View combination in moving objects: The role of motion in discriminating between novel views of similar and distinctive objects by humans and pigeons

Alinda Friedman [a,*], Quoc C. Vuong [b], Marcia L. Spetch [a]

[a] Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9
[b] Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, United Kingdom, NE2 4HH

ABSTRACT

Humans and pigeons were trained to discriminate between views of similar and distinctive objects that rotated in depth coherently or non-coherently. We tested novel views that were either moving or static and were either between the training viewpoints or beyond them. With both types of motion, both species recognized views between the training viewpoints better than views beyond this range. Additionally, for humans, and to some extent for pigeons, when similar objects were learned via coherent motion, dynamic cues facilitated recognition of viewpoints predictable from the direction of motion. Overall, the results suggest that dynamic information may be added to object representations for both species.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Humans and pigeons are both highly visual species, but they differ in their visual experiences and in the morphology of their visual system (see Husband & Shimizu, 2001; Ziegler & Bischof, 1993). For example, because of rapid flight, birds may have different processes for rapid comprehension of the three dimensional (3D) world than do humans. In addition, anatomically, pigeons have two fovea-like retinal areas. One appears to be specialized for near frontal vision, and is the one we presume is engaged in the present study; the other area appears to underlie more distant monocular lateral vision (see Blough, 2001).

To interact with a dynamic environment, both humans and pigeons need to encode relevant features of objects to discriminate among them under novel viewing conditions. Thus, despite many obvious differences between species, there are also striking similarities between their visual systems and the information those systems must process. In our previous work (Friedman, Spetch, & Ferrey, 2005; Spetch & Friedman, 2003; Spetch, Friedman, & Reid, 2001; Spetch, Friedman, & Vuong, 2006; Vuong & Tarr, 2004, 2006), we have shown that humans and pigeons alike are remarkably good at adapting to the stimulus information that is available to learn and subsequently discriminate among objects. For example, both species seem to use static (shape) and dynamic (motion) cues for recognition. This implies that both species have dedicated

neural mechanisms to process and ultimately integrate both cues. For the primate visual system, Giese and Poggio (2003) have proposed that shape and motion cues may be processed in parallel pathways, each along a hierarchy of increasingly complex features that takes into account global shape and movement patterns (see also Ungerleider & Mishkin, 1982). A similar division of function has also been proposed for the avian brain (e.g., Nguyen et al., 2004; see Dittrich & Lea, 2001, for a review of motion discrimination and recognition in avian vision). These system similarities lead us to expect similar performance in humans and pigeons when they are discriminating between dynamic objects.

Whereas the role of static cues in object recognition has received a great deal of attention in both the human and non-human animal literature (e.g., Biederman, 1987; Biederman & Gerhardstein, 1993; Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Friedman, Spetch, & Ferrey, 2005; Logothetis & Pauls, 1995; Peissig, Wasserman, Young, & Biederman, 2002; Spetch & Friedman, 2003; Tarr, Bülthoff, Zabinski, & Blanz, 1997; Spetch, Friedman, & Reid, 2001), less attention has been paid to the role of dynamic cues in recognition, particularly with respect to comparisons across species (but see Cook & Roberts, 2007; Cook, Shaw, & Blaisdell, 2001; Loidolt, Aust, Steurer, Troje, & Huber, 2006; Spetch et al., 2006). Thus, in the present study, we investigate the usefulness of both static and dynamic cues when humans and pigeons attempt to recognize novel views of previously learned objects that are either structurally similar or distinctive. In doing so, we augment an important class of view combination theories of object recognition (Edelman, 1999; Poggio & Edelman,

* Corresponding author. Fax: +1 780 492 1768.
E-mail address: alinda@ualberta.ca (A. Friedman).

1990; Ullman, 1998) that is based on static information alone to situations in which objects are moving. This is the first attempt we are aware of to investigate how humans and pigeons generalize dynamically-learned objects to novel views of those objects (both dynamic and static), in the context of the view combination predictions.

1.1. View combination models and static cues

Empirical studies in the static domain have shown that both humans and pigeons recognize some novel views as well as they recognize familiar, learned views, whereas recognition of other novel views is less efficient (Bülthoff & Edelman, 1992; Edelman, 1999; Edelman, Bülthoff, & Bülthoff, 1999; Friedman & Waller, 2008; Friedman et al., 2005; Kourtzi & Nakayama, 2002; Spetch & Friedman, 2003; Spetch et al., 2001). This kind of result is predicted by a model of object recognition in which familiar ("prototype") shapes are hypothesized to be represented in a multidimensional *shape space* (Edelman, 1999); objects close together in this space share similarities among many of the metric parameters that make up the dimensions of the shape space (e.g., curvature; length; number and location of parts; etc.). As objects decrease in similarity on these shape parameters, they also become more distant in the space. This model can be contrasted with other models that posit explicit representations of part structures that are much less sensitive to viewing conditions (e.g., Biederman, 1987; Marr & Nishihara, 1978).

In the *view combination* or *view interpolation* approach, recognition is essentially a form of generalization (cf. Shepard, 1987). When a stimulus is presented, all of the prototypes that share parametric similarities to that particular stimulus view are activated; the amount of activation is a decreasing function of the metric similarity between the parameters of the novel view and the parameters of the stored prototypes. This activation is then used to construct a new view which is compared to the novel input view. To the extent that the activation is sufficient (i.e., above a threshold), the constructed view will be sufficiently similar to the input view that the input view will be recognized.

The view combination model predicts that some novel views can be potentially recognized at least as efficiently as familiar views whereas others will not be recognized as efficiently as familiar views, or even at all (Wong & Hayward, 2005). In particular, novel views that map into the space spanned by the prototypes (interpolated views) are predicted to be recognized better than novel views that map outside this space (extrapolated views). This is the signature result that provides evidence for view combination processes. These predictions have been supported in several previous discrimination learning paradigms with both humans (e.g., Bülthoff & Edelman, 1992; Friedman & Waller, 2008; Friedman, Waller, Hodgson, & Greenauer, in press; Wong & Hayward, 2005) and in some conditions with pigeons (Friedman et al., 2005; Spetch et al., 2006). However, it is not clear how these predictions might apply to moving objects because motion reveals many views of a dynamic object.

1.2. View combination models and dynamic cues

Edelman (1999) claimed that an object rotating in depth gives rise to a two-dimensional *view space* of the object. The view space spans all of the views through which the object has rotated. So, in principle, a view combination model can accommodate moving objects. However, previous research with humans and pigeons suggests that both rigid and non-rigid motion of objects provide a rich variety of dynamic visual cues, *in addition* to shape, that might be used to recognize those objects, categorize them, or ascribe different intentions and emotions to them (e.g., Grossman et al., 2000; Knappmeyer, Thornton, & Bülthoff, 2003; Lander & Bruce, 2000;

Liu & Cooper, 2003; Newell, Wallraven, & Huber, 2004; Pollick, Paterson, Bruderlin, & Sanford, 2001; Spetch et al., 2006; Troje, 2002; Vuong & Tarr, 2004, 2006). For example, moving objects may give rise to unique spatiotemporal signatures that can be directly encoded for the purpose of recognition (Stone, 1998; Vuong & Tarr, 2006), as when different objects move along different paths. Stone (1999), for instance, conceptualized this signature as a directed sequence of views.

Furthermore, predictable motion trajectories of objects (e.g., smooth clockwise rotation in depth) may affect how unfamiliar views are recognized. For example, the perception of motion can lead to representations of views that are not explicitly shown (Kourtzi & Nakayama, 2002). Similarly, Freyd's work on representational momentum has shown that observers encode dynamic representations that anticipate novel views of objects that follow a predictable path (Freyd, 1987; Kelly & Freyd, 1987; see also Mitsumatsu & Yokosawa, 2003; Vuong & Tarr, 2004). That is, when objects are viewed with motion, observers can predict views of them that are "coming up" but have yet to be seen. In the context of the view combination predictions, these prior findings suggest that motion may eliminate some of the differences previously found in the recognition of unfamiliar views; for example, extrapolated novel views that are in the predicted direction may be recognized as well as interpolated views (e.g., Vuong & Tarr, 2004).

Our own cross-species work suggested that although humans and pigeons used motion information in addition to shape to learn to discriminate objects, there were subtle species differences (Spetch et al., 2006). In this previous study, both humans and pigeons learned to discriminate between two differently-shaped objects that each had a characteristic motion. Thus, either the shape or the motion could be used to perform the discrimination. In addition, the objects were either easy to decompose into simple parts or had no clear part structure. On the test trials, the learned objects could appear in their learned motion, the reverse of the learned motion, an entirely new motion, or, an entirely new object could appear in a learned motion. For humans, any change in motion resulted in poorer performance for both object types; however, humans did not respond differentially to new objects that appeared in the learned motions. In contrast, when the objects had no distinctive parts, pigeons (but not humans) relied almost exclusively on motion cues to make their discriminations.

We also obtained evidence to suggest that pigeons processed the two cues independently (Spetch et al., 2006). We therefore suggested that for some kinds of objects, pigeons weighted motion cues more than humans did (relative to shape), but the underlying object-recognition mechanisms are the same. Notably, however, the motion used in previous experiments was informative regarding the objects' identities (e.g., the original learning could have been based on motion cues or shape cues alone). Moreover, the literature is inconsistent with respect to the role of motion on pigeons' recognition of objects when motion is not informative about object identity. Although Cook and Katz (1999) reported that motion facilitated discrimination between a cube and a pyramid, a recent study on pigeons' recognition of human faces showed no facilitation of motion on recognition of novel depth rotations of the faces (Loidolt et al., 2006). In the present study, the motion cues were not informative with respect to identity. Thus, it is an open question whether humans and pigeons will perform similarly with these types of dynamic stimuli.

