used for the pigeons in Experiment 1.

Subjects. Fourteen experimentally naive Silver King pigeons (*Columba livia*) were assigned in random order to the object ($n = 8$) and picture ($n = 6$) groups. Two of the pigeons in each group failed to reach the accuracy criterion and were dropped from the experiment. Of the remaining pigeons in each stimulus type group, half were trained with the 0° and 90° poses ($n = 3$ for the object group and $n = 2$ for the picture group), and the remainder were trained with the 90° and 180° poses.

Results

The percentage correct and correct RT data were computed for each pose (training, interpolated, extrapolated, and far) and stimulus type (objects or pictures). RTs greater than 10 s were eliminated (4.3% and 7.3% of the data for the pictures and objects, respectively). The measures were analyzed in a mixed design with stimulus type between subjects and pose within subjects. We also computed two contrast scores for each subject: One was for the difference between the training and interpolated conditions, and

the second was for the difference between the interpolated and extrapolated conditions.

The ANOVA on correct RTs yielded a main effect of pose, $F(3, 24) = 3.33$, $MSE = 398.65$, $\eta_p^2 = .294$, and the cubic component of the stimulus type by pose interaction approached significance, $F(1, 8) = 3.91$, $MSE = 163,772$, $\eta_p^2 = .328$. This indicates that the form of the latency function differed as a function of stimulus type. Further, in the analysis of the contrast scores, the contrast by stimulus type interaction also approached significance, $F(1, 8) = 3.90$, $MSE = 341,122$, $\eta_p^2 = .328$. The difference between the training and interpolated conditions was only 43 ms for the birds that learned the actual objects, but it was 456 ms for the birds that learned the photographs. In contrast, the difference between the interpolated and extrapolated conditions was 704 ms for the birds responding to the actual objects but was only 63 ms for the birds responding to the pictures. In other words, as seen in Figure 6, for picture stimuli, performance was fastest to the training poses, intermediate and equivalent for the interpolated and extrapolated poses, and slowest to the far pose. In contrast, for the actual object stimuli, performance was fast and approximately equivalent for the training and interpolated poses and was slower and roughly equivalent for the extrapolated and far poses. It is also clear that, as in Experiment 1 in the present study, the birds responded more quickly to the interpolated object stimuli than they did to their images. They also responded more slowly to the pictures of the far novel poses than they did when the objects were real.

For comparability to Experiment 1 and our earlier research (Spetch et al., 2001), we also computed and analyzed relative latency scores (i.e., the mean correct RT for each condition was expressed as a percentage of each bird's maximum RT). These proved to be less variable than the mean RTs, so that the cubic

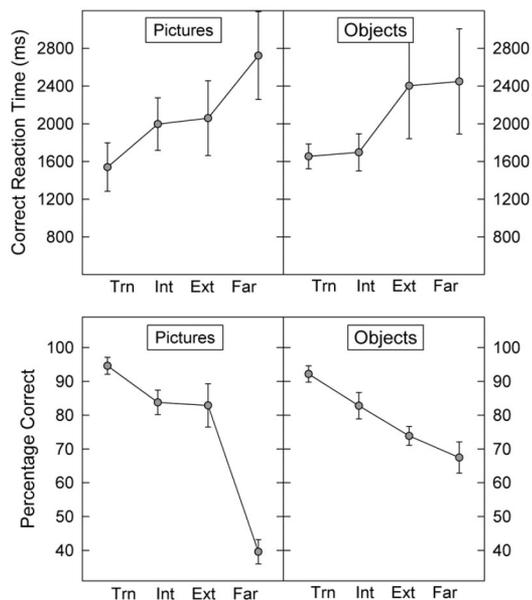


Figure 6. Mean correct reaction times and mean percentage correct for Experiment 2 as a function of stimulus type and pose condition. Error bars are standard errors of the means, computed separately for each condition. Trn = training views; Int = interpolated view; Ext = extrapolated view; Far = far view.

effect of the view by stimulus type interaction, $F(1, 8) = 5.50$, $MSE = 110.06$, $\eta_p^2 = .408$, and the contrast type by stimulus type interaction, $F(1, 8) = 5.30$, $MSE = 299.57$, $\eta_p^2 = .399$, were both statistically significant.

