Limits of dynamic object perception in pigeons: Dynamic stimulus presentation does not enhance perception and discrimination of complex shape

MICHAELA LOIDOLT, ULRIKE AUST, and MICHAEL STEURER
University of Vienna, Vienna, Austria

NIKOLAUS F. TROJE
Queens University, Kingston, Ontario, Canada

and

LUDWIG HUBER
University of Vienna, Vienna, Austria

A go/no-go procedure was used to train pigeons to discriminate pictures of human faces differing only in shape, with either static images or movies of human faces dynamically rotating in depth. On the basis of experimental findings in humans and some earlier studies on three-dimensional object perception in pigeons, we expected dynamic stimulus presentation to support the pigeon's perception of the complex morphology of a human face. However, the performance of the subjects presented with movies was either worse than (AVI format movies) or did not differ from (uncompressed dynamic presentation) that of the subjects trained with a single or with multiple static images of the faces. Furthermore, generalization tests to other presentation conditions and to novel static views revealed no promoting effect of dynamic training. Except for the subjects trained on multiple static views, performance dropped to chance level with views outside the training range. These results are in contrast to some prior reports from the literature, since they suggest that pigeons, unlike humans, have difficulty using the additional structural information provided by the dynamic presentation and integrating the multiple views into a three-dimensional object.

Most of the objects that have adaptive value for an organism, whether as food, prey, enemies, or social partners, are normally seen in motion (Dittrich & Lea, 2001; Dittrich, Lea, Barrett, & Gurr, 1998), due to both movements of the perceived objects and self-induced movements of the perceiving subject. Therefore, motion cues should play a prominent role in successful discrimination, recognition, and categorization of objects in almost all animals but should be of particular importance to highly mobile ones, such as birds and primates. Regarding pigeons, there is evidence that these birds are able to discriminate the velocity of visual stimuli (Hodos, Smith, & Bonbright, 1975; Mulvanny, 1978; Siegel, 1970), to track moving targets (Pisacreta, 1982), and even to extrapolate the movement of a stimulus that is no longer visible (Neiworth & Rilling, 1987). They are furthermore shown to be able to discriminate complex motions, such as Lissajous figures

produced by compound sine-wave functions on an oscilloscope (Emmerton, 1986). Martinoya and Delius (1990) found evidence for the perception of optic flow in pigeons. In several experiments in which they were presented with video images of moving conspecifics, these birds not only detected the presence of motion as such, but also discriminated different types of movement (Dittrich & Lea, 1993; Dittrich et al., 1998; Jitsumori, Natori, & Okuyama, 1999; Shimizu, 1998). However, it is far from clear whether any of those tasks actually requires discrimination of a coherently moving three-dimensional (3-D) object.

The issue of whether and how complex stimuli are viewed as stable 3-D objects by pigeons has become a topic of interest during the last few years. Early investigations on this subject were restricted to static stimuli. Planar rotations were shown to pose no problem to the pigeons' recognition ability (Hollard & Delius, 1982), whereas rotations in depth seem to be more problematic, because different features may be gained or lost as they come into or move out of view. Accordingly, Cerella (1990) reported that pigeons consistently failed to generalize over depth rotations of drawings of cube-like objects. He therefore concluded that pigeons solve image classification problems by means of a simple mechanism of local feature detection, without recovery of the 3-D object. In contrast,

Wasserman et al. (1996) found good evidence for depth-invariant visual performance in pigeons, using drawings of man-made artifacts, such as airplanes and flashlights.

A few years ago, Cook and Katz (1999) examined how pigeons process motion cues by training them to discriminate between moving projections of a cube and a pyramid. They found a dynamic superiority effect, with dynamic conditions always leading to better discrimination performance than did static conditions. Furthermore, the discrimination was found to be invariant across variations in stimulus size, rate of motion, combinations of different motions and rates, direction of motion, and changes in surface color. Importantly, the pigeons also showed above-chance (although not perfect) transfer when tested with a new axis of rotation. Those results led the authors to argue that pigeons discriminated the dynamically presented projections of cube and pyramid objects by experiencing them as invariant, coherent 3-D representations, rather than as collections of flat two-dimensional (2-D) features or templates. They argued that the dynamic change of perspective provided additional information not available in static views or helped integrate the views into an object-like 3-D image. In the human literature, this is a key idea of many theories of object perception, known as *structure from motion* (e.g., Cutting, 1986; Gibson, 1979; Johansson, 1975; Ullman, 1979; Wallach & O'Connell, 1953).

We think, however, that the achievement of object constancy must be examined in relation to the difficulty of the discrimination task as experienced by the subjects. This caveat has rarely received the necessary attention in the animal literature on object recognition. For instance, if we simply have to distinguish an airplane from a chair (Wasserman et al., 1996) or a cube from a pyramid (Cook & Katz, 1999), conspicuous contour differences, such as the legs of the chair, the wings of the airplane, or the pointed top of the pyramid, might be sufficient to determine the correct response, irrespective of the views presented. There is no need to acquire a complex structural representation of those objects if the same simple objects are presented in recurrent encounters. Detection of simple 2-D object-distinguishing features or the storage of 2-D templates in memory is likely to be sufficient. In more complex tasks, the visual system must discriminate between stimuli that are similar in their appearance. The conflict for the visual system is particularly evident if variation across viewpoints, as a consequence of object or observer motion, is larger than variation across object classes.

In general, animate objects may be harder to recognize than man-made artifacts because, as a category, animate objects are visually more similar to each other than are artifacts (Humphreys, Riddoch, & Quinlan, 1988), although it certainly depends on which objects or artifacts are being compared. In most previous studies on perspective invariance, only simple forms (Cerella, 1977, 1990; Cook & Katz, 1999; Lumsden, 1970, 1977) or line drawings of man-made artifacts, such as airplanes and flashlights (Wasserman et al., 1996) or wedges, arches, and cones (Peissig, Wasserman, Young, & Biederman, 2002; Peissig, Young, Wasserman, & Biederman, 1999, 2000; Spetch &

Friedman, 2003; Spetch, Friedman, & Reid, 2001; Spetch, Kelly, & Reid, 1999), have been used as stimuli. Only recently, Jitsumori and Makino (2004) reported an experiment on recognition of depth-rotated human faces in pigeons in which they first used complex natural stimuli in order to examine to what extent pigeons show rotation generalization and whether seeing the objects dynamically rotating in depth promotes perspective invariance. Such stimuli not only contain sufficiently complex shape information—constituting so-called *morphology* (Hill, Schyns, & Akamatsu, 1997; Troje & Bülthoff, 1996)—but also form an object class that has probably undergone more investigation than any other object class (Troje & Bülthoff, 1998). Moreover, two different views of the same face are less similar to each other in absolute terms than are two faces seen from the same viewing angle (Moses, 1993; Moses, Ullman, & Edelman, 1996). However, Moses et al. did not find that dynamic presentation increased the degree of viewpoint independence (Experiment 2), although they reported substantial generalization to novel viewpoints in general (Experiments 1 and 2).

In previous studies, we have used static images of human faces to investigate the pigeon's categorization strategies in complex natural categorization tasks (Huber, Troje, Loidolt, Aust, & Grass, 2000; Loidolt, Aust, Meran, & Huber, 2003; Troje, Huber, Loidolt, Aust, & Fieder, 1999). There, the pigeons successfully learned to sort images of 200 faces into categories of male and female. Using photorealistic images of laser-scanned models of human faces, we investigated which stimulus aspects the birds employed. Vetter and Troje (1997) had developed a representation of human faces that permitted separation of the information contained in the images of human faces by addressing the spatial arrangements of stimulus features, on the one hand—the authors used the term *shape* to address these features—and their particular appearance, on the other hand, the latter subsumed under the term *texture*. In Troje et al.'s study, three groups of pigeons were confronted with different versions of the same human face stimuli, differing with respect to the amount of shape and texture information they contained. The pigeons showed a strong preference for visual texture and color information over facial shape information. Only the pigeons that were presented with texture information (with or without additional shape information) successfully sorted the faces into male and female. This was interesting, since shape contains more information useful for classification by sex than does texture, as has been shown by training artificial neural networks, using different types of information (Troje & Vetter, 1998). The few pigeons that managed to discriminate between males and females on the basis of shape information alone used the size of the head as the major cue governing classification behavior.

One reason for the pigeons' failure to exploit shape information for classification might have been the absence of motion in the static images we presented. Pigeons might be able to use complex shape information under natural conditions but may have difficulties with static 2-D pictures. Research to date suggests that smooth mo-

tion of a head does provide beneficial information for the human viewer. Pike, Kemp, Towell, and Phillips (1997) have shown that dynamic presentation during learning improves recognition of faces. They suggested that this advantage is evoked by the enhanced possibility of forming an object-based 3-D structural representation of the face. The additional structural information provided by the smooth transformational motion of the head offers more opportunities for encoding this information. If pigeons are also capable of forming 3-D structural representations of complex objects, it is conceivable that they may also benefit from dynamic presentation.

This hypothesis was tested in the present study. We used images of human faces as stimuli to investigate the possible effects of dynamic versus static stimulus presentation on the discrimination of complex natural objects. However, in contrast to Jitsumori and Makino (2004), we presented human face stimuli displaying only shape information, in order to constrain the pigeons' attention to configural information. The stimuli were created using the 3-D surface models of laser-scanned human faces, with the hair region of the head being digitally removed (Figure 1).