1.3. A dynamic view interpolation paradigm

In the present experiments, we adapted the view interpolation paradigm used to investigate the role of static cues for object recognition to test dynamic motion cues across different types of objects (Bülthoff & Edelman, 1992). Observers learned to discriminate

between rotating objects from two viewpoints. They were then tested on these *trained views* and on at least two kinds of novel views: *interpolated novel views* were between the shortest angular range spanned by trained views and *extrapolated novel views* were outside of that range. The extrapolated views were both "before" and "after" the trained views in the trained motion sequence, thus providing the means to test whether motion facilitates recognition of certain extrapolated views. As in our previous work (Spetch et al., 2006; Vuong & Tarr, 2006), we used both structurally distinctive objects that could be decomposed into simple parts and structurally similar "amorphous" shapes that were non-decomposable and difficult to discriminate. The stimuli used in the present experiment are illustrated in Fig. 1.

Second, we trained both species to discriminate between temporal segments of rigidly rotating objects. That is, the training stimuli – which were always dynamic sequences – were structured as if observers were standing in one of two places on the circumference of a circle and the objects were rotating at the circle's center through segments that spanned 30°, as shown in Fig. 2. Importantly, the objects always rotated in depth with a global clockwise direction (cf. Bülthoff & Edelman, 1992). We then tested observers on their ability to recognize these objects from novel viewpoints that were either between or beyond the trained viewpoints. To our knowledge, this is the first time that this type of object motion has been used in a comparative study across the two species.

Finally, we used smooth clockwise rotations or we scrambled the rotation by randomizing the frame sequences that made up the coherent motion. Scrambling is a critical manipulation because it preserves the 3D structure of the objects and specific view infor-
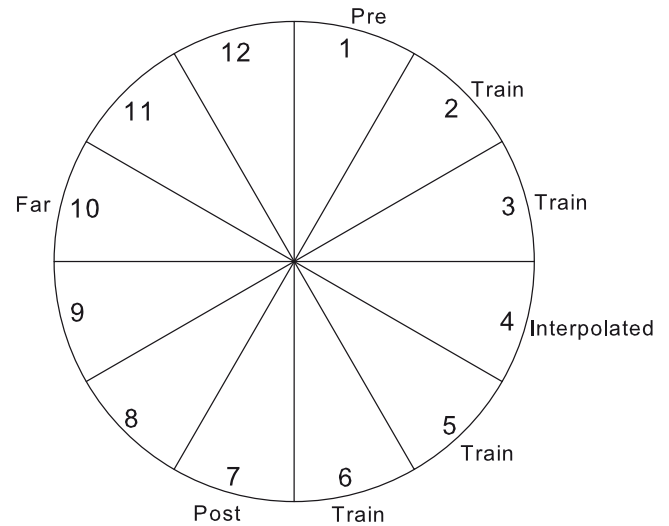

Fig. 2. Example of the "top view" of the experimental conditions for Experiments 1a (humans) and 1b (pigeons). The objects were located in the center of the circle and were rotated around their vertical axes through the different 30° segments shown in the figure. The segments are labeled for one counterbalancing condition only.

mation, while disrupting any dynamic cues and the predictability of the rotation trajectory. This manipulation has been used extensively in human research (e.g., Harman & Humphrey, 1999; Lawson, Humphreys, & Watson, 1994; Liu, 2007; Vuong & Tarr, 2004;
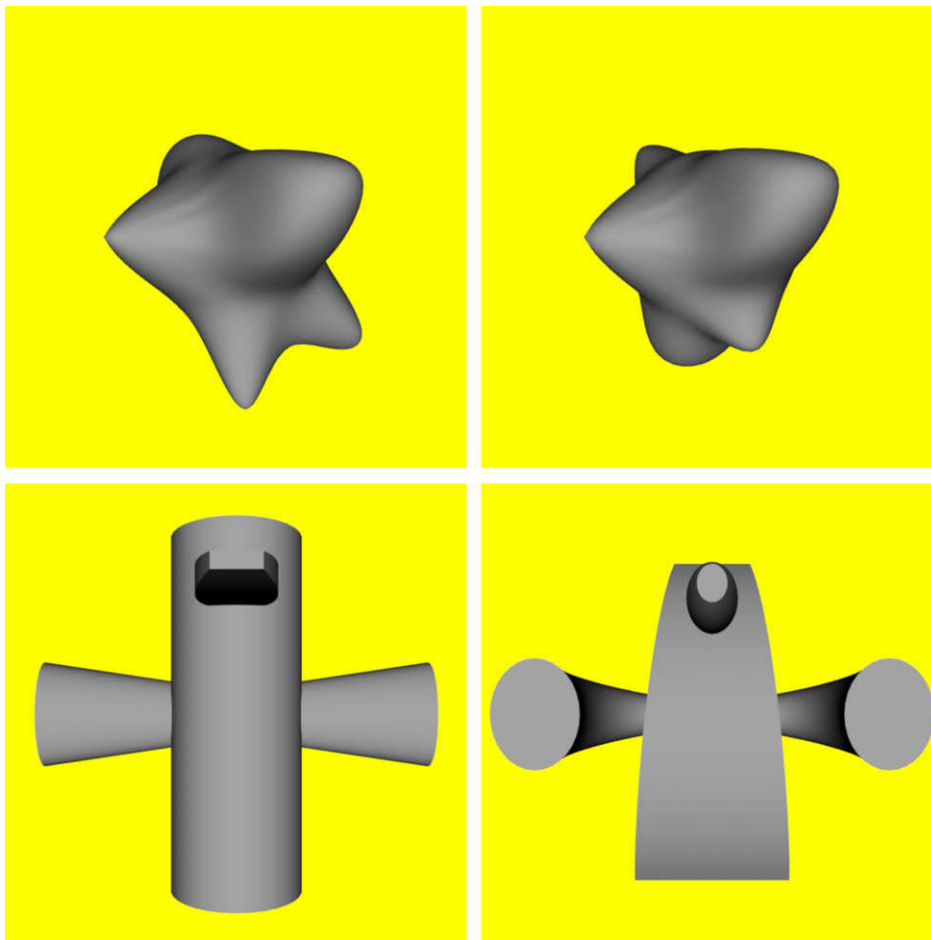

Fig. 1. The similar (top row) and distinctive (bottom row) stimuli used in the present experiments.

Wallis & Bülthoff, 2001), but its precise effect on the role of motion in object recognition remains unclear. In particular, there has not been a consistent advantage when coherent and scrambled motions were directly compared. For example, Harman and Humphrey (1999) found that observers were *better* with scrambled versus coherent motion. They suggested that this may be due to an attentional factor: Coherent motion was predictable so observers did not attend to the individual static images in the series. Lawson et al. (1994), on the other hand, showed an accuracy advantage for coherent motion. We have also shown an advantage for coherent motion in predicting novel views (Vuong & Tarr, 2004).

To our knowledge, scrambled motion of a rotating object has never been used to test object recognition with pigeons. However, Cook et al. (2001) and Cook and Roberts (2007) used scrambled motion to investigate the effect of motion coherence in pigeons' discrimination of motion trajectories. In their studies, the motion trajectories showed a camera view that either navigated through a donut-shaped object, or navigated around the same object. This is the kind of motion that is experienced in navigation, and they found that pigeons discriminated between these trajectories better with coherently ordered frames than with scrambled frames of the motion trajectory. In our current study, the objects themselves carried the motion cues but the motion itself was non-discriminative with respect to shape identity.

In sum, we are comparing across species the effects of dynamic information on recognizing novel views of rotating objects, using a paradigm that has previously been used only to test predictions of a view combination model of static object recognition. In a pure view combination model, if the dynamic trained views are sufficiently close together, objects rotating at interpolated viewpoints should be better recognized than objects rotating at extrapolated viewpoints. Furthermore, objects at all extrapolated viewpoints equidistant to the training viewpoints should be recognized equivalently well (or poorly). By comparison, if motion contributes to recognition in addition to view combination, coherent motion should confer an advantage for some extrapolated viewpoints relative to other such viewpoints. For example, smooth motion can lead to better recognition performance on views that follow the learned views compared to views that precede them (Vuong & Tarr, 2004). This is so even though the two types of novel views are equally distant from *all* of the learned views within the training set. Based on our previous work (Spetch et al., 2006; Vuong & Tarr, 2004), we expect that any advantage of coherent motion may also depend on the stimulus structure and the species. These kinds of findings would imply that the view combination model needs to be augmented to account for the effect of dynamic information, above and beyond the effects that might be attributable to the static view space of a dynamic object.

2. Experiment 1a

In Experiments 1a (humans) and 1b (pigeons), we trained and tested participants with visually similar or distinct objects that were shown either rotating coherently about their vertical axes or with scrambled versions of this motion. Bülthoff and Edelman (1992) used a similar paradigm, but they used only visually similar objects and only coherent motion during training, with static images at test. Moreover, their objects rotated in depth back and forth during training, which eliminated any global motion direction. Thus, with our design, we can address whether humans and pigeons can use motion cues to recognize dynamic objects from novel viewpoints, in line with the view combination predictions. Furthermore, we can test whether both species are sensitive to the global clockwise rotation direction that is evident with coherent motion but not with scrambled motion.

2.1. Method

2.1.1. Participants

The human participants were 34 volunteers (16 males, 18 females) from the University of Alberta participant pool. They received partial course credit as well as performance-based payment for their participation. They were assigned randomly, in groups of four participants to one of the eight experimental conditions formed by the factorial combination of type of motion (scrambled or coherent), whether the stimulus pair that was learned first was distinctive or similar, and which of two sets of particular views were used as the training and test movies (see below). The data from one male and one female were not used because they did not reach the criterion of 70% correct on all four of the training segments during the test trials, leaving 32 participants in the experiment (16 per motion group).

2.1.2. Stimuli, design, and apparatus

One pair of structurally distinctive objects and one pair of similar objects from our previous study were used (Spetch et al., 2006). For each stimulus type (distinctive or similar), one member of the pair was arbitrarily assigned to be the S+ and the other member of the pair was assigned to be the S− for all participants. Both the S+ and S− objects were presented on each trial as animated movies.