As in Experiment 1, the accuracy data mimicked the RT data in most important respects (see Figure 6). It is clear that for the picture stimuli, the highest accuracy was to the training stimuli, accuracy dropped by the same amount for both the interpolated and extrapolated poses, and it made a further drop for the far pose. This is exactly the pattern we observed in our previous studies with picture stimuli (Spetch et al., 2001; Spetch & Friedman, 2003). In contrast, for the actual objects there was a smoother decline in performance.

In line with this description, the main effect of pose was significant, $F(3, 24) = 41.00$, $MSE = 67.12$, $\eta_p^2 = .837$, as was its linear component, $F(1, 8) = 69.05$, $MSE = 107.62$, $\eta_p^2 = .896$. Of importance, the pose by stimulus type interaction was also significant, $F(3, 24) = 9.63$, $MSE = 67.12$, $\eta_p^2 = .546$, as were its linear, quadratic, and cubic components: $F(1, 8) = 7.65$, $MSE = 107.62$, $\eta_p^2 = .489$; $F(1, 8) = 20.70$, $MSE = 36.67$, $\eta_p^2 = .721$; and $F(1, 8) = 6.23$, $MSE = 57.07$, $\eta_p^2 = .438$, respectively. All three trends indicate that the form of the accuracy functions differed as a function of stimulus type. However, there was too much variability for the analysis of the contrast scores to reflect the interaction. Nevertheless, the means were in the right direction: For the picture stimuli, the difference in accuracy between the training and interpolated conditions was -10.8% , whereas it was only -0.9% for the interpolated and extrapolated conditions. For the actual objects, the difference between the training and interpolated conditions was -9.4% , and between the interpolated and extrapolated conditions it was -8.9% .

In sum, the accuracy data for the picture stimuli exactly replicated our previous work; there was no evidence of any benefit for the novel interpolated stimulus compared with the novel extrapolated stimulus. However, the accuracy data from the actual objects were a little ambiguous in that the function had a different form than it did in Experiment 1; performance declined gracefully from the training condition to the far pose.

The overall pattern of findings in Experiment 2 warrants two conclusions: First, the birds that received picture stimuli responded fastest and most accurately to the training views and more slowly and less accurately, by equal amounts on each measure, to both the interpolated and extrapolated views. Their worst performance was on the far views. This finding replicates exactly our previous findings with picture stimuli in which the training views were 90° apart (Spetch & Friedman, 2003). Second, however, the birds that received the actual objects were both fast and relatively accurate on the interpolated view, compared with their performance on the extrapolated view. Thus, in contrast to Experiment 1, in Experiment 2 the birds responded differently to actual objects than they did to their images in the comparison between interpolated and extrapolated views.

Discussion

Across two experiments, pigeon subjects that learned two views of actual 3-D objects responded to a novel stimulus that was interpolated within the shortest distance between the training stimuli about as quickly and accurately as they did to the training poses

themselves. In contrast, they were slower and less accurate on novel views that were outside of the training range by an amount equivalent to the interpolated views. This pattern held for both one- and three-geon actual objects (Experiment 1) and for novel views that were either 30° (Experiment 1) or 45° (Experiment 2) away from the nearest training view. However, for pictures, the advantage in performance for novel interpolated views held only when they were 30° from the training views (Experiment 1).

Experiment 3

The goal of Experiment 3 was to rule out one possible cause of the species differences we had observed in Experiment 1. In particular, we had used a chin rest for the human participants in Experiment 1 to maintain a visual angle in the object condition that was constant and that matched the visual angle of the picture condition. However, we thereby introduced a disparity between the humans and the birds: The birds were in a free-viewing situation, but the humans were not. It is possible that the pigeons' head movements could have produced different visual outcomes in the picture and object conditions of both Experiments 1 and 2. In particular, in addition to motion-parallax cues that are meaningful for the actual objects, the pigeons might have glimpsed more of their structure during the training trials. If so, this additional information could have contributed to their ability to recognize the interpolated views and might have provided at least part of the advantage for the actual objects in Experiment 2. It is possible that humans might also benefit from a free-viewing situation. For example, if they could glimpse more of the objects, they might perform more quickly and accurately on the novel extrapolated views, and so they might have less of a difference between the novel interpolated and extrapolated views. We tested this possibility in Experiment 3, using only the three-geon actual object displays from Experiment 2, in which the training stimuli were at 0° and 90° and the interpolated, extrapolated, and far views were at 45° , 135° , and 180° , respectively.