We examined the discrimination performance of four subject groups trained by means of different presentation modes (Part 1). First, we compared discrimination performance with static presentation of only a single picture of each face (Group Static Single) and performance with motion sequences of the faces with audio video interleave (AVI) format (Group Dynamic AVI). AVI is a video format used to present videos or movies with Windows-based computer systems. Another group of pigeons was then trained with dynamic sequences that were created without using a compression algorithm (Group Dynamic Uncompressed). Finally, a fourth group was presented with static
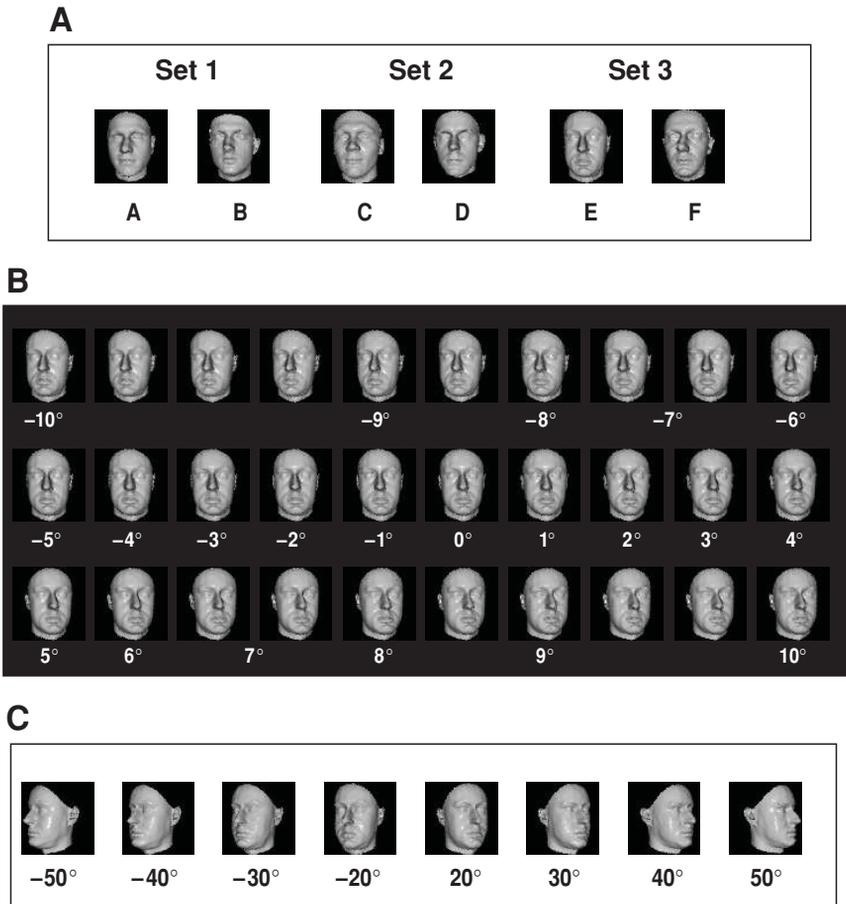
Figure 1. Images of human faces containing only shape information. (A) Six male faces allocated to three stimulus sets. Stimulus Set 1 was used in Phase 1, Stimulus Set 2 was used in Phase 2, and Stimulus Set 3 was used in Phase 3 of the discrimination training and in the generalization tests. The faces are shown under a viewing angle of $-7°$ relative to the frontal view as presented in the static single-presentation condition. (B) Dynamic stimulus presentation: The stimuli were sequences of the faces rotating (between $-10°$ and $+10°$) around the vertical axis at a rate of 30 pictures per second. All 30 image views of Face E are depicted. These views were presented one by one in the static multiple-presentation condition. (C) Test stimuli presented in Test 2: Transfer to novel angular views. Depicted are test images of Face E.

views of the rotated faces in random sequences at the rate of one per trial (Group Static Multiple). This presentation mode allowed us to examine differences between dynamic and static presentations unbiased by any difference in the amount of information available.

The successful birds were then subjected to two series of generalization tests (Part 2). First, we tested for transfer to novel presentation conditions (Test 1). In particular, pigeons that had been trained on the static views were tested with dynamic sequences, and vice versa. Furthermore, we tested the group trained with the uncompressed movies with movies composed of the views used in training, but shown in randomized order. For humans, such a presentation mode destroys the impression of smooth motion and, instead, leads to the perception of an unnatural jerky motion of the heads. Significant performance decrements in the pigeons would support the notion of perception of smooth motion.

Second, we tested the pigeons' recognition of faces rotated up to 40º away from the training view(s) (Test 2). This aspect of the study was consistent with the focus of most current theories of object recognition in the literature on humans that consider recognition across viewpoints as the diagnostic case by which to decide between object-centered, viewpoint-invariant and viewer-centered, viewpoint-dependent recognition mechanisms. Whereas object-centered theories typically expect a rather small decrease in recognition performance with novel views of an object (e.g., Biederman, 1987; Corballis, 1988; Marr & Nishihara, 1978), viewer-centered models predict a large, continuous decrease in performance across changes in viewpoint (Peissig et al., 2000; Tarr & Bülthoff, 1998; Tarr & Kriegman, 2001). Although the present experiments were not designed to decide between the two theories, they might provide some information about whether or not pigeons perceive and respond to complex visual stimuli presented on a computer monitor as if they were 3-D objects.

METHOD

Subjects

We employed 24 pigeons (*Columba livia*) of mixed breed (*C. livia forma urbana* and *C. livia forma domestica*, a traditional livestock breed called *Strasser*) as experimental subjects. Most of the subjects were experimentally naive at the onset of the experiment, but some had previously been used in unrelated experiments. They were housed in several compartments of a large outdoor aviary (54 m³) in groups of 10–15 individuals of mixed sex. During the experiment, the birds were maintained at weights slightly below their free-feeding weights. Water and grit were freely available in the aviary, but food was provided only in the experimental chambers during and shortly after the daily experimental sessions.

Apparatus

The experiment was conducted in four identical one-key chambers located in an indoor lab. Each chamber was connected through a passageway system to an outdoor aviary (see Huber, 1994). The front panel of each $50 \times 30 \times 40$ cm wooden chamber was equipped with an oversized, clear Perspex pecking key (5-cm diameter; ENV-125M, MED Associates) mounted in the center of the panel, 28 cm above the floor of the box. The grain feeder (ENV-205M) used a 28-V DC solenoid-activated hopper. A 6×6 cm aperture for the food hopper

was located directly below the pecking key (8.5 cm above the floor). A hopper light illuminated the aperture of the hopper whenever food was accessible. During the experimental sessions, the chamber was weakly illuminated by a 2-W houselight (ENV-215) located in the rear of the chamber. The stimuli were presented on 15-in. (38-cm diagonal) CRT computer monitors (Panasonic PanaSync 4G) at a distance of 5 cm behind the pecking key. Monitor adjustment was set at true color and a $1{,}024 \times 768$ pixel spatial resolution. Presentation of images or movies, control of experimental procedures, and data collection were managed by four PCs equipped with a PCI-Bus Parallel Digital I/O Board (Keithley, KPCI-PIO24), running the Windows 2000 operating system, one for each chamber.

Stimuli

The stimuli were created using six surface models of male faces from a database of 3-D head models (Figure 1A). The models were derived from laser-scanned heads that were shaved and free of any kind of accessories, such as glasses or earrings (for details, see Troje & Bülthoff, 1996). Before the images were generated, the hair region of the heads was digitally removed from the 3-D models. The ears, however, remained visible. The face images were rendered in perspective projection, as if recorded by a virtual camera at a distance of 2 m from the head. The faces were rendered with constant albedo, instead of the original surface reflection information. We applied an illumination model assuming Lambertian reflection, as well as some degree of specular reflection, to the faces. The illumination consisted of directional light from a single light source 30º above the virtual camera and a small amount of additional ambient light (Figure 1).

From each head model, 30 images were derived that differed from each other only in the viewing angle, which ranged from $-10º$ to $+10º$ (with 0º representing a full face frontal view; see Figure 1B). Within this range, the views were sampled so that if played back as a movie, the rotation would be faster at frontal views and would slow down at the beginning and the end to generate a natural smooth motion.

Images were presented as 64×64 pixel grayscale bitmaps on a dark background. Their size on the monitor was about 2×2 cm. Displays were shown either as static images (Figure 1A) or as *rotation* animations (Figure 1B). By presenting the 30 views in quick succession, we created movies that, to the human eye, appeared as heads smoothly rotating around the vertical axis. Starting at a viewing angle of $-10º$ relative to the frontal view, "rotation" proceeded to an angle of $+10º$ on the other side of the face, thereby showing only one eighteenth of a full rotation, and back again. This animation was repeated until the end of a trial.

We used two presentation types for the rotation animations. In both of them, the face was animated with 30 fps, so that a complete cycle of forward and backward motion took 2 sec. The first was AVI format movies with videocompression, using an Indeo Video 5 Codec. It is the video format most commonly used with Windows-based (Microsoft Inc., Redmond, WA) computer systems. In AVI format movies, the large amount of information inherent in video data is usually compressed for storage. The impression of smooth movement is generated in a human perceiver by using specific compression/decompression algorithms—the Codec. There is no picture-by-picture presentation, as in a cinematic movie.

The second presentation type for the rotation animations was created without using a compression algorithm. It contained successive presentation of uncompressed images (bitmap files) of all 30 views of a head in fast succession, shown at a rate of 30 per second (uncompressed presentation; *UCP movies*). In addition to the 30 frames required for the movies, single images of the faces were rendered at $-50º$, $-40º$, $-30º$, $-20º$, 20º, 30º, 40º, and 50º, to be used as test stimuli in Test 2.

Procedure

Before being exposed to discrimination training, the naive birds were trained to enter the chamber voluntarily and to find food in the

grain hopper and were then autoshaped by standard methods to peck at the illuminated key (for details, see Huber, 1994). We used a standard successive discrimination paradigm, the go/no-go procedure as described in Vaughan and Greene (1984). It requires the birds to peck frequently in the presence of positive stimuli and to refrain from pecking in the presence of negative stimuli. Food reward is delivered only on positive trials.

