Journal of Experimental Psychology: Animal Learning and Cognition
2014, Vol. 40, No. 2, 162–177

© 2013 American Psychological Association
2329-8456/14/$12.00 DOI: 10.1037/xan0000014

Scene Gist Categorization by Pigeons

Kimberly Kirkpatrick and Tannis Bilton
Kansas State University

Bruce C. Hansen
Colgate University

Lester C. Loschky
Kansas State University

Scene gist categorization in humans is rapid, accurate, and tuned to the statistical regularities in the visual world. However, no studies have investigated whether scene gist categorization is a general process shared across species, or whether it may be influenced by species-specific adaptive specializations relying on specific low-level scene statistical regularities of the environment. Although pigeons form many types of categorical judgments, little research has examined pigeons' scene categorization, and no studies have examined pigeons' ability to do so rapidly. In Experiment 1, pigeons were trained to discriminate between either 2 basic-level categories (beach vs. mountain) or a superordinate-level natural versus a manmade scene category distinction (beach vs. street). The birds learned both tasks to a high degree of accuracy and transferred their discrimination to novel images. Furthermore, the pigeons successfully discriminated stimuli presented in the 0.2- to 0.35-s duration range. Therefore, pigeons, a highly divergent species from humans, are also capable of rapid scene categorization, but they require longer stimulus durations than humans. Experiment 2 examined whether pigeons make use of complex statistical regularities during scene gist categorization across multiple viewpoints. Pigeons were trained with the 2 natural categories from Experiment 1 (beach vs. mountain) with zenith (90°), bird's eye (45°), and terrestrial (0°) viewpoints. A sizable portion of the variability in pigeon categorization performance was explained by the systematic variation in scene category-specific statistical regularities, as with humans. Thus, rapid scene categorization is a process that is shared across pigeons and humans, but shows a degree of adaptive specialization.

Keywords: scene perception, scene gist, scene categorization, go/no-go, pigeons

Humans can perform rapid, holistic, semantic processing of a real-world scene within a single eye fixation, known as *scene gist perception*, but what about other divergent animal species? If divergent animal species can accomplish the same task, does the process differ in a qualitatively important way from humans? We report the first study to investigate such issues, examining rapid scene gist categorization by pigeons, a highly divergent species from humans.

This article was published Online First December 23, 2013 under the previous title (*Journal of Experimental Psychology: Animal Behavior Processes*).

Kimberly Kirkpatrick and Tannis Bilton, Department of Psychological Sciences, Kansas State University; Bruce C. Hansen, Department of Psychology and Neuroscience Program, Colgate University; Lester C. Loschky, Department of Psychological Sciences, Kansas State University.

Tannis Bilton is now at the University of Lincoln, Lincoln, United Kingdom.

Parts of the research in this article were previously presented at the 2010 Annual Meeting of the Society for Quantitative Analyses of Behavior (Experiment 1) and at the 2012 Annual Meeting of the Applied Behavior Analysis International Society (Experiment 2). We thank David Ghormley, Michelle Niebres, and Marina Vilardo for their assistance with data collection.

Correspondence concerning this article should be addressed to Kimberly Kirkpatrick, Department of Psychological Sciences, Kansas State University, 492 Bluemont Hall, 1100 Mid-Campus Drive, Manhattan, KS 66506. E-mail: kirkpatr@ksu.edu

In humans, scene gist recognition is extremely rapid, occurring with stimulus durations of 0.027 s or less (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Fei-Fei, Iyer, Koch, & Perona, 2007; Loschky et al., 2007), may occur automatically (Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002), and may precede object recognition (Fei-Fei et al., 2007; Oliva & Torralba, 2001; Schyns & Oliva, 1994). The gist of a scene is often operationally defined as either the superordinate- or the basic-level category to which the scene belongs (Tversky & Hemenway, 1983). The extreme rapidity of gist recognition is consistent with the fact that it appears to be guided by the early low-level perceptual information within a scene (Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Loschky, Hansen, Sethi, & Pydimari, 2010; Loschky & Larson, 2008; Loschky et al., 2007). Gist recognition activates related prior knowledge, which guides attention to relevant areas of the scene in visual search (Eckstein, Drescher, & Shimozaki, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). Having activated a scene's gist may facilitate a viewer's recognition of constituent objects in the scene (Davenport & Potter, 2004; Palmer, 1975) and influence their later memory for the scene (Brewer & Treyens, 1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). Yet, whereas it is clear that scene gist is a fundamental process in human visual perception, to date little is known about whether the ability to rapidly extract the gist of a scene is a general process shared across divergent animal species (Soto & Wasserman, 2012).