We tested two groups of humans. One group used a chin rest to minimize their head movements, as in Experiment 1, and the other did not. If glimpsing more of the objects is beneficial, then the main expectation would be that the group without the chin rest would have faster and more accurate performance than the group who used the chin rest. In addition, it was possible that the group in the free-viewing situation would have faster and more accurate responses to the novel extrapolated views, and perhaps even to the far views, than the group in the chin-rest condition.

Method

The stimuli were identical to those used for the pigeons in the object condition of Experiment 2. The procedure and all other aspects of the experiment were identical to those used for the humans in the object condition of Experiment 1. The only difference between the two groups was that one was instructed to use the chin rest, as in Experiment 1, and for the other group, the chin rest was removed from in front of the apparatus.

The participants were 79 undergraduate volunteers (38 women and 41 men) from the same pool as Experiment 1. They were randomly assigned to one of the two groups. Thirty-nine (25 women and 14 men) were assigned to the free-viewing condition, of whom 7 (5 women and 2 men) did not meet the 80% accuracy criterion. Forty (21 women and 19 men)

were assigned to the chin rest condition, of whom 8 (5 women and 3 men) did not meet the criterion. This left 32 participants in each condition.

Results

Figure 7 shows the mean correct RTs for each condition, as well as the percentage correct. For RT, the main effect of view was significant, $F(3, 186) = 9.48$, $MSE = 52,491$, $\eta_p^2 = .133$, but it did not interact with experimental group, $F(3, 186) < 1$. The contrast between the training and interpolated trials approached significance, $F(1, 62) = 3.55$, $MSE = 59,374$, $\eta_p^2 = .054$; both groups were somewhat faster to respond to the interpolated stimuli than they were to respond to the training stimuli. The participants were also faster to respond to interpolated than to extrapolated stimuli, $F(1, 62) = 13.45$, $MSE = 65,391$, $\eta_p^2 = .178$, and the difference in RTs to the extrapolated stimuli and far stimuli approached significance, $F(1, 62) = 3.89$, $MSE = 130,809$, $\eta_p^2 = .059$. Of importance, none of the contrasts differed between the groups; $F < 1$ in all cases. An analysis of relative latencies showed that the groups were truly indistinguishable: The mean percentage of maximum RT for the group with the chin rest was 82.2%, 79.3%, 88.0%, and 94.3% for the training, interpolated, extrapolated, and far conditions, respectively, and for the free-viewing group the means were 87.1%, 82.2%, 87.9%, and 94.7%.

The findings for percentage correct were similar in all respects to those for correct RTs. There was a main effect of view, $F(3, 186) = 4.82$, $MSE = 70.98$, $\eta_p^2 = .072$, which did not interact with group, $F(3, 186) < 1$. Performance was very accurate, and none of the contrasts were significant, although the contrast between the extrapolated and far conditions approached significance, $F(1, 62) = 3.34$, $MSE = 187.19$, $\eta_p^2 = .051$. Nevertheless, the impor-

tant point is that accuracy in the four conditions was virtually identical across the two groups.

Discussion

There was no evidence that allowing humans to view the stimuli in an unconstrained manner improved their performance. Though performance was very accurate, there was room for it to improve in both the extrapolated and far conditions. Further, the free-viewing group was slightly slower in their RTs than the group that used the chin rest (although this effect was not significant); thus, allowing free viewing did not improve the overall RT to respond to the actual objects, relative to the constrained viewing condition or to the picture condition of Experiment 1. Of course, it is still certainly possible that pigeons' performance on actual objects in Experiments 1 and 2 benefited from free viewing, relative to the picture conditions. However, we think it is unlikely that the use of the chin rest in Experiment 1 was responsible for the observed species differences. This is not surprising, in view of the fact that Bülthoff and Edelman (1992) intentionally gave their human participants more information about the structure of their pictured target objects by rotating them through $\pm 15^\circ$ prior to the start of the training trials; their training stimuli were only 75° apart. They still found large differences in performance between the interpolated and extrapolated conditions.