During discrimination training, as well as generalization testing, sessions consisted of 40 or 60 trials. Each trial started with the presentation of a stimulus. During the first 10 sec of a trial, pecks emitted to the response key were recorded for later assessment of performance. The stimulus remained visible during a further variable interval period (mean, 10 sec; range, 1–20 sec). This variable interval was included to prevent the pigeons from being conditioned on a fixed time schedule. It was followed by a decision phase, during which pecks were recorded again—not for data analysis, but in order to determine whether the response requirement was met. If the stimulus was positive, two responses emitted within a period of 2 sec resulted in 5 sec of food access. Responding to a negative stimulus during the decision phase initiated a delay interval, which was terminated only after 8 sec had passed without any further responses occurring. During generalization testing, test trials were conducted in extinction, and they ended immediately after the first 10 sec of stimulus presentation without food reward. Each trial was followed by a 4-sec intertrial interval, which ended with a short dark phase signaling the presentation of the next stimulus. The pigeons were trained in daily sessions, 5 days a week.

Part 1: Discrimination of human faces under different stimulus presentation conditions. We compared different static and dynamic presentation modes for their effects on discrimination performance. In particular, we examined static presentation of a single view, static presentation of multiple views, and two dynamic presentation modes in three successive phases. In each phase, the subjects were confronted with a different pair of male faces as stimuli (Stimulus Sets 1–3). One face stimulus from each set was arbitrarily defined as positive; the other as negative (Figure 1A). Table 1 gives an overview of the sequence and the presentation modes examined in each phase.

In Phase 1, we compared discrimination performance with static presentation of only a single view of each face (Group Static Single; $n = 5$) and performance with motion sequences of the faces with AVI format movies (Group Dynamic AVI; $n = 5$). Both groups were trained with the faces of Stimulus Set 1 (Faces A and B; Figure 1A), with Face A being the positive and Face B being the negative stimulus for all the subjects. For the static presentation, we chose the $-7°$ view of the faces, which we assumed would provide more 3-D information than would the 0° frontal view. For both groups, each daily training session consisted of 40 trials, involving 20 presentations of the positive and 20 presentations of the negative face.

Assessment of performance during training was based on the ρ (rho) discrimination measure introduced by Herrnstein, Loveland, and Cable (1976), which is derived from the Mann–Whitney U statistic. The birds were trained until they reached a prespecified criterion for the discrimination (two ρ values ≥ 0.8 on 3 successive sessions). This criterion requires discrimination performance far above

the significance level of $\rho \geq 0.68$ ($p \leq .05$, for a 40-trial session). If a subject did not reach the critical discrimination value of $\rho \geq 0.8$ in at least 1 session, training was terminated after 30 sessions. If a subject reached the critical value on at least 1 session, training was continued for another 30 sessions or until criterion was met.

In Phase 2, we introduced a further experimental group (Group Dynamic Uncompressed), consisting of 7 subjects. Group Dynamic Uncompressed was trained with uncompressed motion sequences of the human faces. For reasons of comparability, the subjects of all three groups were now presented with a new set of faces (Stimulus Set 2), where the positive stimulus was Face C and the negative stimulus was Face D (Figure 1A). Training for Group Dynamic Uncompressed was similar to that for the other groups. Sessions consisted of 40 trials and involved 20 presentations of each face. Assessment of performance during training followed exactly the same scheme as that in Phase 1.

In Phase 3, we introduced a fourth presentation group (Group Static Multiple) consisting of 7 subjects, whereas Group Dynamic AVI was excluded from the experiment. All the subjects were presented with a new set of faces (Stimulus Set 3), with Face E being the positive and Face F being the negative stimulus (Figure 1A). The subjects from Group Static Multiple were trained with all 30 of the slightly different views of the faces that were used to create the movies, with each being presented as a static image, 1 per trial. Since there were 30 images of each of the two faces (the positive and the negative stimulus), training sessions for Group Static Multiple consisted of 60 trials. Distribution of the image views during a session was random. Assessment of performance during training followed exactly the same scheme as that in Phases 1 and 2. The birds were trained until they reached the prespecified discrimination criterion (two ρ values ≥ 0.8 within three successive sessions). The significance level for a 60-trial session was at $\rho \geq 0.64$ ($p \leq .05$).

Part 2: Generalization tests. The successful birds from Phase 3 (4 pigeons from Group Static Single, 3 pigeons from Group Dynamic Uncompressed, and 3 pigeons from Group Static Multiple) were subjected to two series of generalization tests. Unfortunately, 1 successful subject from Group Static Single died shortly after the onset of the test series, so only 3 pigeons from this group completed the tests. The series of generalization tests involved transfer to (1) presentation conditions other than that encountered during training and (2) novel angular views. The total number of sessions performed in each test depended on the number of testing conditions. During a test session, 10 test trials were randomly interspersed among 30 training trials (for Groups Static Single and Dynamic Uncompressed) or 50 training trials (for Group Static Multiple). Test stimuli were presented in extinction. All the test stimuli were derived from Face E and Face F in Stimulus Set 3. Each face was presented five times in every testing condition.

Test 1: Transfer to other presentation conditions. (a) *Transfer from dynamic presentation to static views*. The subjects from Group Dynamic Uncompressed were tested on static presentations of the faces and familiar viewing angles—that is, angles that were within the training range ($-10°$, $-7°$, $-1°$, $+6°$, and $+10°$). Test 1a involved 5 sessions. (b) *Transfer from dynamic presentation to other dynamic presentation modes*. The subjects from Group Dynamic

Table 1
Design of Part 1: Discrimination of Human Faces Under Different Stimulus Presentation Conditions

Phase	Stimuli	Presentation Conditions (Groups)		
1	Set 1: Face A and Face B	Static single: Static presentation of 1 image of each face ($-7°$ view)		Dynamic AVI: AVI-format movies of the faces rotating in the range between $-10°$ and $+10°$
2	Set 2: Face C and Face D	Static single	Dynamic AVI	Dynamic uncompressed: Dynamic sequences of the faces composed of 30 images showing views between $-10°$ and $+10°$ presented in rapid succession
3	Set 3: Face E and Face F	Static single	Dynamic uncompressed	Static multiple: Static presentation of 30 images of each face representing views between $-10°$ and $+10°$, one image per trial

Uncompressed were tested for transfer to AVI movies, as well as to uncompressed movies with random sequences of images. In the latter condition, the 30 views of each head were presented in random order at a rate of 30 per second. Test 1b involved two sessions. (c) *Transfer from static presentation to dynamic presentation*. The subjects from Groups Static Single and Static Multiple were tested with the AVI movies and with the uncompressed movies containing regular sequences of images. Test 1c also involved two sessions.

Test 2: Transfer to novel angular views. The subjects of all three groups (Static Single, Dynamic Uncompressed, and Static Multiple) were tested with static presentations of the faces with viewing angles beyond the training ranges. Tested viewing angles were $-50°$, $-40°$, $-30°$, $-20°$, $+20°$, $+30°$, $+40°$, and $+50°$ (Figure 1C). This involved eight test sessions for the subjects of all three groups. In addition, Group Static Single, which had been trained on the $-7°$ view only, was presented with test stimuli showing the faces at viewing angles closer to the training view, at $-10°$, $-4°$, $0°$, $+4°$, $+7°$, and $+10°$. So there were another six test sessions appended only for this group.

Data analysis was based on normalized response rates (NRRs). They were obtained by dividing the pecking rate in response to each stimulus in a session (which was registered automatically during the first 10 sec of stimulus presentation) by the average peck rate measured over all the trials (except test trials) of the respective session. To assess discrimination performance in the various presentation conditions, we used the discrimination ratio (DR), calculated by dividing the sum of (normalized) pecks to all presentations of the positive stimulus face presented in a particular condition by the total sum of (normalized) pecks to presentations of both the positive and the negative faces measured in the respective condition. Because training stimuli were shown repeatedly during the whole test series, whereas test stimuli were presented only five times in each condition, calculations of the training component were based on normalized pecking rates emitted to five randomly chosen training presentations of each face in the respective test. Group performances

were given as mean DRs (for all the subjects of the respective group) $\pm$ standard deviations from the mean. To evaluate performances in the various test conditions, we calculated paired t tests comparing the average pecking rates to positive and negative stimuli for each subject. We used an alpha level of .05 throughout the analyses.

RESULTS

Part 1: Discrimination of Human Faces Under Different Stimulus Presentation Conditions

Figure 2 compares performances of the different presentation groups during the three successive phases of discrimination training. It shows the percentage of subjects that succeeded in learning the discrimination (left), as well as the median number of sessions these successful birds required to reach the learning criterion (right).

Phase 1: Static presentation of a single view versus dynamic presentation with AVI movies. Four of the 5 subjects from Group Static Single learned to discriminate the two faces of Stimulus Set 1. They required between 14 and 20 sessions to reach the learning criterion (median = 16.5). Interestingly, dynamic presentation clearly failed to facilitate acquisition of the discrimination. None of the 5 pigeons in Group Dynamic AVI learned the discrimination. Since no subject reached the critical rho value of $\rho \geq 0.8$, training was terminated for all the birds after 30 sessions.