One way of beginning to address these questions is to explore scene gist processing in pigeons, a popular nonhuman model of

visual cognition (e.g., R. G. Cook, 2001; Fagot, 2000). Comparisons between pigeon and human visual cognition are of theoretical interest because progenitors of the two species diverged during the early Permian period (Jarvis et al., 2005); thus, any similarities between them must either be based on shared evolutionary structures (homology) or due to the environment pushing their evolution along a similar path (homoplasy due to convergent evolution). Both object recognition and categorization by pigeons have been well studied, and it appears that pigeons rely on features similar to those used by humans (see Kirkpatrick, 2001, for a review). Given these similarities, it is possible that pigeons, like humans, rapidly perceive the gist of scenes "at a glance." If this is the case, then this would suggest that the ability to rapidly categorize scenes in a single eye fixation may be a general process shared across divergent animal species. Experiment 1 investigated this broader question and, more specifically, the speed with which pigeons were able to categorize real-world scenes, as measured by the stimulus durations needed for accurate performance. In addition, Experiment 1 addressed whether pigeons could more easily discriminate between scenes differing at the superordinate categorical level (e.g., natural vs. manmade scenes, such as beaches vs. streets) than those from the same superordinate category (e.g., natural scenes, such as beaches vs. mountains), a factor that has proven important in human scene gist recognition (Joubert et al., 2009; Loschky & Larson, 2010; Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009).

The incredible speed of scene categorization in humans suggests that a majority of the category-specific signal may arise from critical "low-level" image structures. Such structures have been characterized by specific second-order and higher order statistical regularities of luminance contrast (i.e., global and local image properties, respectively, described in detail in Experiment 2) within different scene categories. Interestingly, such statistical regularities have been shown to selectively modulate different neural populations in the early human visual system (e.g., Felsen & Dan, 2005; Felsen, Touryan, Han, & Dan, 2005; Hansen, Jacques, Johnson, & Ellemberg, 2011; Hansen, Johnson, & Ellemberg, 2012) and therefore may serve as the primary guiding signal in rapid processing via a parallel (Rousselet et al., 2002) feedforward system (Thorpe, 2002; VanRullen & Koch, 2003). Furthermore, recent behavioral work (e.g., Hansen & Loschky, 2013; Loschky, Hansen, et al., 2010) has demonstrated the relative importance of both types of scene statistical regularities in human rapid scene categorization. Experiment 2 of the current study therefore explored whether pigeon rapid scene categorization abilities similarly relied on low-level scene properties. Alternatively, the highly divergent evolutionary histories of pigeons and humans may have led to species-specific adaptive specializations in their abilities to recognize scene gist. Given that pigeons have evolved specialized systems for flight (whereas humans have evolved on the ground), one would expect that if pigeons do rely on scene statistical regularities, then they would rely on these regularities at several different viewpoints. This was explored in Experiment 2.

Experiment 1

The present experiment assessed scene gist categorization in pigeons as measured by their discrimination and generalization of scene categories and the rapidity of those categorization processes. Numerous studies have shown that pigeons are adept at categorizing real-world photographs of objects, faces, and abstract shapes (e.g., Bhatt, Wasserman, Reynolds, & Knauss, 1988; Herrnstein, Loveland, & Cable, 1976; Huber, Roje, Loidolt, Aust, & Grass, 2000; Von Fersen & Lea, 1990). However, such studies have not examined pigeons' ability to categorize entire real-world scenes, and the vast majority of previous studies of pigeons' image categorization abilities have used far longer stimulus durations (e.g., 5 s) than those typically used for studies of human scene gist recognition ($\leq$ 0.027 s). Nevertheless, a handful of recent pigeon vision studies have begun to examine pigeons' ability to rapidly process visual images.

Kramer (2010) examined rapid discrimination of simple geometric stimuli and naturalistic slides (containing humans vs. nonhumans) with randomly intermixed exposure times of 0.1–10 s. Discrimination accuracy was reduced with display times of 1 s or less, but some birds were able to perform above chance at durations less than 0.5 s. In addition, pigeons have been shown to engage in texture segregation with as little as 0.1 s of exposure time (R. G. Cook, Cavoto, Katz, & Cavoto, 1997). Finally, pigeons have been shown to be able to discriminate same versus different scene image sequences (e.g., AAA vs. ABA vs. ABC) with only a 0.5-s stimulus duration for the first item in the series, although shorter durations significantly reduced discrimination performance (R. G. Cook & Blaisdell, 2006). All of these previous studies suggest that visual recognition in various tasks can occur fairly rapidly in pigeons, although it is not known whether they can rapidly process complex scene image information in terms of scene category distinctions. In the current experiment, we tested not only whether pigeons can rapidly categorize real-world scenes, but also determined the minimum stimulus duration that was required for accurate performance.