General Discussion

The new findings in the present study can be summarized as follows. First, training with actual objects did not help the pigeons take advantage of the presence of a distinctive geon. In general, they responded in the same way to one-geon stimuli that they did to three-geon stimuli in Experiment 1, regardless of stimulus type. This finding contrasts with the human data. The difference humans showed in their responses to interpolated and extrapolated views of both objects and pictures characterized their performance with the three-geon stimuli but not the one-geon stimuli. The results for both species replicate previous findings from studies that have used only pictures of objects (Spetch et al., 2001; Tarr & Bülthoff, 1998). Object recognition in humans clearly benefits from the presence of a single diagnostic component, but object recognition in pigeons does not. This suggests that there may be a fundamental difference between pigeons and humans in the processes used to decompose objects into components.

Second, pigeons displayed differences in speed and accuracy of recognition as a function of stimulus type when the training stimuli were 90° apart (Experiment 2) but not when the same shapes were trained at 60° apart (Experiment 1). When the training stimuli were 90° apart, RTs to novel interpolated views of actual objects were faster than to novel extrapolated views, but RTs to novel interpolated and extrapolated views of pictures were equivalent and were slower than the times to the training views. In contrast, when the training stimuli were 60° apart, RTs (and accuracy) for the interpolated views benefited relative to the extrapolated views for both objects and pictures. Third, even in Experiment 1, in which there were no stimulus type differences in terms of the observed advantage for interpolated views relative to extrapolated views, the pigeons still responded faster and more accurately overall to the actual objects than they did to their images.

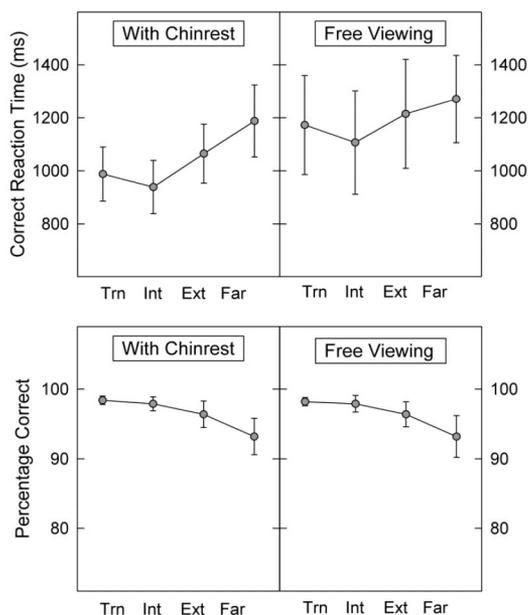


Figure 7. Mean correct reaction times and mean percentage correct for Experiment 3 as a function of viewing condition and pose. Error bars are standard errors of the means, computed separately for each condition. Trn = training views; Int = interpolated view; Ext = extrapolated view; Far = far view.

Thus, the key finding across Experiments 1 and 2 is that the pigeons' recognition performance to novel interpolated views of actual objects was different from their performance with the photographs of those views, but only when the training views were relatively far apart. There are several implications of this finding, and they are not all mutually exclusive. First, the process of generalizing from independent representations at each training view would be expected to produce superior recognition of novel interpolated views only if there was overlap between the two generalization functions. Because the view combination approach is a form of generalization, the same implication holds: Superior recognition of a novel view would occur only if the two radial basis functions from which a predicted view was formed were sufficiently close to each other, so that the novel and predicted views were relatively similar. Thus, because successful generalization (or view combination) appeared to occur in the 90° training condition for actual objects but not for their images, this implies that the representations that resulted from seeing actual objects were more broadly tuned than were the representations that resulted from seeing photographs, at least for pigeons.

A second, related possibility is that in both the 60° and 90° training conditions, the pigeons represented the two training views as distinctly different objects when they were presented as pictures but as the same object when they were real. This could be because the pictures required interpretative processes that pigeons do not have. If two views are determined to be views of the same object, that may contribute to the breadth of their representations or to an increase in the connections between representations and, hence, to the ability to generalize to representations "in between."