The pigeons' failure to discriminate the two heads in the dynamic presentation condition—although the same faces could be discriminated when presented statically—was rather surprising, given previous reports that have yielded
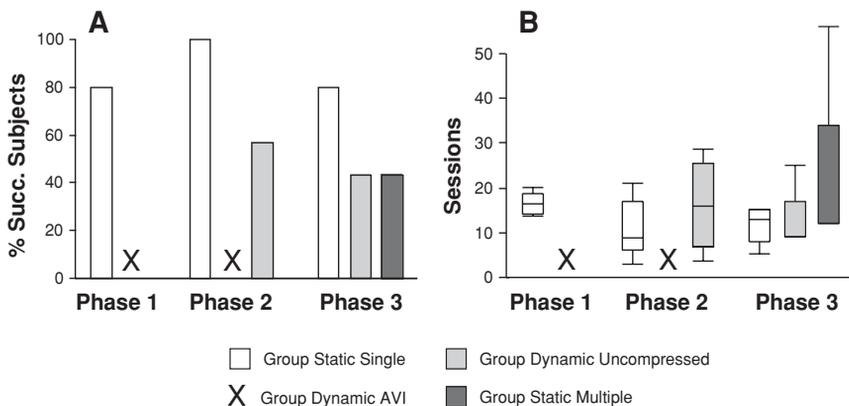
Figure 2. Performance of the four groups of pigeons discriminating between two human faces under different stimulus presentation conditions (Part 1). (A) Percentage of subjects in each group that successfully learned to discriminate between the positive and the negative stimuli. Numbers of subjects in each group were $n = 5$ in Group Static Single, $n = 5$ in Group Dynamic AVI, $n = 7$ in Group Dynamic Uncompressed, and $n = 7$ in Group Static Multiple. (B) Boxplots showing the number of sessions the successful birds in each group required for reaching the learning criterion of two rho values ≥ 0.8 within three consecutive sessions. The boxplots represent the median value (horizontal line across the box), as well as the minimum and maximum number of sessions needed within a group. The number of successful subjects ($n_{succ.}$) contributing to the calculation of each boxplot varied. In Phase 1, $n_{succ.} = 4$ for Group Static Single, and $n_{succ.} = 0$ for Group Dynamic AVI; in Phase 2, $n_{succ.} = 5$ for Group Static Single, $n_{succ.} = 0$ for Group Dynamic AVI, and $n_{succ.} = 4$ for Group Dynamic Uncompressed; in Phase 3, $n_{succ.} = 4$ for Group Static Single, $n_{succ.} = 3$ for Group Dynamic Uncompressed, and $n_{succ.} = 3$ for Group Static Multiple.

evidence to the contrary, with pigeons being successfully trained to discriminate dynamic objects (e.g., Cook & Katz, 1999; Dittrich & Lea, 1993, 2001; Dittrich et al., 1998; Jitsumori & Makino, 2004; Jitsumori et al., 1999). We suspected that our pigeons' failure to discriminate the dynamic stimuli could have been specifically related to AVI movie presentation. To validate this assumption, in Phase 2 of the experiment, we trained another group of pigeons with dynamic sequences of human faces that were created without using a compression algorithm.

Phase 2: Dynamic presentation with uncompressed movies. Acquisition of the discrimination and learning speed of Group Dynamic Uncompressed were examined in comparison with Groups Static Single and Dynamic AVI. All 5 subjects from Group Static Single learned to discriminate the new faces of Set 2. They required between 3 and 21 sessions (median = 9) to reach the learning criterion. Again, all the birds from Group Dynamic AVI failed to reach the critical rho value of $\rho \geq 0.8$ at least once within 30 sessions. However, Group Dynamic Uncompressed performed better. Four out of 7 subjects managed to learn the discrimination. They needed between 4 and 29 sessions to reach criterion (median = 16; see Figure 2).

Although acquisition of the discrimination was better with uncompressed movies than with AVI movies, we found no advantage of dynamic over static presentation. This result did not support our expectation that smooth motion of the heads may facilitate learning by providing additional structural information, an effect reported by Cook and Katz (1999) for pigeons discriminating moving projections of a cube and a pyramid. On the other hand, Jitsumori and Makino (2004), who also trained pigeons to discriminate human faces, likewise found that dynamic presentation of the faces did not make the discrimination easier for the pigeons, in comparison with a single static presentation of only one view of each face. However, their data cannot be directly compared with ours, due to several differences in the experimental designs of the two studies. For example, their pigeons were trained on four faces (two positive, two negative) simultaneously, and rotation of the faces to the left and to the right ranged from −67° to +67°.

Group Dynamic AVI's repeated failure to acquire the discrimination indicated that it was indeed the specific AVI presentation technique that gave rise to the pigeons' problems in discriminating the faces. Although the AVI movie generated the impression of a smooth movement in human observers despite information compression, pigeons may be more sensitive to unnatural deformations caused by compression (see the Discussion section for a more detailed discussion on this topic). Therefore, we terminated training with AVI movies and restricted training in Phase 3 to uncompressed dynamic presentation.

Phase 3: Multiple static presentation. Acquisition of the discrimination and learning speed of Group Static Multiple were examined in comparison with Groups Static Single and Dynamic Uncompressed. All the groups were presented with the two faces of Stimulus Set 3.

Four of the 5 subjects from Group Static Single learned the discrimination. It took them between 5 and 15 sessions (median = 13). In Group Dynamic Uncompressed, 3 out of 7 subjects learned the discrimination. Two of them reached the criterion after 9 sessions, and 1 subject needed 25 sessions (median = 9). Also, in Group Static Multiple, 3 out of 7 subjects learned to discriminate the two faces, with 2 birds reaching criterion after 12 sessions. One bird reached the critical rho value of $\rho \geq 0.8$ only once (in Session 23), so training was continued beyond 30 sessions until the criterion of two rho values ≥ 0.8 was met after 56 sessions (median = 12; see Figure 2).

Although we cannot draw strong conclusions from these results because group sizes were rather small, there are nevertheless some interesting trends apparent. Static presentation of a single view of the faces seems to be the easiest type of presentation. Dynamic presentation of the faces rotating in rapid succession between −10° and +10° seems to be as challenging to the pigeons as multiple static presentation of the same 30 facial views. There was no conspicuous difference in learning speed among the successful birds in the three groups, with the exception of 1 subject in Group Static Multiple that needed 56 training sessions until it showed stable discrimination performance. Overall, learning was as fast for Group Static Multiple as it was for Groups Static Single and Dynamic Uncompressed.

Of course, it may be argued that these results are not directly comparable, because Group Static Multiple received more training trials (60 per session) than did the other two groups, Static Single and Dynamic Uncompressed (40 per session). But it has to be considered that the subjects from Group Static Multiple were presented with 30 different views of each face (Faces E and F) within one session. The subjects from Group Static Single saw 1 single view of each face 20 times per session, and the subjects from Group Dynamic Uncompressed saw all 30 views of the two faces repeatedly within a single trial. Therefore, the subjects from the two latter groups had access to the total information provided by their respective presentation conditions, one might say, more often than did the subjects from Group Static Multiple. Presenting the birds with more trials per session than were received by the subjects of the two other groups may (at least in part) have compensated for that disadvantage.

Part 2: Generalization Tests

Test 1: Transfer to other presentation conditions. Figure 3 shows mean discrimination ratios for the presentation of the faces in the various generalization conditions in comparison with the training condition (left), as well as mean normalized response rates to positive and negative stimuli (right) for Groups Dynamic Uncompressed, Static Single, and Static Multiple.

(a) *Transfer from dynamic presentation to static views*. The subjects from Group Dynamic Uncompressed were tested for generalization to static presentation of the faces with familiar viewing angles within the range covered by the training movies. The five tested angles were −10°,