Bitmaps of each stimulus were made at each degree of viewing angle by moving a virtual camera clockwise around a circle whose radius was an arbitrary number of units from the center of the objects. The objects were rendered with a matte gray surface, and placed against a uniform yellow background. When displayed side-by-side on the screen, each object in a pair was displayed in an area that was 450 by 450 pixels (approximately 13.1 by 13.1 cm).

The 360 bitmaps for each object were divided into 12 segments of 30 consecutive bitmaps each. Participants received all 30 views for one segment (both S+ and S−) simultaneously, during both training and testing. Half the participants received all of their stimuli in their consecutive order, so that the resulting movie showed smooth motion (the *coherent* group). For the remainder of the participants, the bitmaps within each segment were first divided into 10 clusters of three bitmaps each; these clusters were then randomized anew for each presentation throughout both practice and test trials (the *scrambled* group). We clustered the frames so that local motion processing could still occur (see Vuong & Tarr, 2004). The resulting movies still had some apparent motion, but it was very choppy.

Fig. 2 shows the viewing conditions for four of the eight experimental groups (one group in each motion type x stimulus order condition). For participants in these groups, the training movies were made from the 30 bitmaps in each of Segments 2, 3, 5, and 6 in the figure. Thus, through the course of training, participants saw a full 1/3 of the structure of the stimuli.

The test stimuli consisted of the four training segments, as well as the segment that showed the 30 views that preceded Segment 2 (Pre), the segment that followed Segment 6 (Post), the segment that showed the views in between the training views (Interpolated), and the segment that was taken from the other side of the figure (Far). The remaining four groups were trained with segments 8, 9, 11, and 12 and had corresponding assignments of segments to the other conditions. In particular, the Pre, Interpolated, Post, and Far segments for these groups were Segments 7, 10, 1, and 4, respectively. Thus, the views that were in the Interpolated segment for half the participants were in the Far segment for the other half; similarly, the views that were in the Post segment for half the participants were in the Pre segment for the other and vice versa. This counterbalancing ensured that there was nothing idiosyncratic about the particular viewpoints that could have caused differences in performance for the critical segments.

All participants received two blocks of 40 training trials followed by one block of 80 test trials for each object type (distinctive and similar), for a total of 320 trials. The 30° S+ and S− for a given segment were shown simultaneously on both training and test trials, but only one segment at a time was shown. For each block of 40 training trials, the four training segments (e.g., Segments 2, 3, 5, and 6) were each presented twice in randomized blocks of 8 trials; the S+ in each pair was randomly-selected to be on the right for half the time and on the left for the other half. For the 80 test trials, the eight different test segments (e.g., Segments 1–7 and 10) were each presented twice, randomized in blocks of 16 trials; the S+ in each pair was on the right half the time and on the left the other half. Each movie segment (e.g., Pre, Train, etc.) was thus seen 10 times during the test trials.

The experiment was conducted on a computer that had an NVidia GeForce 7600GS Video card. The stimuli were displayed on a 19" Samsung Syncmaster 940BF LCD monitor that had a 2 ms gray-to-gray response rate, a resolution of 1280 × 1024 pixels and a 60 Hz refresh rate. The frame rate was 30 frames/sec. The stimuli were presented as pairs of animations, centered side-by-side on the screen, which was approximately 60 cm from the participant. There was a button box in front of the monitor with two push button switches that were 8 cm from center to center. Participants responded by pushing the button on the side of the response box that corresponded to the object they thought was the S+.

2.1.3. Procedure

When a participant arrived, he or she was seated in a small room in front of the experimental computer. After signing the consent form, the initial instructions were presented on the computer screen with the experimenter present. The instructions informed the participants that their task was "to learn a discrimination between two stimulus displays of novel objects that are shown as animated movies." They were also told that they could earn money for accurate and fast responding and that if they scored perfectly the amount they would earn would be $8.00. They were told their total earnings at the end of each training and test block. The experimenter left the room after the first few trials of each learning block and after the participant had read the instructions for the test block.

During the first training block, participants could not respond until the movies had been shown for three cycles. Correct responses earned 1¢ and incorrect responses were penalized 1¢. Participants were asked to look at both objects as much as possible; they were told they would have to guess at first which was the correct object but that they would get feedback on each trial and they should use it to figure out which object was the correct object. The participants received visual feedback in the form "You earn 1¢" or "You lose 1¢" after each trial. We used a reward scheme to make the human procedure similar to the pigeon procedure; we have done this in our previous work on cross-species motion perception (Spetch et al., 2006).

For each training trial in this phase, a beep sounded simultaneously with the onset of a fixation point, which remained on for 750 ms. Then the S+ and S− movies for a given training segment were shown simultaneously for three full cycles (3 sec total). A beep sounded at the end of the third cycle to signal that the participant could respond. After the participant responded, the feedback for that trial was displayed for 1 sec. This response was followed by a 1 sec inter-trial interval (ITI).

For the second block of training trials, the procedure was the same, but the participants were additionally told that each time they responded correctly and "are fast enough" they would earn 3¢; otherwise, if they were correct they would earn 1¢, and if they were incorrect 1¢ would be subtracted from their total. The 3¢ reward was given for responses that were made in under 1 sec, but

the participants did not know the exact time that was being used as the criterion. During these trials the movie was still repeated for three full cycles but participants could respond any time after the onset of the stimuli.

For the test trials, the procedure was the same as for the second block of learning trials, except that no feedback was given, although there was still a 1 sec ITI. In addition, participants were warned that some of the animations they would see would be different than those they had previously seen, and that they "should try to decide whether to respond to the right or left side based on which object is the correct one, given your previous feedback." They were told that the same earning scheme was in place as had been for the previous block of trials: 3¢ for correct, fast responses; 1¢ for correct responses that were not fast, and −1¢ for incorrect responses, but that they would not get feedback.

After finishing the test trials for the first stimulus type, the participants were given a short break, and then they proceeded to the second stimulus type. The procedure was identical to that for the first stimulus type.

2.2. Results

2.2.1. Data trimming and analysis

We averaged each participant's correct reaction times (RTs) separately over the distinctive and similar objects and omitted RTs that were more than three *SDs* above these means from further consideration. The omitted trials were counted as errors and comprised 1.5% of the data. There were two participants in the scrambled group who had no correct responses for one test segment each. Their RTs for that condition were replaced by the group means.

For both measures, we first examined performance on the training viewpoints during testing in object type (similar, distinctive) by motion type (coherent, scrambled) analyses of variance (ANOVAs). We next conducted an ANOVA on both measures, with motion type as the between-subjects factor and object type (distinctive or similar) and novel viewpoint (Pre, Interpolated, Post, and Far) as within-subjects factors. Finally, using planned contrasts, we examined more specifically the differences in performance between interpolated and extrapolated novel tests and the effects of motion on performance. For all analyses throughout this study, we used $p < .05$ as the criterion for significance and report η_p^2 as the measure of effect size.

2.2.2. Overall ANOVAs

Fig. 3 shows the mean percent correct and Fig. 4 shows the mean correct RTs for each segment as a function of object type and motion type. For the training viewpoints, distinctive objects were recognized more accurately than similar objects, $F(1,30) = 4.75$, $MSE = 18.48$, $\eta_p^2 = .137$, although the differences were small in an absolute sense (98.3% and 96.8%, respectively). The distinctive objects at the training viewpoints during test were also recognized more quickly than the similar objects, $F(1,30) = 41.63$, $MSE = 45,856.69$, $\eta_p^2 = .581$. The means were 509 and 854 ms, respectively.

For the ANOVA on percentage correct for the novel test segments, there were main effects of viewpoint, $F(3,90) = 13.07$, $MSE = 186.98$, $\eta_p^2 = .30$, and object type, $F(1,30) = 62.89$, $MSE = 214.90$, $\eta_p^2 = .68$. The mean percent correct for the distinctive objects was 97.7% and for the similar objects it was 83.1%. The means for the Pre, Interpolated, Post, and Far viewpoints were 90.8%, 96.1%, 93.0%, and 80.5%, respectively. There was also an interaction between the two factors, $F(3,90) = 16.20$, $MSE = 197.40$, $\eta_p^2 = .35$, which is shown in Fig. 3. None of the remaining effects were significant, $Fs < 1.10$.

The RT data for the omnibus ANOVA mirrored the accuracy data. There were main effects of viewpoint, $F(3,90) = 14.38$, $MSE = 33,076.29$, $\eta_p^2 = .32$, and object type, $F(1,30) = 45.77$,
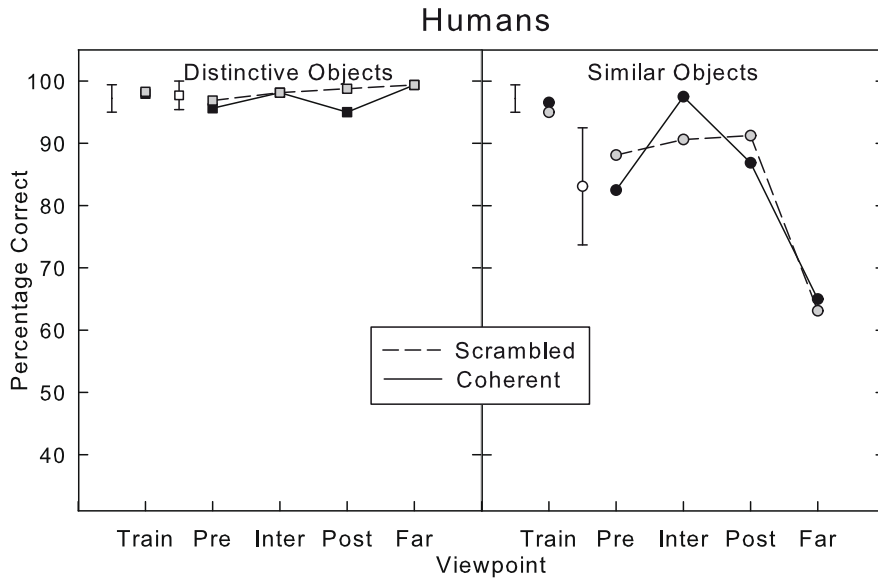
Humans