The third possibility is that actual objects may allow more of their structure to be inferred more readily than pictures, either because of binocular or motion-parallax cues or because of other, more subtle differences between them. Related to this is the possibility that the pigeons saw more of the structure of the actual objects than they did of the pictures in Experiments 1 and 2. If so, this could have helped to create more broadly tuned representations, which in turn was beneficial to performance on the novel interpolated views, especially when the training views were relatively far apart. However, it is notable that the performance of humans who were permitted to see more of the 3-D structure of an object when the training views were 75° apart (Bülthoff & Edelman, 1992) did not necessarily benefit from this.

The first three possible explanations of the data all presuppose that the difference in performance between pictures and objects observed in Experiment 2 is a result of differences in the similarity of, or overlap between, either the generalization gradients or representations of each of the training stimuli and the novel views to be recognized. A fourth possibility is that rather than generalizing between the two representations of the training stimuli, the pigeons normalized from the novel interpolated picture to one of the two training stimuli in the picture condition of Experiment 2, but not in the object condition of that experiment or in either stimulus type condition of Experiment 1. We think this alternative is the least parsimonious and the least likely to be correct. Indeed, we think our data may imply that normalization is not usually the means by which familiar objects in novel views are recognized. Rather, the data provide evidence for the interpretation that both pigeons and humans can combine information from two trained

views of the same object, and from their pictures, provided that there is overlap in the tuning functions of the represented stimuli.

Tarr and Pinker (1989) and Tarr (1995) obtained evidence for what they claimed was a normalization process after training on multiple views of, respectively, pictures of 2-D letterlike stimuli and pictures of 3-D stimuli made from cubes. However, we think that instead, they may have set up a learning situation that encouraged the formation of very narrowly tuned representations in order to perform correctly. For example, in Tarr and Pinker (1989), participants learned to discriminate three very similar 2-D characters presented in both standard and mirror-image versions, in each of four orientations. They had a total of 12 practice blocks of 192 trials each with these stimuli before being tested on "surprise" stimuli at new orientations. A similar design was used by Tarr (1995) with pictures of 3-D objects. In both cases, the difficulty of the discrimination as well as the degree of overlearning afforded by the practice blocks may have contributed to the formation of multiple representations of each object that were very narrowly tuned. If so, it is not surprising that it was difficult to use information from any two of the views to facilitate performance on any of the novel views. That is, if participants must distinguish among very similar stimuli, it will be to their advantage to have narrowly tuned representations. When representations are narrowly tuned, they are functionally less similar to each other, so that information from the different views cannot be effectively combined to facilitate recognition.

On the basis of the data from Experiments 1 and 2, we conclude that under certain circumstances, pigeons can successfully combine information from two pictured views of an object. However, when the two views were far apart, pigeons combined the information for actual objects but not for their pictures. It is therefore possible that the processes used for creating representations in the two domains differ or, at a minimum, that the representations themselves differ. They may differ quantitatively, in terms of the breadth of their tuning, or they may differ qualitatively, in terms of the kinds of features that are represented. An implication of this conclusion is that humans, too, may use different processes to encode and recognize actual objects and their photographs but that they are so adept at processing photographs by the time they are adults that our task and/or measures are insensitive to the differences.

At this stage we can only conjecture about the kinds of cues that actual objects afford but that are absent in their photographs. But in the human literature it is beginning to become evident that there are differences in the behavior observed when conditions are closer to the real-world environment than to static photographs (Waller, Beall, & Loomis, 2004). It is likely that such behavioral differences would also be observed in other species; indeed, they might even be magnified owing to a lack of experience with pictures. The present data underline the importance of using actual objects as stimuli in assessing and modeling human and animal object recognition.