Group Dynamic Uncompressed

Group Static Single

Group Static Multiple
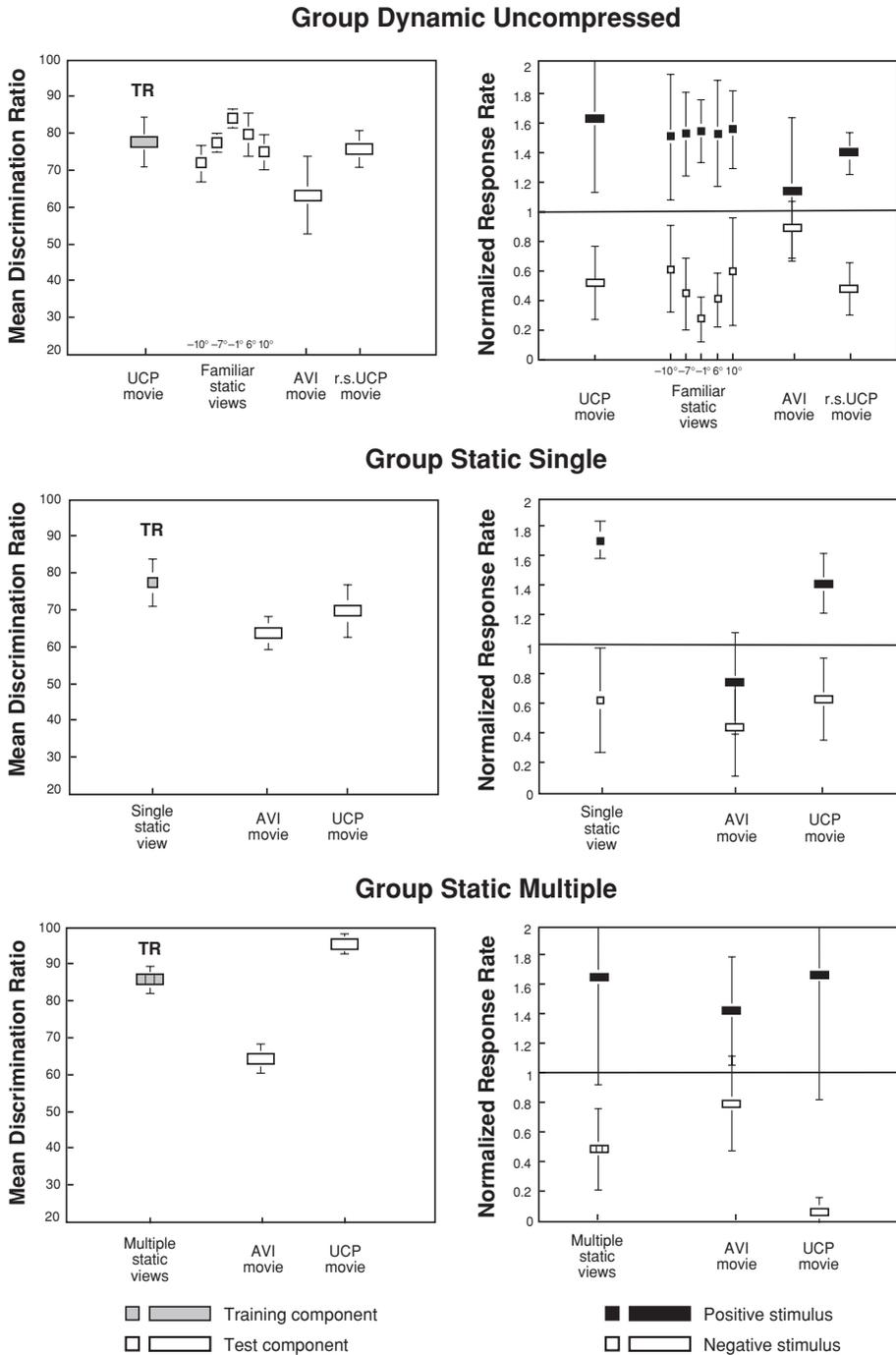
Figure 3. Transfer to other presentation conditions. Left: Mean discrimination ratios ± standard deviations for training and test stimuli for Groups Dynamic Uncompressed, Static Single, and Static Multiple. The symbol TR indicates performance on the training component. Right: Mean normalized response rates ± standard deviations to the positive (filled symbols) and negative (open symbols) stimuli. Squares symbolize static presentation of a single view, blank rectangles symbolize dynamic presentation, and rectangles with lines inside symbolize static presentation of multiple views. Transfer conditions were static presentations of familiar angular views within the training range (Group Dynamic Uncompressed; tested viewing angles were $-10°$, $-7°$, $-1°$, $+6°$, and $+10°$), AVI movies, uncompressed movies with regular sequences of images (UCP movie; Groups Static Single and Static Multiple), and random presentation sequence of the faces (r.s.UCP movie; Group Dynamic Uncompressed).

$-7°$, $-1°$, $+6°$, and $+10°$. The average discrimination performance with the test stimuli (mean DR = 77.3 ± 5.1) was as high as performance with the dynamic training stimuli (DR = 75.7 ± 8.7). Paired t tests comparing the average response rates emitted to positive and negative test faces for each subject presented under the same viewing angles were significant [all $ts(2) \geq 5.05$, $p < .05$] for the $-7°$ (DR = 77.2 ± 2.8), the $-1°$ (DR = 84.4 ± 2.7), the $+6°$ (DR = 79.5 ± 6.4), and the $+10°$ (DR = 74.3 ± 5.1) views, but not for the $-10°$ view [DR = 70.9 ± 5.3; $t(2) = 3.8$, n.s.]. Performance with the static test images varied with best discrimination on presentation angles close to the frontal view and decreasing performance on the angles further away (Figure 3, left). The average response rates emitted to the static test images were quite similar to the response rates emitted to the training movies (for the positive test stimuli, mean NRR = 1.51 ± 0.42, and for the positive training stimuli, NRR = 1.61 ± 0.49; for the negative test stimuli, mean NRR = 0.46 ± 0.36, and for the negative training stimuli, NRR = 0.52 ± 0.24). But although responding to the positive test images was rather constant over all the tested views (range of NRR, 1.49–1.54), responding to the negative test images varied substantially (range of NRR, 0.27–0.61). Obviously, only the individual images that constituted the negative movie were of different salience to the pigeons (depending on viewing angle), but not the ones that constituted the positive movie.

(b) *Transfer from dynamic presentation to other dynamic presentation modes*. The subjects from Group Dynamic Uncompressed were tested for transfer to AVI movies and to movies with the 30 views of each face presented in randomized order. The pigeons generalized to the randomized uncompressed movies without any drop in performance [DR = 74.8 ± 8.8; $t(2) = 5.37$, $p < .05$; for the training stimuli, DR = 76.1 ± 11.3]. There was only a slight decrease of responses emitted to the positive randomized uncompressed movie (NRR = 1.38 ± 0.14; for the positive nonrandomized training movie, NRR = 1.61 ± 0.49; for the negative randomized test movie, NRR = 0.47 ± 0.17; for the negative training movie, NRR = 0.52 ± 0.24). On the other hand, discrimination decreased substantially, with the AVI movies displaying the regular sequence of images [DR = 62.1 ± 19.7; $t(2) = 0.95$, n.s.]. This failure was the result of both a decrease of responses emitted to the positive AVI test movies (NRR = 1.14 ± 0.48) and an increase of responses emitted to the negative AVI movies (NRR = 0.87 ± 0.19; the training values are the same as before).

(c) *Transfer from static presentation to dynamic presentation (different presentation modes)*. The subjects in the two static groups (Static Single and Static Multiple) were tested for transfer to both AVI movies and uncompressed movies. In both groups, there was a substantial drop in performance with the AVI movies (Group Static Single, DR = 62.6 ± 9.4, and for the training stimuli, DR = 76.1 ± 12.7; Group Static Multiple, DR = 63.5 ± 8.8, and for the training stimuli, DR = 86.1 ± 8.8) to a nonsignificant level [both $ts(2) < 2.37$, n.s.]. In Group Static Single, this failure was due to a substantial decrease of responses to the positive AVI test movies (NRR = 0.73 ± 0.34; for the positive training stimuli, NRR = 1.69 ± 0.13; for the negative AVI test movies, NRR = 0.43 ± 0.34; for the negative training stimuli, NRR = 0.62 ± 0.35). In Group Static Multiple, classification of both the positive and the negative stimulus faces was only slightly impaired. However, that decrease in performance was sufficient to make the difference nonsignificant (NRR of the positive AVI test movies = 1.41 ± 0.36; for the positive training stimuli, mean NRR = 1.64 ± 0.71; for the negative AVI test movies, NRR = 0.79 ± 0.32; for the negative training stimuli, mean NRR = 0.49 ± 0.27).

Regarding the uncompressed movies, generalization performance in Group Static Single varied substantially among the subjects. Although 1 subject discriminated the test movies significantly (DR = 83.8), performance of the group did not exceed chance level [DR = 68.7 ± 13.1; $t(2) = 2.27$, n.s.]. This was due to slightly fewer responses to the positive test movies (NRR = 1.40 ± 0.20; for the negative test movies, NRR = 0.62 ± 0.28; the training values were the same as before). In Group Static Multiple, by contrast, not only was discrimination with uncompressed movies significant [DR = 95.9 ± 4.1; $t(2) = 5.88$, $p < .05$], but it was even better than with the static training stimuli. This outstanding performance was due to particularly low response rates with the negative test movies (NRR = 0.08 ± 0.09), whereas responding to the positive test movies was similar to the positive training stimuli (NRR = 1.66 ± 0.71; the training values were the same as before).

What emerges from those tests is quite interesting. The pigeons' failure to discriminate the AVI movies in all three groups confirmed the result of discrimination training—namely, that the pigeons found it difficult to extract the relevant information from such movies. Even the pigeons in Group Dynamic Uncompressed failed to generalize to the AVI movies, although human observers could not see any difference between those two presentation modes. However, the pigeons did not show any deterioration in performance when confronted with uncompressed movies showing randomized sequences of images, although for humans these movies generate the perception of an unnatural jerky motion of the head, thereby destroying the perception of smooth motion.

Regarding transfer to uncompressed movies, test performance of Group Static Multiple even exceeded that shown during training. For the pigeons in Group Static Multiple, the motion itself was the only additional information in the movies, since the individual images constituting the test movies were exactly the same as the static stimuli presented on a one-by-one basis during training. Therefore, it seems that dynamic presentation during the test actually promoted discrimination in Group Static Multiple, perhaps because all previously learned exemplars were presented together (i.e., in immediate succession within each trial). Although in Group Static Single test performance with the uncompressed movies dropped to a nonsignificant level, it did not break down completely, as was the