Fig. 3. Percent correct as a function of motion condition, segment, and type of object for Experiment 1a (human participants). For this and all subsequent figures with data, the error bars are 95% confidence limits (Loftus & Masson, 1994) centered on the means for which they are relevant. Those to the immediate left of the data for the training views were computed from the error term from the motion type by object type ANOVA on only the training stimuli; they are thus the same in both panels. Those to the left of the data for the novel test views were computed separately for each object type in two motion type by viewpoint ANOVAs which excluded the training stimuli.
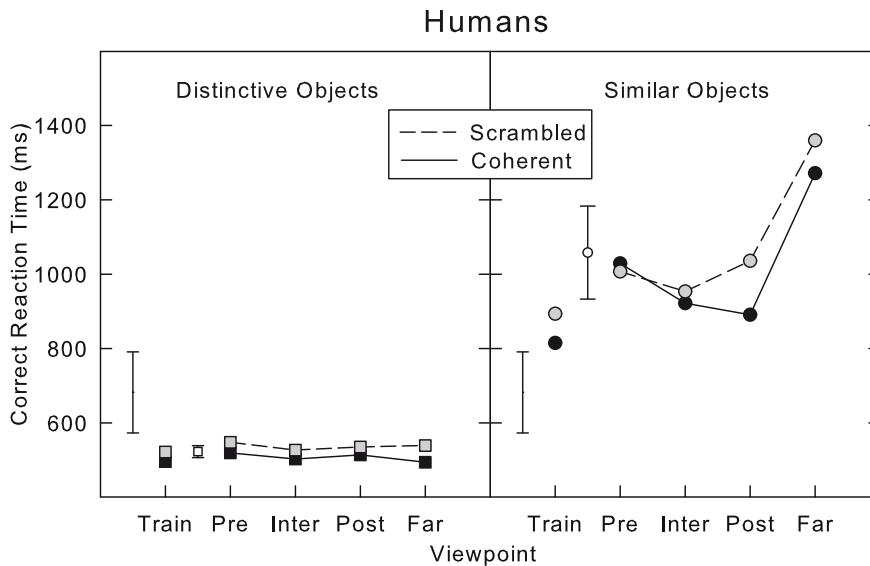
Humans

Fig. 4. Correct reaction times as a function of motion condition, segment, and type of object for Experiment 1a (human participants).

MSE = 402,035.47, η_p^2 = .60. The mean correct RT for the distinctive objects was 523 ms and for the similar objects it was 1059 ms. The means for the Pre, Interpolated, Post, and Far views were 776, 727, 745, and 916 ms, respectively. There was again an object type by viewpoint interaction, $F(3, 90)$ = 15.17, MSE = 33,057.46, η_p^2 = .34 (see Fig. 4). None of the remaining effects were significant, $Fs < 1.00$. There were higher-order polynomial effects for both measures, but as these could have been unduly influenced by the Far pairs (cf., Wong & Hayward, 2005), we do not report them here.

It is clear from Figs. 3 and 4 that there was no effect of viewpoint for the distinctive objects on either measure; the distinctive objects were responded to quickly and accurately at all viewpoints. In contrast, there was a large effect of viewpoint for the similar objects. Moreover, in this and subsequent experiments, there was no indication of any speed–accuracy trade-offs.

2.2.3. Effects of interpolation vs. extrapolation

To examine the predictions of the view combination model directly, we conducted ANOVAs on both measures with just the Pre, Interpolated, and Post novel viewpoints. Motion type was the between-subjects factor and object type and view were within-subjects. We omitted the Far condition because the Pre and Post conditions were equidistant to the Interpolated condition, which enabled us to specifically examine the quadratic components of viewpoint and its interactions (which could not be influenced by the Far condition); these a priori contrasts should be significant if the Interpolated condition is more accurate and/or faster than the Pre and Post conditions.

For the percent correct data, there was a significant quadratic effect of view, $F(1, 30)$ = 9.09, MSE = 83.52, η_p^2 = .233. There was also a significant quadratic component in the viewpoint by motion type interaction, $F(1, 30)$ = 6.60, MSE = 83.52, η_p^2 = .180. For the

coherent motion group, the percent correct for Pre, Interpolated, and Post views was 89.1%, 97.8%, and 90.9%, respectively, and for the scrambled motion group the percent correct was 92.5%, 94.4%, and 95.0%. This interaction indicates that the interpolation effect occurred with coherent motion but not with scrambled motion. This is a unique finding for human participants.

For the RT data, the same ANOVA did not produce significant quadratic effects; however, a separate ANOVA on each motion type condition revealed that for coherent motion, there was a linear effect of viewpoint, $F(1, 15) = 9.50$, $MSE = 8656.40$, $\eta_p^2 = .388$, as well as a linear component to the viewpoint by object type interaction, $F(1, 15) = 11.30$, $MSE = 6226.33$, $\eta_p^2 = .430$. For similar objects undergoing coherent motion, the means for the Pre, Interpolated, and Post viewpoint conditions were 1029, 922, and 891 ms, respectively, and for the distinctive objects, the means were 520, 503, and 514 ms. This interaction indicates that with coherent motion, whereas performance was fast at all of the viewpoints for the distinctive objects it was still fastest for the interpolated views. However, the absolute differences among the distinctive objects were small. In contrast, for the similar objects undergoing coherent motion, there was a 107 ms facilitation in the expected direction for the Interpolated viewpoint, and a further 31 ms facilitation for objects seen in the Post viewpoint. This is some evidence for facilitative effects of interpolation and additional facilitative effects of motion, which will be explored further below.

There were no higher-order polynomial effects for the scrambled motion condition, $Fs < 1.11$. However, the means for both object types were consistent with the view combination predictions. For the similar objects undergoing scrambled motion, the mean RTs for the Pre, Interpolated, and Post conditions were 1007, 954, and 1036 ms, respectively, and for the distinctive objects they were 548, 527, and 535 ms.

In general, the data indicate that for humans observing rotating objects, successful interpolation between familiar parts of their structure can be achieved with both coherent and scrambled motion, but the effects are stronger when the motion is coherent and the objects are difficult to discriminate. In addition, there is a hint in the RT data that coherent motion provided a speed advantage for viewpoints that followed the training views. We explore this further in the next section.

2.2.4. Effects of motion type

The linear effects of view in the RT analysis of the similar objects indicate that there was an effect of motion specific to the coherent motion condition. Motivated by earlier work (e.g., Kelly & Freyd, 1987; Vuong & Tarr, 2004) we examined this further through planned tests. For coherent motion, the difference in RT between the Pre and Post segments was 138 ms for the similar stimuli, $t(15) = 3.28$, $SD_{diff} = 168.55$, but only 5 ms for the distinctive stimuli, $t(15) < 1.00$. In contrast, for the scrambled motion condition, the difference in RT between the Pre and Post segments was 29 ms in the wrong direction for the similar stimuli, $t(15) < 1.0$, and was 13 ms for the distinctive stimuli, $t(15) < 1.0$. These effects were not significant for either type of motion in the error data.

Thus, there was a clear Post vs. Pre segment advantage in RT for similar stimuli that were moving coherently. Nevertheless, apart from this difference, there was not a general advantage for coherent motion when compared to scrambled motion; perhaps this is because the training views comprised a total of 1/3 of the objects' structure.

3. Experiment 1b

In this experiment, we replicated Experiment 1a with pigeon subjects, using the same stimuli and design as in Experiment 1a

and altering the procedure only as necessary to accommodate the different species.

3.1. Method

3.1.1. Subjects

Twelve adult pigeons (*Columba livia*) with varied experimental histories served in the experiment. None had previously been trained with moving stimuli or with static versions of the stimuli used in the present study. The birds were maintained at 85–90% of their free-feeding weights by pigeon pellets obtained during experimental sessions and supplemental feedings in the home cages. The birds were housed in large individual cages under a 12:12-h light:dark cycle (with light onset at 6:00 a.m.). Grit and water were freely available in the home cages.

3.1.2. Apparatus

The experiment was conducted in a large custom-built operant chamber, 44 cm high, 32 cm deep, and 74 cm wide (inside dimensions). The LCD monitor and video card were identical to those used in Experiment 1a and also identical to the apparatus used in Spetch et al. (2006) which showed that pigeons were sensitive to motion cues. The monitor was equipped with a 17" Carroll Touch infrared touch frame that recorded the x- and y-coordinates of the pigeons' pecks, and two solenoid-type bird feeders, one on each side of the monitor. Lamps located within each feeder illuminated feeder presentations, and photocells measured the duration of head entries into the hoppers to limit feeding durations to 1 or 2 sec per food presentation (depending on the bird's weight). The chambers were connected to microcomputers located in an adjacent room. These computers controlled all of the experimental contingencies and recorded the responses.

3.1.3. Stimuli and design

The stimuli and design were identical to those used in Experiment 1a. Half of the pigeons in each motion group were trained with segments 2, 3, 5 and 6 whereas the remaining pigeons were trained with segments 8, 9, 11, and 12. Again, only one segment of S+ and S− was shown at a time. Within each group, half of the birds were first trained and tested with distinctive objects and then were trained and tested with similar objects. The reverse order was used for the remaining birds. With each object type, the pigeons were trained to an accuracy criterion and then were given eight sessions of testing.