References

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint in-

- variance. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1506–1514.
- Broomhead, D. S., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321–355.
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, USA*, 89, 60–64.
- Bülthoff, H. H., Edelman, S., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, 3, 247–260.
- Cabe, P. A. (1976). Transfer of discrimination from solid objects to pictures by pigeons: A test of theoretical models of pictorial perception. *Perception & Psychophysics*, 19, 545–550.
- DeLoache, J., & Burns, N. (1994). Early understanding of the representational function of pictures. *Cognition*, 52, 83–110.
- Deregowski, J. B. (1989). Real space and represented space: Cross-cultural perspectives. *Behavioral and Brain Sciences*, 12, 51–119.
- Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, 8, 302–307.
- Edelman, S. (1995). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics*, 72, 207–220.
- 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.
- Edelman, S., Bülthoff, H. H., & Bülthoff, I. (1999). Effects of parametric manipulation of inter-stimulus similarity on 3D object categorization. *Spatial Vision*, 12, 107–123.
- Fagot, J. (Ed.). (2000). *Picture perception in animals*. Philadelphia: Psychology Press/Taylor & Francis.
- Fagot, J., Martin-Malivel, J., & Depy, D. (1999). What is the evidence for an equivalence between objects and pictures in birds and nonhuman primates? *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 18, 923–949.
- Friedman, A., Spetch, M. L., & Lank, I. (2003). An automated apparatus for presenting depth-rotated three-dimensional objects for use in human and animal object recognition research. *Behavior Research Methods, Instruments, & Computers*, 35, 343–349.
- Honig, W. K., & Urciuoli, P. J. (1981). The legacy of Guttman and Kalish (1956): 25 years of research on stimulus generalization. *Journal of the Experimental Analysis of Behavior*, 36, 405–445.
- Jolicoeur, P., Regehr, S., Smith, L. B. J. P., & Smith, G. N. (1985). Mental rotation of representations of two-dimensional and three-dimensional objects. *Canadian Journal of Psychology*, 39, 100–129.
- Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, 4, 401–414.
- Lumsden, E. A. (1977). Generalization of an operant response to photographs and drawings/silhouettes of a three-dimensional object at various orientations. *Bulletin of the Psychonomic Society*, 10, 405–407.
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203–220.
- McNeel, R., & Associates. (2001). *Rhinoceros 1.1*. [Computer software]. Seattle, WA: Author.
- Nicholson, J. R., & Seddon, G. M. (1977). The influence of secondary depth cues on the understanding by Nigerian schoolboys of spatial relationships in pictures. *British Journal of Psychology*, 68, 327–333.
- Peissig, J. J., Wasserman, E. A., Young, M. E., & Biederman, I. (2002). Learning an object from multiple views enhances its recognition in an orthogonal rotational axis in pigeons. *Vision Research*, 42, 2051–2062.
- Poggio, T., & Edelman, S. (1990, January 18). A network that learns to recognize three-dimensional objects. *Nature*, 343, 263–267.
- Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception*. Indianapolis, IN: Bobbs-Merrill.
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly spaced data. *Proceedings of the 23rd National Conference of the ACM*, 517–524.
- Spence, K. W. (1937). The differential response of animals to stimuli differing within a single dimension. *Psychological Review*, 44, 430–444.
- Spetch, M. L., & Friedman, A. (2003). Recognizing rotated views of objects: Interpolation versus generalization by humans and pigeons. *Psychonomic Bulletin & Review*, 10, 135–140.
- Spetch, M. L., Friedman, A., & Reid, S. L. (2001). The effect of distinctive parts on recognition of depth-rotated objects by pigeons (*Columba livia*) and humans. *Journal of Experimental Psychology: General*, 130, 238–255.
- Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, 2, 55–82.
- 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. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32, 193–254.
- Waller, D., Beall, A. C., & Loomis, J. M. (2004). Using virtual environments to assess directional knowledge. *Journal of Environmental Psychology*, 24, 105–116.
- 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 stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 205–221.
- Watanabe, S. (1993). Object-picture equivalence in the pigeon: An analysis with natural concept and pseudoconcept discriminations. *Behavioral Processes*, 30, 225–231.
- Watanabe, S. (1999). Enhancement of viewpoint invariance by experience in pigeons. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 18, 321–335.
- Watson, B., Friedman, A., & McGaffey, A. (2000, April). Using naming time to evaluate quality predictors for model simplification. *Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems*, 113–120.

Received August 13, 2003

Revision received December 15, 2004

Accepted December 28, 2004 ■