case with the AVI movies. It may be that the poorer performance with the uncompressed movies is attributable to a novelty effect caused by motion. However, regarding the excellent transfer to uncompressed movies in Group Static Multiple, we rather suspect that the movies contained too much additional information for the pigeons in Group Static Single to transfer the discrimination acquired by means of a single static stimulus to a moving stimulus consisting of 30 different views of the face (and thus containing 29 additional views, besides the 1 with which the subjects had been trained).
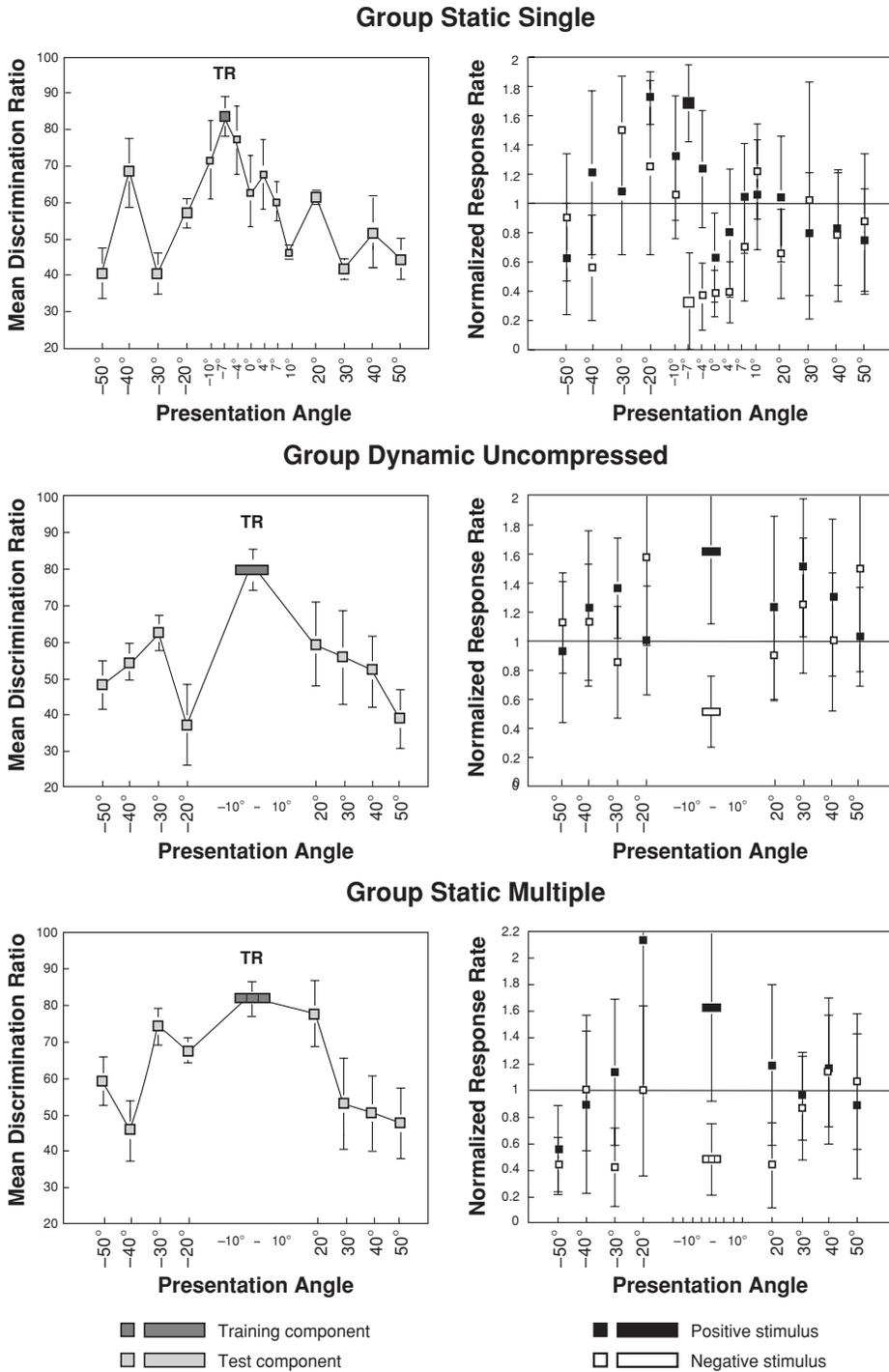
Test 2: Transfer to novel angular views. Groups Static Single, Static Multiple, and Dynamic Uncompressed were tested for transfer to static images showing the faces at viewing angles of 50º, 40º, 30º, and 20º in both rotation directions. Group Static Single was additionally tested with images showing the faces at viewing angles of −10º, −4º, 0º, +7º, and +10º. Figure 4 depicts mean discrimination ratios as a function of viewing angle in comparison with the training condition (left), as well as mean normalized response rates emitted to positive and negative stimuli (right) for all three groups.

Average discrimination rates with the test stimuli showing viewing angles between 20º and 50º substantially decreased in all three groups (Group Static Single, mean for all tested viewing angles, mean DR = 50.8 ± 14.6, and for the training stimuli, DR = 83.9 ± 9.6; Group Dynamic Uncompressed, mean DR = 51.1 ± 14.5, and for the training stimuli, DR = 78.5 ± 2.0; Group Static Multiple, mean DR = 59.5 ± 16.4, and for the training stimuli, DR = 78.2 ± 14.1). Groups Static Single and Dynamic Uncompressed did not generalize the discrimination to any of the tested views [all $ts(2) < 2.36$, n.s.]. In both groups, this was due to a decrease in average responses emitted to the positive test images, as well as an increase in average responses emitted to the negative test images (Group Static Single, with the positive test stimuli for all tested viewing angles, mean NRR = 1.00 ± 0.35, and for the positive training stimuli, NRR = 1.67 ± 0.26; for the negative test stimuli for all the tested viewing angles, mean NRR = 0.94 ± 0.31, and for the negative training stimuli, NRR = 0.33 ± 0.33; Group Dynamic Uncompressed, for the positive test stimuli, mean NRR = 1.20 ± 0.20, and for the positive training stimuli, NRR = 1.57 ± 0.42; for the negative test stimuli, mean NRR = 1.20 ± 0.26, and for the negative training stimuli, NRR = 0.43 ± 0.22). In Group Static Single, discrimination performances, with the test stimuli deviating only a few degrees from the training angle, decreased as well (mean DR = 61.6 ± 14.5; the training value is the same as before). However, discrimination of the faces was significant on several test views in at least some pigeons. In particular, 2 subjects significantly discriminated the faces with the −4º test view (DR = 82.1, and 90.4), 2 subjects also discriminated with the 0º test view (DR = 72.7, and 73.2), and 1 subject discriminated with the +4º view (DR = 86.7). But regarding Group Static Single as a whole, performances to all these views fell to a nonsignificant level [all $ts(2) \leq 3.48$, n.s.]. Response rates to the positive test images de-

creased substantially (for all tested viewing angles, mean NRR = 1.02 ± 0.26), but responses to the negative test stimuli increased only slightly (for all tested viewing angles, mean NRR = 0.69 ± 0.37; the training values are the same as before).

In Group Static Multiple, however, transfer of the discrimination to some novel angular views of the faces was significant not only for single subjects, but also for the group as a whole. Discrimination was significant with the +20º test view (DR = 77.7), the −20º test view (DR = 67.6), and the −30º test view [DR = 74.3; all $ts(2) \geq 4.78$, $p < .05$]. Good generalization to the −20º view was due to an extraordinarily high response rate to the positive test stimulus face that overcompensated for increased responding to the negative stimulus. By contrast, significant generalization performances with the −30º and the +20º views resulted from persisting low response rates with the negative test faces, whereas responses to the positive faces also decreased. On average, however, there was a decrease in responses emitted to the positive test images and an increase in responses emitted to the negative test images (for the positive test stimuli, mean NRR = 1.11 ± 0.50, and the positive training stimuli, NRR = 1.64 ± 0.71; for the negative test stimuli, mean NRR = 0.80 ± 0.31, and the negative training stimuli, NRR = 0.49 ± 0.27).

All in all, the tests for generalization to novel views also failed to demonstrate any advantage of dynamic over static stimulus presentation during training. On the contrary, it was Group Static Multiple that showed some rotational invariance at viewing angles up to a distance of 20º from the closest angle used in training. One possible explanation lies in the specific training conditions encountered by the pigeons in Group Static Multiple. During training, those birds were confronted with 30 slightly different views of each of the two heads, each view presented in a separate trial. Actually, the birds faced a categorization task, rather than a discrimination task, which required them to classify 30 instances of the positive and 30 instances of the negative categories. In both the human (Homa & Chambliss, 1975; Homa & Vosburgh, 1976) and the nonhuman (Overman & Doty, 1980; Wright, Cook, Rivera, Sands, & Delius, 1988; Wright, Santiago, & Sands, 1984) literature, there is ample evidence that concept learning depends on the number of exemplars involved. For example, when pigeons were trained with only 1 exemplar, they showed no evidence of transfer, but when they were trained with 152 different exemplars, each presented once per session, they showed excellent transfer (Wright et al., 1988). Likewise, the increase in exemplars from 1 to 30 should have enhanced generalization in the present experiment. Considering the underlying classification mechanism, either feature learning, based on the extraction of one or more class-invariant or category-specific features from the images (see, e.g., Herrnstein et al., 1976; Huber & Lenz, 1993; Huber et al., 2000; Jitsumori, 1993; Troje et al., 1999) or exemplar learning, which requires remembering all exemplars as templates, together with their class contingencies (see, e.g., Medin & Schaffer, 1978; Smith & Medin, 1999), could have led to the present results. The

Figure 4. Transfer to novel angular views. Left: Mean discrimination ratios ± standard deviations for training and test stimuli for Groups Static Single, Dynamic Uncompressed, and Static Multiple. The symbol TR indicates performance with the training component. Right: Mean normalized response rates ± standard deviations to the positive (filled symbols) and negative (open symbols) stimuli. Squares symbolize static presentation of a single view, blank rectangles symbolize dynamic presentation, and rectangles with lines inside symbolize static presentation of multiple views. Tested viewing angles were −50º, −40º, −30º, −20º, +20º, +30º, +40º, and +50º. Group Static Single was additionally tested with −10º, −7º, −4º, 0º, +4, +7º, and +10º views.

more class members (i.e., different stimuli) the pigeons encountered during training, the more likely they would have been to detect similarities between test stimuli and previously stored training stimuli.