3.1.4. Training procedure

The pigeons received one session per day, 5 or 6 days per week. Each pigeon first received one or more sessions of S+ only training to establish reliable pecking at the moving S+. On these trials, the S+ was presented by itself on either the left or the right half of the screen. After 8 sec or if the bird pecked on that side of the screen, a food reward was presented. Next the pigeon received several sessions of training in which the S+ and S− animations were presented simultaneously on the left and right halves of the screen. The side containing the S+ was counterbalanced across trials. The animations remained on for a minimum of two complete cycles (60 bitmaps at 30 bitmaps/sec) and then the first peck to either side terminated the trial. If there was no peck the trial timed out at 60 sec. A peck to the side containing the S+ was followed by food reward; no reward was presented if the bird pecked the S− side. Training continued until the bird chose with 80% or higher accuracy for two consecutive sessions, and then the percentage of reinforcement for correct choices was reduced to 50% until the bird reached the 80% accuracy criterion for five consecutive sessions. Correct choices on non-reinforced trials ended the same way as incorrect choices. This partial reinforcement was designed to

encourage persistent pecking during subsequent unreinforced probe tests.

3.1.5. Testing procedure

Test sessions consisted of a mixture of reinforced baseline trials and non-reinforced probe trials. The probe trials presented animations of the same S+ and S− objects as on training trials, but the viewpoint was varied (see Fig. 2). For most birds, each of the eight test sessions provided three cycles in which there were 16 reinforced trials with the training segments, four probe (non-reinforced) trials with the training segments, and two probe trials with each of the test segments. For each segment type, the S+ was equally often presented on the left or right. Across all test sessions, the birds received 192 reinforced trials and 48 non-reinforced trials with each of the four training and four test segments. For three of the birds in the coherent motion condition, a programing error during testing with the distinctive objects resulted in a slightly different number of trials (range of 168–221 non-reinforced training trials over the four training segments and 42–56 probe trials for each novel test segment). Results for these birds were not noticeably different from the other birds in their condition, and they received the standard number of trials during testing with the similar objects.

3.2. Results

3.2.1. Acquisition

Acquisition of the discrimination was assessed in terms of sessions to criterion, starting with the first simultaneous session through to meeting the accuracy criterion (80% or higher for two consecutive sessions) for moving to testing. An ANOVA with motion type (coherent or scrambled) and order (distinctive or similar first) as between-subjects factors and object type (distinctive or similar) as the within-subjects factor showed only a significant effect of object type, $F(1,8) = 27.7$, $MSE = 105.63$, $\eta_p^2 = .776$. Pigeons took significantly more sessions to acquire the discrimination with similar objects ($M = 31.8$, $SD = 14.0$) than they did with distinctive objects ($M = 9.8$, $SD = 3.2$). No other factors and no interactions were significant.

3.2.2. Data trimming and analysis

Test results were analyzed in terms of mean accuracy for each test segment type, averaged across the eight sessions of testing.

We eliminated from the accuracy data any trials for which the first peck was less than 200 msec. We analyzed the accuracy data identically to that of Experiment 1a. We also examined reaction time on correct trials, measured as the time from onset of the stimuli to the first choice peck, again omitting trials for which RTs were less than 200 msec. However, the remaining RT data were very variable between birds (e.g., from a minimum of 1.15 sec for one bird to a maximum of 21.18 sec for another). Consequently, we do not report the RT data.

3.2.3. Overall ANOVAs

Fig. 5 shows the mean percent correct for the pigeons for each segment as a function of object type and motion type. First, for the non-reinforced tests at the training viewpoints, we conducted motion type by object type ANOVAs. There was an effect of object type, $F(1, 10) = 24.03$, $MSE = 24.93$, $\eta_p^2 = .706$. The means for the similar and distinctive objects were 85.3% and 95.3%, respectively, replicating the human results.

We next conducted object type (distinctive, similar) by novel viewpoint (Pre, Interpolated, Post, Far) by motion type (coherent, scrambled) ANOVAs in which object type and viewpoint were within-subjects and motion type was between-subjects. There was a main effect of object type, $F(1, 10) = 49.46$, $MSE = 187.63$, $\eta_p^2 = .832$ (see Fig. 5). As we have found on previous occasions (Spetch et al., 2006), pigeons responded more accurately to the distinctive stimuli than to the similar stimuli (87.4% vs. 67.4%, respectively). There was also a main effect of viewpoint, $F(3, 30) = 23.52$, $MSE = 98.96$, $\eta_p^2 = .702$. The means for the Pre, Interpolated, Post, and Far conditions, respectively, were 78.2%, 86.9%, 81.3%, and 63.8%. As with the human participants, there was also a significant object type by viewpoint interaction, $F(3, 30) = 8.56$, $MSE = 126.11$, $\eta_p^2 = .461$. Basically, performance with the distinctive objects was almost uniformly accurate across all views, but for the similar objects, performance tended to be worse in the Pre, Post, and Far conditions than it was in the Interpolated condition. This pattern of performance is similar, in general, to what was observed in the human data.

3.2.4. Effects of interpolation vs. extrapolation

An ANOVA on the percent correct data with only the Pre, Interpolated, and Post views with motion type as the between-subjects factor and object type and viewpoint as the within-subjects factors showed a significant quadratic effect of viewpoint, $F(1, 10) = 8.89$,
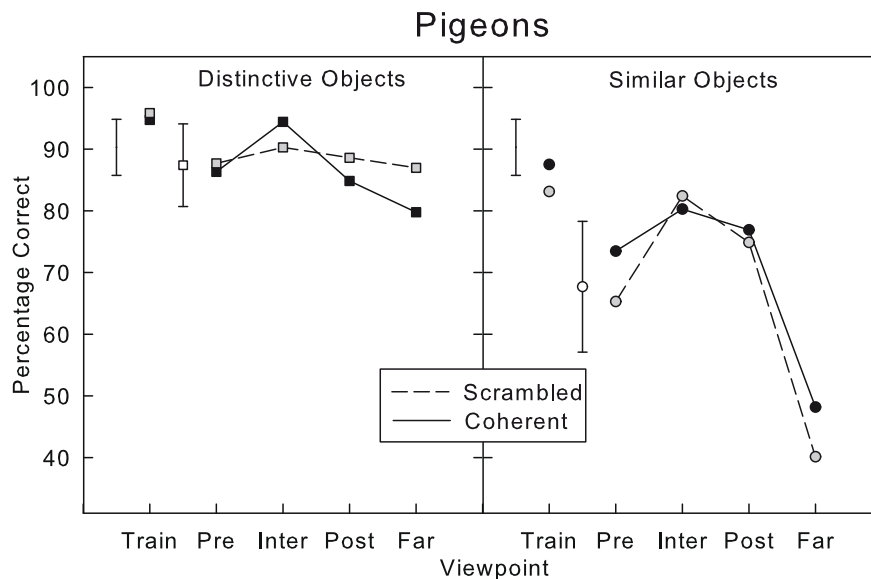

Fig. 5. Percent correct as a function of motion condition, segment, and type of object for experiment 1b (pigeon subjects).

$MSE = 90.73$, $\eta_p^2 = .470$, but no quadratic component to the viewpoint by motion type interaction. Across motion types, the means for the Pre, Interpolated and Post views for the similar objects were 69.4%, 81.3%, and 75.9%, respectively, and for the distinctive objects, they were 87.0%, 92.4%, and 86.7%. Similarly, across object types, the means for the Pre, Interpolated, and Post views undergoing coherent motion were 79.9%, 87.4%, and 80.9%, respectively, and for objects undergoing scrambled motion the means were 76.5%, 86.3%, and 81.7%. Thus, the birds were more accurate on the interpolated views than on either of the extrapolated views for both motion types and both object types.

3.2.5. Effects of motion type

A priori comparisons of accuracy on the Pre and Post viewpoints showed no significant effects for either object type or motion type. We address the potential reasons for this in Experiment 2.

3.3. Discussion

The significant effects in the omnibus ANOVA for accuracy on both training and novel test viewpoints were identical across species. In addition, both species provided evidence that they interpolated between the training viewpoints. Finally, the scrambling manipulation did not produce a general decrement in discrimination performance for either species; however, the humans did have an advantage for views that followed the training views in the coherent motion condition, especially for the similar objects. Further, the effects of motion observed for human participants are not likely due to differential attention to the two types of motion (cf. Harman & Humphrey, 1999). This is because there was no difference in performance among the trained views for the two different motion conditions during the test phase, and previous work has shown that participants are more likely to be attentive to scrambled rather than coherent image sequences (e.g., Harman & Humphrey, 1999).

It is important to re-emphasize that the four training segments were presented individually in random order during both training and test. With similar objects, humans, but not pigeons, seemed able to extrapolate the consistent rotation direction in each of these training segments to novel segments that continued the global rotation direction (i.e., the Post segment) but not to the preceding novel segments. There is some limit to this extrapolation; performance on the Far segment was extremely poor (cf. Wong & Hayward, 2005). The lack of motion effects for pigeons may have been due to the stimulus and training regimen. We address this further in Experiment 2.

3.4. Experiment 2

The type of training used in Experiment 1b, as well as the method of scrambling, may have been responsible for the lack of difference between motion conditions in the Pre vs. Post comparisons for the pigeons. Recall that we presented the training views one segment at a time, and the training viewpoints were presented in randomized order (e.g., any combination of segments 2, 3, 5, or 6 could be presented from trial to trial; see Fig. 2). This might have reduced the perception of the objects as moving in a global clockwise direction in the coherent motion condition. In addition, to scramble the motion, we combined every three views (3°), in proper sequential order, within a 30°-view sequence, and then randomized the resulting 10 sequences. However, we did not change the clockwise motion within each of the 3° segments. Thus, pigeons may still be sensitive to this local rotation direction. Our previous work has shown that pigeons are highly sensitive to motion cues, particularly for similar objects (Spetch et al., 2006). Furthermore, Lawson et al. (1994) have shown that, under some conditions, hu-

man observers can also be sensitive to local motion cues. The combination of presenting one segment at a time (potentially diminishing global coherence) and allowing the 3° segments to all move in the same direction (potentially diminishing the disruptive effect of scrambling) could both have served to lessen the perceptual differences between the coherent and scrambled motion conditions for pigeons.