A second aim of the current experiment was to establish whether pigeons were better able to discriminate between scene categories that differ at the superordinate level than those that share the same superordinate category. Several studies have recently shown that when human subjects are allowed only very brief processing times, they are more accurate at discriminating scene categories that differ at the superordinate level, for example, beach (natural) versus street (manmade), than they are at discriminating categories that are from the same superordinate category, for example, beach and mountain (both natural; Joubert et al., 2009; Loschky & Larson, 2010; Macé et al., 2009). In the current study, we also compared pigeons' performance on cross- versus within-superordinate categorization.

Method

Subjects. The subjects were eight experimentally naïve, mixed-sex, captive-bred homing pigeons (*Columba livia*; Double-T Farms, Glenwood, IA) that were approximately 2 years of age at the start of the experiment. The birds were housed in individual cages in a colony room on a 12:12 light–dark cycle with light onset at 8 a.m. Each bird was maintained at 85% of its free-feeding weight by the delivery of individual 45-mg pigeon pellets (Test Diet, Richmond, IN) in the experimental apparatus and supplementary access to regular mixed grain (Des Moines Feed Company, Des Moines, IA) in the home cage, ranging from 5 to 15 g per day. The birds were allowed free access to red pigeon grit (Des Moines Feed Company) and water in the home cage. One bird

(Y7) was removed from the study because of chronic low response rates.

Apparatus. The pigeons were trained and tested in two 35 × 32 × 24 cm operant chambers housed inside of a sound- and light-attenuating box (Med Associates, St. Albans, VT). One wall of the chamber was fitted with a 15-in. touch screen (Elotouch Solutions, Accutouch, Menlo Park, CA) that was in front of a 15-in. flat-panel monitor that was turned on its side. The monitors were set at a resolution of 1,024 × 768 pixels for the duration of the experiment and were calibrated to the following values: Kelvin = Whitepoint (6500), $\gamma = 2.200$, White-Luminance (brightness) = 80.600, Black-Luminance (contrast) = 0.420, Color (*x*-value) = 0.314, and Color (*y*-value) = .324. On the opposite wall of the chamber were a magazine pellet dispenser (Med Associates, ENV-203) and clicker (Med Associates, ENV-135M) that were activated at the same time as the pellet dispenser. Individual 45-mg pigeon pellets were delivered through a tube into a food cup (Med Associates, ENV-200-R1M) that was located 2 cm above the grid floor. A houselight was located on the top-left wall and delivered diffuse illumination to the pigeon chamber at an intensity of approximately 200 lux (Med Associates, ENV-227M). Responses were recorded from the touch screen via a USB touch screen controller (Elotouch Systems, 3000U USB controller). Control of the feeder and houselight was accomplished by a digital I/O card (National Instruments, Austin, TX; PCI-6503). A video splitter (Rextron, Taipei, Taiwan, BSA12) allowed for the simultaneous presentation of images to the operant chamber and control computer. Two Dell P4 computers located adjacent to each operant chamber delivered the experimental procedures and recorded data using MATLAB Version 7.1. The location of the peck was recorded in the form of *x*, *y* coordinates with a time tag.

Stimuli. The stimuli consisted of digital photographs of beaches, mountains, and streets. The basic-level categories can be further grouped into superordinate categories of "natural" (beach and mountain) and "manmade" (street). All images were sampled in their native form (high-resolution RGB color format, 8 bits per channel) from freely available sources on the Web and cropped to 512 × 512 pixels. All images were then converted to grayscale using the standard National TV Standards Committee formula (i.e., luminosity = $0.299 * R(x) + 0.587 * G(x) + 0.114 * B(x)$). The grayscale stimuli were adjusted to have the same root mean squared contrast, 0.25, defined as the standard deviation of all image pixel grayscale values normalized by the mean of all pixel grayscale values (e.g., Pavel, Sperling, Riedl, & Vanderbeek, 1987). Finally, all root mean squared-normalized images were set to possess the same mean grayscale value (127, on a 0–255 scale). There were 20 images randomly assigned to each category (S+ and S–) used during the training phase and a further 40 novel images (S+ and S–) that were used in generalization testing.