Why, then, did the pigeons from Group Dynamic Uncompressed fail to show generalization at a level comparable to that found in Group Static Multiple, since both groups had access to the same number of exemplars? We suspect that the answer lies in the different modes of stimulus presentation. Actually, the training conditions for the two groups differed in one important aspect. Group Static Multiple explicitly responded to every single exemplar, because the individual images were presented in separate trials. By contrast, Group Dynamic Uncompressed was presented with all exemplars on every trial but may well have learned the discrimination by using only a few or, in the extreme case, only one. If so, the kind of discrimination learned by Group Dynamic Uncompressed and Group Static Single should be virtually the same. Actually, it was similar, since there were no differences in learning speed or discrimination performance between the two groups. Neither group generalized the discrimination to novel views of the faces. The only intergroup difference that could be revealed concerned transfer to uncompressed dynamic stimuli. Whereas Group Static Single did not generalize to the uncompressed dynamic presentation of the faces (Test 2c), Group Dynamic Uncompressed showed significant transfer to the randomized uncompressed movies (Test 2b). In fact regarding performance with the test movies, both groups behaved quite similarly, with slightly decreased response rates with the positive test movies, as compared with those with the positive training stimuli, and unchanged response rates with the negative test movies (Figure 3).

Group Dynamic Uncompressed showed good transfer to the static stimuli within the range covered by the training movies (Test 1a), whereas Group Static Single even failed to show good transfer to viewing angles that were closest to the training angle (Test 2), which were nearly the same angles. This may, in fact, be interpreted in terms of a dynamic superiority effect. However, such an effect can account for the results only if the birds of Group Dynamic Uncompressed not only based their discrimination decisions on only 1 single image out of all 30 images that were available, but also would have failed to perceive the other 29 views or would have forgotten them immediately after presentation. Only then would those images have been as novel to them as were the test images for the subjects in Group Static Single. Although that possibility cannot be ruled out, it does not seem very likely that such a mechanism should account for the pigeons' responding behavior.

DISCUSSION

The present experiment did not reveal an advantage of dynamic over static image presentation for both discrimination training and generalization of complex shape in pigeons. The discriminative performance with uncom-

pressed movies of human faces that repeatedly rotated back and forth between $-10°$ and $+10°$ was similar to that of the pigeons trained with multiple static images of the faces (Part 1: Phase 3); however, it was still better when the subjects were trained with only a single static image of each face (Phases 2 and 3). Likewise, tests for generalization to novel views of the faces did not demonstrate any advantage of training with dynamic over training with static images (Test 2). Only Group Static Multiple showed evidence of some viewpoint invariance.

The pigeons' complete failure to discriminate the faces when presented as AVI movies (Phases 1 and 2, Test 1) was rather surprising, given other studies in which AVI format presentation of videos led to quite satisfying discriminative responding in that species (e.g., Cook, Shaw, & Blaisdell, 2001). We suspect that this result could have been an artifact of the way in which the AVI movies were created. Usually (but not always) in AVI format movies, the information is compressed for storage. We cannot know how pigeons perceive a video that has been modified by the specific compression/decompression algorithm we used. These algorithms have been created to generate the impression of smooth movement in a human perceiver. But due to their higher flicker-fusion threshold, pigeons may be more sensitive to unsteady development of the picture and possibly unnatural deformations of the object caused by AVI compression. Physiological measurement indicates that the critical flicker-fusion frequency for pigeons is approximately 140 Hz under optimal conditions (Dodt & Wirth, 1953; Emmerton, 1983), whereas it is only 60 Hz for humans (Hecht & Verrijp, 1933). Thus, it seems likely that pigeons do not perceive such movies as smooth motion videos, but in a stroboscopic manner (D'Eath, 1998; see also Jitsumori & Makino, 2004, and Jitsumori et al., 1999, for related discussions).

Basically, the main findings of the present study are consistent with the results reported by Jitsumori and Makino (2004), who did not find any superiority effects of dynamic view training of human face stimuli, in comparison with static view training. However, there were also findings in their study that differ from our results. For example, they reported that after training on a single static view of the faces (the training view was the 0° frontal view, Experiment 1), their pigeons showed substantial generalization to novel viewpoints up to at least ±45°, whereas in the present study, except for individual subjects, Group Static Single did not even generalize to views that were closest to the training view. Apart from possible effects of procedural differences between the two studies, we suspect that such a fundamental discrepancy in transfer performance must be a consequence of some critical differences in the stimuli that had to be discriminated. Assuming a feature-learning account, Jitsumori and Makino suggested that the surprisingly good transfer to the rotated views of the faces in their study could be explained by the use of multiple 2-D features that were more or less distinctive over some range of rotation. The stimuli in Jitsumori and Makino were still video pictures containing all the information inherent in human faces, includ-

ing textural surface information. As has been shown by Troje et al. (1999), pigeons are especially sensitive to surface features, such as color, brightness, and other textural cues, which do not change with rotation in depth. These highly salient features from the texture domain had been removed in the present experiment (i.e., the face stimuli contained solely shape information). We suggest that this major difference in feature content is responsible for the differences in transfer performance to novel angular views found in the two studies.

An interesting finding reported by Jitsumori and Makino (2004) was their pigeons' failure to show transfer from static to dynamic view conditions (Experiment 2). The authors concluded that this finding, together with the birds' failure to transfer the discrimination to novel orientations outside the range covered by the dynamic views (Experiment 3), indicated that the pigeons did not perceive different views as belonging to the same object. In other words, it was suggested that the pigeons did not form three-dimensional structural representations of each face but, instead, "lumped different views together on the basis of similarity judgments" (p. 155). Regarding the results of the present experiment, we concur with that assumption. If this notion is correct, an increase in different views already presented during training should enhance generalization to dynamic presentation. This is exactly what we found. The pigeons in Group Static Multiple that were trained with all the static views constituting the movies showed excellent transfer, whereas the subjects presented with only a single view of the faces during training (Group Static Single) showed, at best, some moderate transfer to (uncompressed) dynamic presentation of the faces. In addition, similar to the findings reported by Jitsumori and Makino, the pigeons trained with dynamic presentations of the faces (Group Dynamic Uncompressed) failed to generalize the discrimination to novel orientations outside the range covered during training. Only Group Static Multiple showed some rotational invariance at viewing angles up to a distance of 20º from the closest angle used in training. Thus, the presentation of multiple static views during training enhanced generalization.

All these findings strongly support the hypothesis that pigeons group together different 2-D views of an object on the basis of similarity. This fits with the viewer-centered models in the human literature (e.g., Peissig et al., 2000; Tarr & Bülthoff, 1998; Tarr & Kriegman, 2001), which assume that the structure of a 3-D object is represented by a collection of its 2-D views stored in memory and that recognition of an object varies with similarities to the stored views. Of course, the more stored views that are available, the easier will be the recognition of a novel view, because the probability of a good fit with any of the stored views will increase. The findings of neither study (Jitsumori & Makino's [2004], as well as the present one) support the alternative hypothesis that the pigeons formed object-like 3-D representations of the faces, the view held by object-centered models (e.g., Biederman, 1987; Corballis, 1988; Marr & Nishihara, 1978). This is in sharp contrast to the findings of Cook and Katz (1999), who found evidence

that the pigeons in their study may have generalized mental 3-D representations of cube and pyramid objects. The most prominent support for their conclusion was the limited drop in performance whenever the subjects were tested with stimuli varying in, for example, size, rate of motion, direction of motion, and surface color. However, the stimuli they presented were simple geometric figures, and we believe that object constancy is strongly dependent on the difficulty of the discrimination. The more similar the correct object is to erroneous alternatives, the higher is the cost of changes in viewpoint (Edelman, 1995). Furthermore, object recognition in pigeons is likely to vary on a continuum between learning specific images and discriminating between individual objects, depending on the familiarity or expertise with a specific object class (Tarr & Cheng, 2003). Pigeons are obviously no experts in human face recognition. Therefore, a strong viewpoint dependency, such as we found in the present experiments, must be expected when they discriminate on the basis of literal images. It is still an open question whether pigeons can become experts in specific visual categories. Recent findings suggest that strong viewpoint dependency also emerges when pigeons are required to discriminate between pictures of different pigeons (Nakamura, Croft, & Westbrook, 2003).

The present experiment had been designed with the goal of testing whether motion would promote the perception of complex shape for pigeons, as is reported for humans. Empirical evidence indicates that for humans, movement is one of the most efficient cues for shape perception (see, e.g., Nawrot, Shannon, & Rizzo, 1996; Richards, 1985). However, it seems as if, at least in this study, the pigeons actually did not perceive the complex configural information of the faces. Rather, they may have relied on a collection of independent simple 2-D features. Most of the measurements in the vertical dimension, such as the height of the head, the height of the forehead, or the length of the nose, remain relatively constant over depth rotation. Given the pigeons' visual acuity, such size differences would be obvious candidates for discriminative features in the present task. The results confirmed our previous findings that complex shape information is generally not well accessible to pigeons and plays only a minor role as a cue in discrimination tasks (Troje et al., 1999). On the basis of the present results, together with those reported by Jitsumori and Makino (2004), we can conclude that despite movement, shape information does not have the same salience as texture cues for pigeons.

A final point is worth emphasizing. When computer or television monitors are used as stimulus generators for avian visual perception, a number of problems have to be considered. These devices have limited resolution, depend on apparent motion, and flicker at rates that are perceptible to birds, but not to the human experimenter. Their color generation is attuned to the human eye (Carnt & Townsend, 1969), not to that of the pigeon or any other bird species, and in particular, they emit no ultraviolet light. Furthermore, they are inherently 2-D. In the present study, the stroboscopic effect of images presented on

monitors with screen refresh rates far below the critical flicker-fusion frequency of pigeons might have been the most critical point. Some authors therefore find it questionable whether pigeons are principally able to interpret movies presented on a computer monitor as "real" moving objects (D'Eath, 1998; Fleishman & Endler, 2000; Patterson-Kane, Nicol, Foster, & Temple, 1997). At least, the assumption that pigeons see such movies in the same way as do humans is certainly unjustified. If the stimuli to be discriminated are simple geometric figures, pigeons might be able to compensate for the stroboscopic motion induced by movies with a presentation rate of 30 frames per second. But if the task involves complex shapes, they might, instead, rely on the simpler strategy of memorizing individual pictures or item-specific features. Research involving more sophisticated technology, such as LCD monitors and high temporal resolution, is needed to elucidate the difficult issue of the perception of structure from motion in pigeons.