In Experiment 2, we therefore explored the effect of motion on pigeons' discrimination performance further by changing the animations used in training and testing in four important ways, as illustrated in Fig. 6. First, the range of views over which the birds saw the objects in training was smaller, spanning only 1/6 rather than 1/3 of the full viewing range. Second, rather than presenting a single training segment on each training trial, each training animation presented both of the training segments on each cycle. For example, for birds in the coherent group, the animation started at the beginning of segment 3 and rotated smoothly from 3 to 5, omitting segment 4. At the end of segment 5, the movie again rotated smoothly from the beginning of segment 3 to the end of segment 5, omitting segment 4. This directional sequencing across segments should provide a more coherent sense of global motion than in the previous experiments because the 30° training segments were always displayed in the correct sequence, instead of being displayed in a random order. Importantly, this manipulation allowed us to further rule out an explanation based purely on temporal associations (e.g., Wallis & Bülthoff, 2001). By immediately proceeding from segment 3 to segment 5, we place very disparate views in close temporal proximity. Therefore, on a "pure" temporal association account, we would not expect any benefit for interpolated views if segments 3 and 5 become associated into a single representation.

For birds in the scrambled group, the animation started at a randomly-selected location in either segment 3 or 5. The animation consisted of a random arrangement of 3° motions from across these segments. Moreover, the three frames, although "stitched together" in sequence, rotated in either a clockwise or counterclockwise direction, so that there was a perception of jerky non-coherent motion with no consistent direction.

The third change we made was that we also tested segments that were closer to the training views. In Fig. 6, the segments labeled Pre-Far and Post-Far are the identical distance to Training segments 3 and 5 as they were in Experiments 1a and 1b. However,
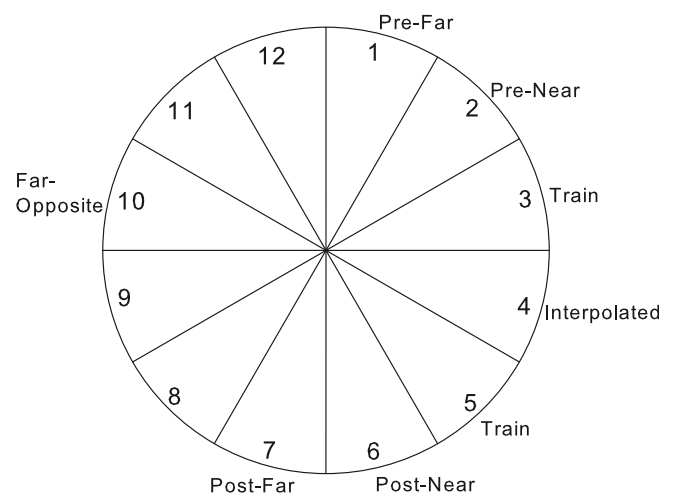

Fig. 6. Example of the "top view" of the experimental conditions for Experiment 2 (pigeon subjects). The objects were located in the center of the circle and were rotated around their vertical axes through the different 30° segments shown in the figure.

the segments labeled Pre-Near and Post-Near were previously training segments; in the present experiment they represent novel test segments. It is possible that for these kinds of objects and motion, pigeons need the testing segments to be closer to the training segments to display robust effects of motion (e.g., Friedman et al., 2005; see also Friedman & Waller, 2008, Experiment 2).

Finally, the test stimuli consisted of static single frames rather than animations. On test trials, the middle frame of a given segment (e.g., Pre-Near, Pre-Far, etc.) was presented as a static image. We did this to provide a more rigorous test of whether coherent motion enhances the ability to extract and generalize the object structure from a limited set of learned views. With static tests, the test stimuli presented to both groups are identical and so any differences between the groups must reflect the differences in the learned object representations to which the test stimuli are compared.

3.5. Method

3.5.1. Subjects

Eight adult pigeons (*C. livia*) with varied experimental histories served in the experiment. None had previously been trained with moving stimuli or with static versions of the stimuli used here. The birds were housed and maintained as described in Experiment 1b.

3.5.2. Apparatus, stimuli, and design

The experimental apparatus and the objects used were identical to those described in Experiment 1b. The training animations differed from Experiment 1a and 1b in that both segments 3 and 5 were presented on each trial. For birds in the coherent group, the animations rotated from segments 3 to 5, skipping over segment 4. For birds in the scrambled groups, randomly-selected sets of three frames from either segment were presented in clockwise or counterclockwise order. Test stimuli consisted of a single static frame from the center of each segment.

The design of the experiment was similar to Experiments 1a and 1b except that only one set of views was used (i.e., the training views were always segment 3 and 5). Four randomly-selected birds were trained with coherent motion and the remaining birds were trained with scrambled motion. Order of exposure to object type

(distinctive and similar) was counterbalanced across birds in each motion type group.

3.5.3. Procedure

The training procedure was the same as in Experiment 1b. The testing procedure was similar to that in Experiment 1b except that the test sessions consisted of a mixture of reinforced trials with the training animations, and non-reinforced probe tests with static images. On these test trials, static images of the S+ and S− were presented until the bird pecked at one of them. Static probe tests were presented with each of 8 segments (see Fig. 6): 1 (Pre-Far), 2 (Pre-Near), 3 (Train), 4 (Interpolated), 5 (Train), 6 (Post-Near), 7 (Post-Far) and 10 (Far-Opposite). The probe tests were the middle (15th) bitmap from each segment. Each session consisted of three cycles that each provided two trials with each type of probe test and 24 reinforced training trials.

3.6. Results

3.6.1. Acquisition

An ANOVA on session to criterion (conducted as described in Experiment 1b) showed only a significant effect of object type, $F(1,4) = 10.57$, $MSE = 9.94$, $\eta_p^2 = .726$. Pigeons took longer to acquire the discrimination with similar objects ($M = 15.1$, $SD = 5.8$) than with distinctive objects ($M = 10.0$, $SD = 3.5$), as before. No other factors and no interactions were significant.

3.6.2. Data trimming and analysis

Test results were analyzed in terms of mean accuracy for each test segment type, averaged across the eight sessions of testing. We again eliminated from consideration any trials for which the RT to the first peck was less than 200 ms. Like Experiment 1b, the remaining average RTs in each segment varied widely across birds (e.g., from a minimum of 1.26 sec for one bird to a maximum of 13.35 sec for another bird; both were in the coherent motion condition), so we report only the accuracy data.

3.6.3. Overall ANOVAs

Fig. 7 shows the mean percent correct for the pigeons for each segment as a function of object type and motion type. Performance
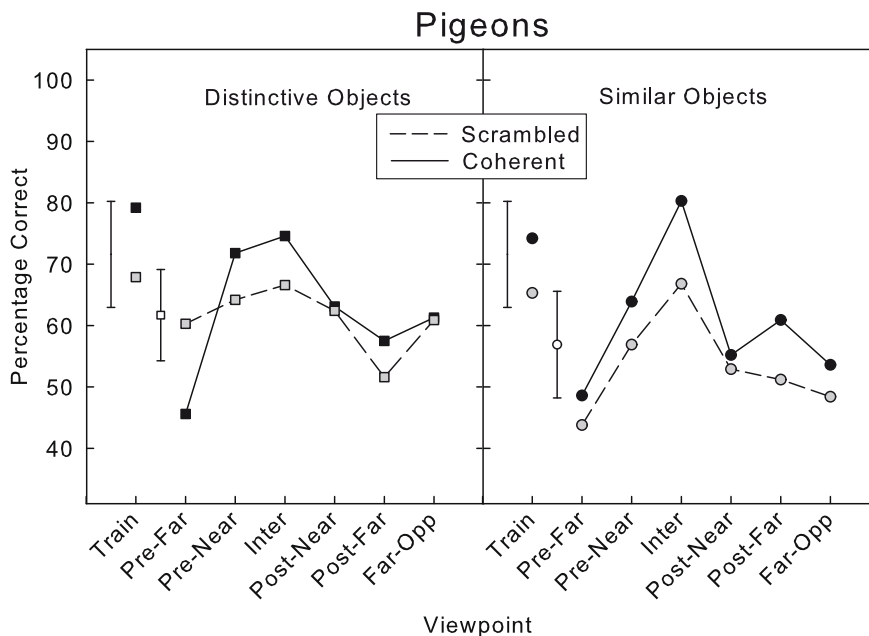

Fig. 7. Percent correct as a function of motion condition, segment, and type of object for Experiment 2 (pigeon subjects).

in the coherent motion condition was worse than it was in Experiment 1b, particularly for the distinctive objects. This could be because there were overall fewer training views (120° in Experiment 1b and 60° in Experiment 2), or because the tests were static, rather than dynamic views. We believe this is likely to be due to the latter, but cannot distinguish between these two possibilities with the present data.

There were no significant effects on accuracy for the training stimuli that were not reinforced during testing, although the effect of motion type approached significance, $F(1, 6) = 4.63$, $MSE = 89.44$, $\eta_p^2 = .436$, $p = .08$. The mean accuracy for the coherent motion group was 76.7% and for the scrambled motion group it was 66.6%.