Procedure.

Pretraining. The pigeons were given initial training to eat from the food cup and to peck at a 512 × 512 pixel gray square presented in the middle of the touch screen on a black background. Once the birds began pecking, they were given training with a series of fixed ratio schedules (FR1, FR2, FR3, and FR5) to peck the square for food reinforcement. The FR schedules were presented in discrete trials (100 trials per session) and reinforced with a single 45-mg food pellet, with a 10-s intertrial interval (ITI) during which the monitor was dark. The pigeons progressed to the next FR once they completed all 100 trials within 1 hr. Once the birds pecked the screen reliably on an FR5, they received a final pretraining phase in which a fixation cue was added prior to the FR5 requirement. Each trial began with a 2.0-cm diameter white circular fixation point that was presented in the middle of the screen. The pigeon had to peck this one time, after which there was a uniformly distributed random delay ($M = 0.6$ s, range = 0.3–0.9 s), and then the gray square was presented and remained on the screen until the pigeon completed the FR5. The circle size was then reduced to 1.0 cm and then to 0.5 cm in successive phases. Pigeons received a minimum of two sessions of training on each of the pretraining procedures until they pecked reliably at each stage before progressing to the next stage.

Discrimination training. The pigeons were randomly assigned to one of two groups and were trained with a go/no-go procedure to discriminate between two different categories. One group received a discrimination task between two natural categories (beaches vs. mountains; Group N-N) and the other received a natural versus a manmade category (beaches vs. streets; Group N-M). There were 20 exemplars from each category used in the training phase (see Figure 1 for examples). One of the categories was designated as the positive (S+ or "go") and the other as the negative (S– or "no-go") category; stimulus assignments were counterbalanced across birds. An example of an individual trial is diagrammed in Figure 2. The trials were structured similarly to the pretraining phase except that the initial 0.5-cm fixation circle was followed by the presentation of a gray square, lasting 0.3–0.9 s, and then a grayscale image from one of the categories was presented in the center of the screen for 5 s and surrounded by a black background. The image was then replaced by the gray square, which remained on the screen for 5 s, during which go/no-go responses were collected. After the 5-s response interval expired, the screen darkened and a 10-s ITI was delivered.

On S+ trials, if the pigeon pecked the gray square during the 5-s response period, it received a single food pellet at the end of the trial as reinforcement. If the pigeon failed to peck, food was omitted. On S– trials, if the pigeon pecked during the 5-s response period, this resulted in a blackout of the houselight, as a mild punishment, during the first half of the ITI. If the pigeon successfully withheld responding on S– trials, the ITI was delivered without a blackout of the houselight. There were 200 trials per session, 100 S+ and 100 S–. The stimuli were presented in five 40-trial blocks. Each block consisted of 20 S+ and 20 S– stimuli that were randomly ordered, with no repeats of individual stimuli within a block, but with repeats of stimuli across blocks. The pigeons were trained until they produced a discrimination ratio (DR; see Data analysis) of at least 0.80 for two consecutive sessions. This criterion is slightly more stringent than in previous studies, which have generally required accuracy rates in the 0.70–0.75 range (Kirkpatrick, Wilkinson, & Johnston, 2007; Kirkpatrick-Steger, Wasserman, & Biederman, 1996, 1998), to ensure a high-level discrimination accuracy prior to rapid discrimination training (see below).

Generalization testing. During the generalization testing phase, the birds continued to receive normal discrimination training trials, but there were 20 novel test trials added per session, 10 S+ and 10 S–. Test trials were randomly intermixed with training trials, with two S+ and 2 S– stimuli added to each training block. The structure of test trials was identical to training trials except



Figure 1. Training images used in the experiment. The images have been reduced from 512 × 512 pixels for presentation purposes.

that the reinforcement contingency was neutral so that there was no reinforcement or punishment. Testing was conducted in blocks of four sessions, with two or more retraining sessions in between test blocks. There were no repeats of test stimuli within a four-session test block. Thus, a total of 40 S+ and 40 S– stimuli were presented over the course of a four-session block for each bird and each block was repeated five times so that each individual test image was presented five times across the generalization phase (but only once within a block of test sessions). Testing was recommenced when the DR was at least 0.80 on two consecutive retraining sessions (see, e.g., Kirkpatrick-Steger et al., 1996). A total of 20 test sessions were delivered for a total of 200 S+ and 200 S– test trials.