REFERENCES

BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

CARNT, P. S., & TOWNSEND, G. B. (1969). *Colour television: N.T.S.C. system, principles and practice*. London: Iliffe.

CERELLA, J. (1977). Absence of perspective processing in the pigeon. *Pattern Recognition*, **9**, 65-68.

CERELLA, J. (1990). Pigeon pattern perception: Limits on perspective invariance. *Perception*, **19**, 141-159.

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., SHAW, R., & BLAISDELL, A. P. (2001). Dynamic object perception by pigeons: Discrimination of action in video presentations. *Animal Cognition*, **4**, 137-146.

CORBALLIS, M. C. (1988). Recognition of disoriented shapes. *Psychological Review*, **95**, 115-123.

CUTTING, J. E. (1986). *Perception with an eye for motion*. Cambridge, MA: MIT Press.

D'EATH, R. B. (1998). Can video images imitate real stimuli in animal behaviour experiments? *Biological Reviews*, **73**, 267-292.

DITTRICH, W. H., & LEA, S. E. G. (1993). Motion as a natural category for pigeons: Generalization and a feature-positive effect. *Journal of the Experimental Analysis of Behavior*, **59**, 115-129.

DITTRICH, W. H., & LEA, S. E. G. (2001). Motion discrimination and recognition. In R. G. Cook (Ed.), *Avian visual cognition*. Retrieved July 25, 2001, from www.pigeon.psy.tufts.edu/avc.

DITTRICH, W. H., LEA, S. E. G., BARRETT, J., & GURR, P. R. (1998). Categorization of natural movements by pigeons: Visual concept discrimination and biological motion. *Journal of the Experimental Analysis of Behavior*, **70**, 281-299.

DODT, E., & WIRTH, A. (1953). Differentiation between rods and cones by flicker electroretinography in pigeon and guinea pig. *Acta Physiologica Scandinavica*, **30**, 80-89.

EDELMAN, S. (1995). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics*, **72**, 207-220.

EMMERTON, J. (1983). Vision. In M. Abs (Ed.), *Physiology and behaviour of the pigeon* (pp. 245-266). London: Academic Press.

EMMERTON, J. (1986). The pigeon's discrimination of movement patterns (Lissajous figures) and contour-dependent rotational invariance. *Perception*, **15**, 573-588.

FLEISHMAN, L. J., & ENDLER, J. A. (2000). Some comments on visual perception and the use of video playback in animal behavior studies. *Acta Ethologica*, **3**, 15-27.

GIBSON, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

HECHT, S., & VERRIJP, C. D. (1933). Intermittent stimulation by light. *Journal of General Physiology*, **17**, 237-282.

HERRNSTEIN, R. J., LOVELAND, D. H., & CABLE, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**, 285-302.

HILL, H., SCHYNS, P. G., & AKAMATSU, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, **62**, 201-222.

HODOS, W., SMITH, L., & BONBRIGHT, J. C. (1975). Detection of the velocity of movement of visual stimuli by pigeons. *Journal of the Experimental Analysis of Behavior*, **25**, 143-156.

HOLLARD, V., & DELIUS, J. D. (1982). Rotational invariance in visual pattern recognition by pigeons and humans. *Science*, **218**, 804-806.

HOMA, D., & CHAMBLISS, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning & Memory*, **1**, 351-359.

HOMA, D., & VOSBURGH, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 322-330.

HUBER, L. (1994). Amelioration of laboratory conditions for pigeons. *Animal Welfare*, **3**, 321-324.

HUBER, L., & LENZ, R. (1993). A test of the linear feature model of polymorphous concept discrimination with pigeons. *Quarterly Journal of Experimental Psychology*, **46B**, 1-18.

HUBER, L., TROJE, N. F., LOIDOLT, M., AUST, U., & GRASS, D. (2000). Natural categorization through multiple feature learning in pigeons. *Quarterly Journal of Experimental Psychology*, **53B**, 341-357.

HUMPHREYS, G. W., RIDDOCH, M. J., & QUINLAN, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, **5**, 67-103.

JITSUMORI, M. (1993). Category discrimination of artificial polymorphous stimuli based on feature learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 244-254.

JITSUMORI, M., & MAKINO, H. (2004). Recognition of static and dynamic images of depth-rotated human faces by pigeons. *Learning & Behavior*, **32**, 145-156.

JITSUMORI, M., NATORI, M., & OKUYAMA, K. (1999). Recognition of moving video images of conspecifics by pigeons: Effects of individuals, static and dynamic motion cues, and movement. *Animal Learning & Behavior*, **27**, 303-315.

JOHANSSON, G. (1975). Visual motion perception. *Scientific American*, **232**, 76-89.

LOIDOLT, M., AUST, U., MERAN, I., & HUBER, L. (2003). Pigeons use item-specific and category-level information in the identification and categorization of human faces. *Journal of Experimental Psychology: Animal Behavior Processes*, **29**, 261-276.

LUMSDEN, E. A., JR. (1970). Implication of the equivalence of mirror-image stimuli for object constancy. *Psychonomic Science*, **19**, 55-56.

LUMSDEN, E. A., [JR.] (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.

MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **200**, 269-294.

MARTINOYA, C., & DELIUS, J. D. (1990). Perception of rotating spiral patterns by pigeons. *Biological Cybernetics*, **63**, 127-134.

MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

MOSES, Y. (1993). *Face recognition: Generalization to novel images*. Unpublished doctoral thesis, Weizmann Institute of Science, Israel.

MOSES, Y., ULLMAN, S., & EDELMAN, S. (1996). Generalization to novel images in upright and inverted faces. *Perception*, **25**, 443-461.

MULVANNY, P. (1978). Velocity discrimination by pigeons. *Vision Research*, **18**, 531-536.

NAKAMURA, T., CROFT, D. B., & WESTBROOK, R. F. (2003). Domestic pigeons (*Columba livia*) discriminate between photographs of individual pigeons. *Learning & Behavior*, **31**, 307-317.

NAWROT, M., SHANNON, E., & RIZZO, M. (1996). The relative efficacy of cues for two-dimensional shape perception. *Vision Research*, **36**, 1141-1152.

Neiworth, J., & Rilling, M. E. (1987). A method for studying imagery in animals. *Journal of Experimental Psychology: Animal Behavior Processes*, **13**, 203-214.

Overman, W. H. J., & Doty, R. W. (1980). Prolonged visual memory in macaques and man. *Neuroscience*, **5**, 1825-1831.

Patterson-Kane, E., Nicol, C. J., Foster, T. M., & Temple, W. (1997). Limited perception of video images by domestic hens. *Animal Behavior*, **53**, 951-963.

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.

Peissig, J. J., Young, M. E., Wasserman, E. A., & Biederman, I. (1999). The pigeon's perception of depth-rotated shapes. *Current Psychology of Cognition*, **18**, 657-690.

Peissig, J. J., Young, M. E., Wasserman, E. A., & Biederman, I. (2000). Seeing things from a different angle: The pigeon's recognition of single geons rotated in depth. *Journal of Experimental Psychology: Animal Behavior Processes*, **26**, 115-132.

Pike, G., Kemp, R., Towell, N., & Phillips, K. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition*, **4**, 409-437.

Pisacreta, R. (1982). Stimulus control of the pigeon's ability to peck a moving target. *Journal of the Experimental Analysis of Behavior*, **37**, 301-309.

Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A*, **2**, 343-349.

Shimizu, T. (1998). Conspecific recognition in pigeons (*Columba livia*) using dynamic video images. *Behaviour*, **135**, 43-53.

Siegel, R. K. (1970). Apparent movement detection in the pigeon. *Journal of the Experimental Analysis of Behavior*, **14**, 93-97.

Smith, E. [E.], & Medin, D. [L.] (1999). The exemplar view. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 207-221). Cambridge, MA: MIT Press.

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.

Spetch, M. L., Kelly, D. M., & Reid, S. [L.] (1999). Recognition of objects and spatial relations in pictures across changes in viewpoint. *Current Psychology of Cognition*, **18**, 729-764.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey, and machine. *Cognition*, **67**, 1-20.

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences*, **7**, 23-30.

Tarr, M. J., & Kriegman, D. J. (2001). What defines a view? *Vision Research*, **41**, 1981-2004.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, **36**, 1761-1771.

Troje, N. F., & Bülthoff, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research*, **38**, 79-89.

Troje, N. F., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999). Categorical learning in pigeons: The role of texture and shape in complex static stimuli. *Vision Research*, **39**, 353-366.

Troje, N. F., & Vetter, T. (1998). Representations of human faces. In C. Taddei-Ferretti & C. Musio (Eds.), *Downward processes in the perception representation mechanisms: Proceedings of the International School of Biocybernetics* (pp. 189-205). Singapore: World Scientific.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Vaughan, W. J., & Greene, S. L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 256-271.

Vetter, T., & Troje, N. (1997). Separation of texture and shape in images of faces for image coding and synthesis. *Journal of the Optical Society of America A*, **14**, 2152-2161.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217.

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.

Wright, A. A., Cook, R. C., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning & Behavior*, **16**, 436-444.

Wright, A. A., Santiago, H. C., & Sands, S. F. (1984). Monkey memory: Same/different concept learning, serial probe acquisition, and probe delay effects. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 513-529.