For the novel test viewpoints, the main effect of object type approached significance, $F(1, 6) = 5.61$, $MSE = 97.79$, $\eta_p^2 = .483$, $p = .06$. The distinctive objects were marginally easier overall than the similar objects (61.7% vs. 56.9%, respectively). The main effect of viewpoint was significant, $F(5, 30) = 30.94$, $MSE = 32.09$, $\eta_p^2 = .838$, and was mitigated by a viewpoint by motion type interaction, $F(5, 30) = 3.90$, $MSE = 32.09$, $\eta_p^2 = .394$, as shown in Fig. 7. Pigeons were more accurate with coherent motion than with scrambled motion at almost every test viewpoint, but particularly so for the Interpolated views (e.g., for the Interpolated views of similar objects, the mean for the coherent motion condition was 80.3% and the mean for the scrambled motion condition was 66.8%; for distinctive objects the mean for the coherent motion condition was 74.6% and for the scrambled motion condition it was 66.6%). For the coherent motion condition, the means averaged across object types for the Pre-Far, Pre-Near, Interpolated, Post-Near, Post-Far, and Far-Opposite conditions were 47.1%, 67.9%, 77.4%, 59.2%, 59.2%, and 57.4%, respectively, and for the scrambled motion condition, the means for these viewpoints were 52.0%, 60.6%, 66.7%, 57.6%, 51.4%, and 54.6%. There were significant higher-order polynomial effects in this analysis, but as with the previous experiments, these might have been influenced by the Far-Opposite condition, so they will not be discussed.

3.6.4. Effects of interpolation vs. extrapolation

We conducted an ANOVA using only the Pre-Near, Interpolated, and Post-Near views, with motion type as the between-subjects factor and object type and viewpoint as the within-subjects factors. There was a significant linear effect of view, $F(1, 6) = 8.85$, $MSE = 30.95$, $\eta_p^2 = .596$, as well as a significant quadratic component, $F(1, 6) = 55.03$, $MSE = 22.51$, $\eta_p^2 = .902$, and a significant quadratic component to the object type by viewpoint interaction, $F(1, 6) = 7.92$, $MSE = 41.78$, $\eta_p^2 = .569$. Across motion types, the means for the similar objects were 60.4%, 73.6%, and 54.0%, respectively, for the Pre-Near, Interpolated, and Post-Near views, and for the distinctive objects, the means were 68.0%, 70.6%, and 62.8%. Thus, the Interpolated views were more accurate than either of the nearest extrapolated views for both object types, but the effect was more striking for the similar objects.

The quadratic component of the viewpoint by motion type interaction approached significance, $F(1, 6) = 4.70$, $MSE = 22.51$, $\eta_p^2 = .439$, $p = .08$. Again, both types of motion showed some facilitation for interpolated viewpoints, but in this experiment, the effect was stronger for coherent motion (see Fig. 7). This may mean that activation due to interpolation and motion type combine.

3.6.5. Effects of motion type

To further examine the effect of motion type, we conducted an object type by Pre/Post by Near/Far by motion type mixed ANOVA; motion type was between-subjects. The new results of interest are the Pre/Post and Near/Far variables and their interactions; the main effect of object type, $F(1, 6) = 6.91$, $MSE = 67.19$, $\eta_p^2 = .535$, did not interact with any other variables in this analysis.

There was a main effect of Near/Far, $F(1, 6) = 42.71$, $MSE = 29.51$, $\eta_p^2 = .877$. Responses to the Near views were more accurate than responses to Far views, 61.3% vs. 52.4%, respectively. There was also an interaction between Pre/Post and Near/Far, $F(1, 6) = 25.62$, $MSE = 21.10$, $\eta_p^2 = .810$, and importantly there was a significant interaction between those two variables and motion type, $F(1, 6) = 16.09$, $MSE = 21.10$, $\eta_p^2 = .728$. The simple interaction between Pre/Post and Near/Far was significant for the coherent motion group, $F(1, 3) = 44.95$, $MSE = 19.32$, $\eta_p^2 = .937$, and was not significant in the scrambled motion group, $F(1, 3) < 1.00$. Collapsing across object types (because the variable did not interact with any other), for the coherent motion condition, the 8.7% ($p = .10$) difference between the Pre-Near and Post-Near views was not significant, but the 12.1% facilitation for the Post-Far relative to Pre-Far viewpoints was significant, $t(3) = 5.03$, $SD_{diff} = 4.80$, $p < .02$. There were no Pre-Post differences in the scrambled motion condition ($ps > 0.12$). Thus, the coherent motion condition showed the expected pattern of facilitation, but only for the Post-Far stimuli. Moreover, it should again be noted that the Pre-Far and Post-Far viewpoints in the present experiment are the same segments as the Pre and Post viewpoints in Experiments 1a and 1b.

3.7. Discussion

In Experiment 2, we provided pigeons with stronger global direction cues by having the stimuli "sweep" in the same direction across the two training segments on each training trial. At the same time, we eliminated both local and global directional coherence in the scrambled condition (see also Lawson et al., 1994). We again obtained significantly better performance on the interpolated views, and consistent with the human data of Experiment 1a, this effect was larger for the similar objects than it was for the distinctive objects. Also consistent with the human data in Experiment 1a, the changes in how the coherent and scrambled motion were implemented in Experiment 2 resulted in an effect of motion type in the coherent motion condition, but only for the Post-Far views relative to the Pre-Far views. No similar interaction was found for Pre-Near and Pre-Far views (although the effect was in the expected direction). It may be that both of the Near views receive sufficient activation from the training views to be relatively well recognized (see Wong & Hayward, 2005).

4. General discussion

The present study directly tested view-based mechanisms across humans and pigeons in the context of dynamic stimuli. We found that under some conditions, humans and pigeons performed similarly: both species used the global rotation direction to help them recognize novel views of learned dynamic objects that continued the rotation trajectory (Experiments 1a and 2), and both species could interpolate novel views of learned dynamic objects, irrespective of whether the motion was coherent or scrambled (Experiments 1a, 1b, and 2), although the interpolation effects were somewhat stronger with coherent motion. Finally, both humans and pigeons found novel views of distinctive objects easier to recognize than novel views of similar objects (see also Spetch et al., 2006; Vuong & Tarr, 2006).

Despite the species similarities we found, some species differences were also apparent. For example, pigeons did not show effects of motion type until we made the motion direction more salient and removed all local directional cues from the scrambled motion. These results highlight how the role of motion in generalizing to novel views depends, perhaps critically, on both stimuli and species.

4.1. Role of coherent motion for view extrapolation

One critical finding was the differential effects of coherent motion for recognizing novel extrapolated views across species and object type. For humans (Experiment 1a), there was evidence that smooth, coherent motion of structurally similar objects provides additional information in long-term memory that is advantageous, insofar as it enables observers to predict views that follow the training views. No similar advantage was found with scrambled motion or with structurally distinct objects. This finding extends earlier work that provides evidence for similar short-term memory predictions for upcoming views of rotating objects (e.g., Vuong & Tarr, 2004).

For pigeons (Experiment 2), when we presumably strengthened the motion coherence of the training segments, we also obtained evidence that coherent motion facilitated accurate recognition for the Far-Post viewpoints relative to Far-Pre viewpoints. It appears that the combination of the new training conditions and scrambling method were effective in causing the birds to be able to take advantage of motion, above and beyond the advantage that accrues to the interpolated views via view combination. The effect of motion did not seem to depend on object type for pigeons as it did for humans.

4.2. Role of coherent motion for view interpolation

Our second main finding was that for both species, interpolated views were generally recognized more accurately (and for humans, faster) than extrapolated views in both motion conditions, but the effects were larger for coherent motion. This finding for humans replicates earlier studies with static objects (e.g., Bülthoff & Edelman, 1992; Friedman et al., 2005; Spetch & Friedman, 2003) and extends it to the case of dynamic stimuli. For pigeons, better recognition of interpolated than extrapolated views has also been found previously, but only when training views were sufficiently close, or when pigeons viewed real 3D objects (Friedman et al., 2005; Spetch & Friedman, 2003; Spetch et al., 2001). The present results add to the small literature showing an interpolation effect in pigeons, and they extend this effect to the case of moving objects.

Although interpolation effects tended to be larger for coherent motion than for scrambled motion, both species showed some evidence of interpolation in the scrambled condition (e.g., on RT for humans in Experiment 1a and on accuracy for pigeons in Experiment 2). This is theoretically interesting because it implies that they inferred the structure of the interpolated portion from having seen the two sets of training views, even though they were moving in a scrambled manner. Therefore, whatever the underlying mechanisms for view combination may be, our results show that disrupting the spatiotemporal contiguity of the stimuli does not impair the ability to extract the necessary features for these mechanisms to successfully function. This is a novel finding and it suggests that view extraction may be a rapid process, because we preserved local coherence to within only three frames (100 ms) in the scrambled condition. This finding is striking, given the very different visual systems and demands on those systems across the two species.

4.3. Augmenting view combination models

Across Experiments 1 and 2, there is evidence that coherent motion cues affect recognition for both species under some conditions, beyond the facilitative effect of view combination alone. For humans in Experiment 1a and pigeons in Experiment 2, the direction of coherent motion alters the ease with which a novel view is recognized, with views that follow the direction of motion generally being recognized more readily than views that precede the direction of motion when both Pre and Post views are sufficiently far from the training views; for pigeons, it seems also necessary that there are no local motion cues during learning. This finding implies that the view combination model needs to be augmented to account for the role of dynamic cues. We believe that, as a model of the representation and recognition of static objects and scenes, view combination probably accounts for more data than any other available model (such as more part-based models, e.g., Hummel & Biederman, 1992). However, the view combination model does not currently accommodate the effects of motion cues on recognition reported here and elsewhere (e.g., Spetch et al., 2006). Further, our results suggest that additional mechanisms such as temporal association (e.g., Wallis & Bülthoff, 2001) cannot completely account for the data.