Rapid categorization training. During rapid categorization training, the stimulus duration was reduced across sessions: 4.00, 2.00, 1.00, 0.75, 0.50, 0.35, 0.20, and 0.10 s. Pigeons were trained

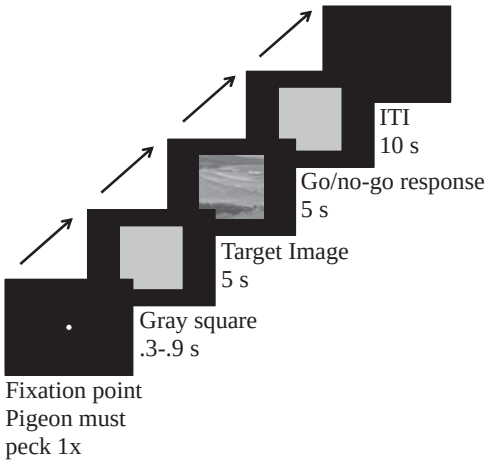


Figure 2. A schematic of an individual trial. Training and generalization test trials were identical except that training trials contained feedback in the form of food reinforcement or punishment by blackout of the houselight. During the rapid categorization training phase, the stimulus duration was decreased while all other aspects of the trial remained the same. ITI = intertrial interval.

with each of the stimulus durations until achieving a DR of 0.80 for two consecutive sessions. In cases in which birds failed to reach criterion, their duration was not reduced. All other aspects of training were the same as in the original discrimination training phase.

Data analysis. Pecks were analyzed over the 5-s response period on each trial in all phases of the experiment. The measure of discrimination performance was a DR during the go/no-go period, which was equal to $A/(A + B)$, where A is the response rate on S+ trials and B is the response rate on S– trials.

Statistical analyses were conducted in SPSS Version 20. For F tests, measurements of effect size used a partial eta-squared statistic, and those for paired-sample and between-subjects t tests used a Cohen's d statistic. The 95% confidence interval (CI) of the difference of two means is reported for F tests involving two means and for follow-up t tests of F tests with more than two means.

Results

Discrimination training. Figure 3 shows DRs as a function of sessions of training for individual birds that were trained on beach versus mountain (top panel) or beach versus street (bottom panel) discriminations. All seven birds acquired the initial training discrimination with no significant difference in speed of learning between Group N-N (M = 39 sessions, SD = 18) and Group N-M (M = 35 sessions, SD = 8), $t(5) = 0.3$. During the last two sessions, the average DR was 0.86 (SD = 0.06) for Group N-N and 0.84 (SD = 0.003) for Group N-M; there was no group difference in asymptotic performance, $t(5) = 0.7$.

Generalization testing. As shown in Figure 4, both groups of birds successfully generalized their discrimination learning to novel images while also maintaining a high level of accuracy on the training stimuli. The average generalization decrement for Group N-N was 0.09 and for Group N-M was 0.05. There was a significant decrease in accuracy between the training and test trials, $F(1, 5) = 9.1$, $p = .030$, $\eta_p^2 = .64$, power = .68, 95% CI = [.01, .13], but no group main effect, $F(1, 5) < 1$, or any Group × Trial type interaction, $F(1, 5) = 1.1$. A one-sample t test (comparing against DR = 0.50) revealed that both the training and test stimuli resulted in above-chance performance in both Group N-N, $t(3) = 12.4$, $p = .001$, for training, and $t(3) = 6.6$, $p = .007$, for testing, and Group N-M, $t(2) = 28.7$, $p = .001$, for training, and $t(2) = 8.3$, $p = .014$, for testing.

Rapid categorization training. Table 1 displays the number of sessions for each bird to reach the performance criterion (DR ≥ 0.80 for two consecutive sessions) at each stimulus duration. All of the birds easily met criterion for durations of 4.00, 2.00, 1.00, and 0.75 s, but from 0.50 s onward, the birds required more sessions to