Stone (1999) proposed that coherent motion can lead to view-bias due to the encoding of a directed sequence of views, and provided some statistical support for this idea. Given the interpolated/extrapolated tests we used, we found more direct support for this idea. This view-bias may be one way that view combination models could be augmented. That is, it is already acknowledged that static trained views may activate extrapolated as well as interpolated views, especially if the extrapolated views are sufficiently close to the trained views (e.g., Wong & Hayward, 2005). This may be why the pigeons in Experiment 2 did not show effects specific to motion for the Near views – they were receiving sufficient activation from the training views to be reasonably well responded to. With dynamic objects at farther distances from the trained views, global directional motion appears to have biased novel views that followed this direction (i.e., the Post-Far test viewpoints) rather than views that did not (i.e., the Pre-Far test viewpoints). This proposal is consistent with results obtained from other recognition paradigms that have used dynamic objects (e.g., Freyd, 1987; Kourtzi & Nakayama, 2002; Vuong & Tarr, 2004). Interestingly, any view-bias in the present experiments remained the strongest for novel Interpolated views, which is consistent with the basic mechanisms proposed by view combination models (Edelman, 1999). However, as noted above, pigeons required a stronger manipulation of motion coherence than humans to show this effect when motion was not discriminative between the S+ and S−. For pigeons, it may be that characteristic motion is a more significant source of discriminative information than is motion per se. For example, Cook et al. (2001) found an effect of coherent and scrambled motion when the motion carried discriminative information relevant to navigation (i.e., moving through or around an object).

4.4. Underlying neural mechanisms for view combination

Overall, the data indicate that humans and pigeons incorporated coherent motion into the dynamic object representation, in addition to any static structure. It is possible that static cues (from a metric shape space) and motion cues (from a metric motion space) are processed in parallel by different systems in both species. In the primate visual system, shape cues are processed along a ventral pathway from the occipital lobe extending to the temporal pole. Motion cues, on the hand, are processed along a more dorsal pathway from the occipital lobe to parietal and motor cortices (see Giese & Poggio, 2003; Ungerleider & Mishkin, 1982). In the avian brain, there is a rostral pathway extending from the superficial layers of the optic tectum to the neostriatum frontale which processes shape (and color); and a more caudal pathway extending from the deep layers of the optic tectum to the neostriatum intermediale which processes motion. Lesions to the rostral pathway impaired shape perception but spared motion perception; by comparison, lesions to the caudal pathway impaired only motion perception (Nguyen et al., 2004). Our results are consistent with this functional organization.

However, if motion cues add to the activation that is summed during recognition, then it is necessary to determine how this happens. It is unlikely that the Pre-Post differences observed in Experiments 1a and 2 happened because the motion exposed more of the objects' structure or other visual features such as parts, because the facilitation only occurred during coherent motion. More likely, the coherent motion may have provided additional input from the dorsal pathway to a generalization mechanism which allowed it to better assess the similarities or differences between the S+ and S− stimuli. For instance, the object dynamics acquired during learning may contribute to the activation of novel viewpoints during recognition by enhancing certain novel viewpoints (e.g., Freyd, 1987; Kourtzi & Nakayama, 2002; Vuong & Tarr, 2004).

Another possibility is that visual systems have specialized regions that integrate shape and motion information. There is evidence, for example, that the posterior regions of the superior temporal sulcus in humans and monkeys perform such integration (Giese & Poggio, 2003). Importantly, this region seems to be sensitive to the global coherence of the motion rather than to local motion signals (e.g., they respond to point-light walkers moving coherently, yet not when the points are mixed up but preserve local motion trajectories; Grossman et al., 2000). Therefore, these regions may further facilitate stronger view combination with coherent motion. To our knowledge, there is no corresponding region that integrates shape and motion cues in the pigeon brain, which may help to explain the greater degree of independence of these cues for birds (e.g., Spetch et al., 2006).

5. Conclusions

The kind of motion in the present experiments was rigid motion, and it did not distinguish among the objects in a uniquely characteristic manner, as we did in our previous research (Spetch et al., 2006). In that study, each training object had a unique structure as well as a unique rigid rotation trajectory assigned to it. Often, unique object motion is produced by non-rigid motion, such as facial expressions or other body movements (e.g., Knappmeyer et al., 2003). Thus, future research is needed to explore how humans and birds react when objects are moving non-rigidly. This question is important because non-rigid motion often deforms the 3D shape of an object, yet at least human observers are still capable of recognizing the object. Further research is needed to explore species differences when non-rigid motion is a unique characteristic of the object (e.g., the differences in movement between butterflies and snakes). There is some evidence that birds can recognize non-rigid motion (e.g., point-light displays; Regolin, Tommasi, & Vallortigara, 2000). However, these experiments are not tests of the role of motion in object recognition per se.

Across diverse species, the visual system has adapted to the stimulus information in a dynamic environment. It should therefore not be surprising that both humans and pigeons use motion cues to generalize to novel viewing conditions, such as a novel viewpoint. Perhaps what is surprising is the similarity of the mechanism for processing shape and motion cues in both species (i.e., view combination). Future work is needed to further explore these similarities in primate and avian brains, given the importance of object-recognition mechanisms for survival.

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 invariance. *Journal of Experimental Psychology: Human Perception & Performance, 19*, 1506–1514.

Blough, P. M. (2001). Cognitive strategies and foraging in pigeons. In R. G. Cook (Ed.), Avian visual cognition. <www.pigeon.psy.tufts.edu/avc/pblough>.

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, 89*, 60–64.

Cook, R. G., & Katz, J. S. (1999). Dynamic object perception by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 194–210.

Cook, R. G., & Roberts, S. (2007). The role of video coherence on object-based motion discriminations by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 33*, 287–298.

Cook, R. G., Shaw, R., & Blaisdell, A. P. (2001). Dynamic object perception by pigeons: Discrimination of action in video presentations. *Animal Cognition, 4*, 137–146.

Dittrich, W. H., & Lea, S. E. G. (2001). Motion discrimination and recognition. In R. G. Cook (Ed.), Avian visual cognition. <www.pigeon.psy.tufts.edu/avc/dittrich>.

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 3-D object categorization. *Spatial Vision, 12*, 107–123.

Freyd, J. (1987). Dynamic mental representations. *Psychological Review, 94*, 427–438.

Friedman, A., Waller, D., Hodgson, E., & Greenauer, N. (in press). Learning scenes from multiple views: Novel views can be recognized more efficiently than learned views. Memory & Cognition.

Friedman, A., Spetch, M. L., & Ferrey, A. (2005). Recognition by humans and pigeons of novel views of 3-D objects and their photographs. *Journal of Experimental Psychology: General, 134*, 149–162.

Friedman, A., & Waller, D. (2008). View combination in scene recognition. *Memory & Cognition, 36*, 467–478.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Review Neuroscience, 4*, 179–192.

Grossman, E. D., Donelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., et al. (2000). Brain areas involved in the perception of biological motion. *Journal of Cognitive Neuroscience, 12*, 711–720.

Harman, K., & Humphrey, G. K. (1999). Encoding "regular" and "random" sequences of views of novel three-dimensional objects. *Perception, 28*, 601–615.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

Husband, S. & Shimizu, T. (2001). Evolution of the avian visual system. In R.G. Cook (Ed.), Avian visual cognition. <www.pigeon.psy.tufts.edu/avc/husband>.

Kelly, M. H., & Freyd, J. (1987). Explorations of representational momentum. *Cognitive Psychology, 19*, 369–401.

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research, 43*, 1921–1936.

Kourtzi, Z., & Nakayama, K. (2002). Distinct mechanisms for the representation of moving and static objects. *Visual Cognition, 9*, 248–264.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology, 12*, 259–272.

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint-specific recognition procedures. *Perception, 23*, 595–614.

Liu, T. (2007). Learning sequence of views of three-dimensional objects: The effect of temporal coherence on object memory. *Perception, 36*, 1320–1333.

Liu, T., & Cooper, L. A. (2003). Explicit and implicit memory for rotating objects. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 554–562.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review, 1*, 476–490.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex, 5*, 270–288.

Loidolt, M., Aust, U., Steurer, M., Troje, N., & Huber, L. (2006). Limits of dynamic object perception in pigeons: Dynamic stimulus presentation does not enhance perception and discrimination of complex shape. *Learning & Behavior, 34*, 71–85.

Marr, D., & Nishihara, H. K. (1978). Represenetation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, B, 200*, 269–294.

Mitsumatsu, H., & Yokosawa, K. (2003). Efficient extrapolation of the view with a dynamic and predictive stimulus. *Perception, 32*, 969–983.

Newell, F. N., Wallraven, C., & Huber, S. (2004). The role of characteristic motion in object categorization. *Journal of Vision, 4*, 118–129.

Nguyen, A. P., Spetch, M. L., Crowder, N. A., Winship, I. R., Hurd, P. L., & Wylie, D. R. W. (2004). A dissociation of motion and spatial-pattern vision in the avian

telencephalon: Implications for the evolution of "visual streams". *Journal of Neuroscience, 24*, 4962–4970.

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). A network that learns to recognize three-dimensional objects. *Nature, 343*, 263–266.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*, B51–B61.

Regolin, L., Tommasi, L., & Vallortigara, G. (2000). Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Animal Cognition, 3*, 53–60.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317–1323.

Spetch, M. L., & Friedman, A. (2003). Recognizing rotated views of objects: Interpolation versus generalization by humans and pigeons. *Psychological 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.

Spetch, M., Friedman, A., & Vuong, Q. C. (2006). Dynamic object recognition in pigeons and humans. *Learning & Behavior, 34*, 215–228.

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research, 38*, 947–951.

Stone, J. V. (1999). Object recognition: View-specificity and motion-specificity. *Vision Research, 39*, 4032–4044.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoints? *Psychological Science, 8*, 282–289.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*, 371–387.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition, 67*, 21–44.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: The MIT Press.

Vuong, Q. C., & Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research, 44*, 1717–1730.

Vuong, Q. C., & Tarr, M. J. (2006). Structural similarity and spatiotemporal noise effects on learning dynamic novel objects. *Perception, 35*, 497–510.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, 98*, 4800–4804.

Wong, A. C.-N., & Hayward, W. G. (2005). Constraints on view combination: Effects of self-occlusion and differences among familiar and novel views. *Journal of Experimental Psychology: Human Perception & Performance, 31*, 110–121.

Ziegler, H. P., & Bischof, H. J. (1993). *Vision, brain, and behaviour in birds*. Cambridge, MA: MIT Press.