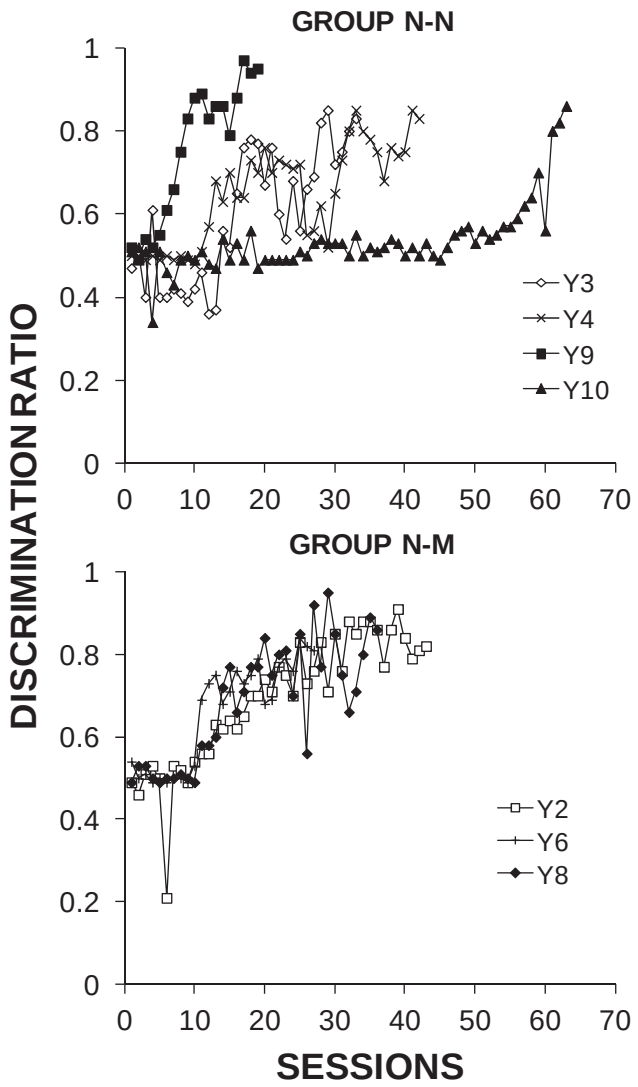


Figure 3. Discrimination ratios as a function of sessions for individual birds (pigeons Y3, Y4, Y9, and Y10) in Group N-N (top panel) and Group N-M (pigeons Y2, Y6, and Y8; bottom panel) during the discrimination training phase. N-N = natural–natural (beach vs. mountain); N-M = natural–manmade (beach vs. street).

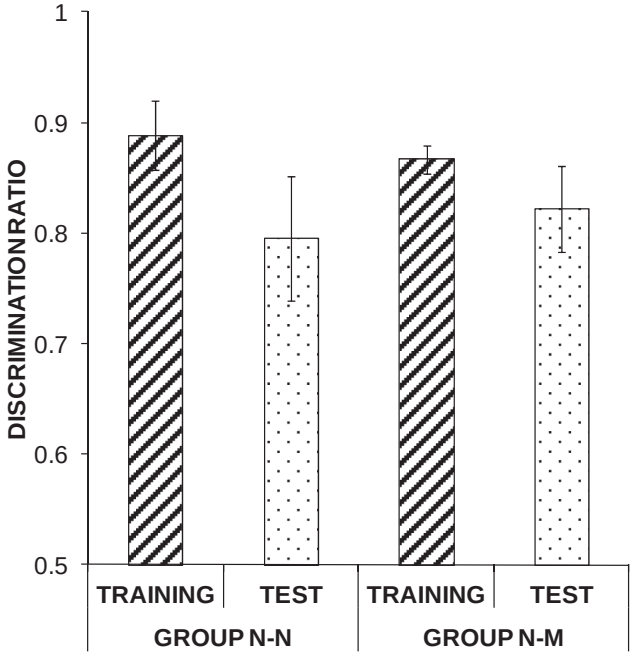


Figure 4. Mean (±*SEM*) discrimination ratio during the generalization testing phase on training and test trials in Groups N-N and N-M. N-N = natural–natural (beach vs. mountain); N-M = natural–manmade (beach vs. street).

meet criterion, and some birds began to fail to meet criterion from the 0.35-s duration onward. A mixed analysis of variance (ANOVA) with the variables of stimulus duration (within-subjects) and group (between-subjects) was conducted on the sessions to criterion data for the five durations that all seven birds successfully mastered (4.00, 2.00, 1.00, 0.75, and 0.50 s). This revealed a near-significant effect of stimulus duration, $F(4, 20) = 2.8$, $p = .053$, $\eta_p^2 = .36$, power = .66, that was reflective of the general tendency for increased numbers of sessions to criterion as duration decreased. There was no effect of group, $F(1, 5) = 2.3$, nor was there any Group × Stimulus duration interaction, $F(4, 20) < 1$.

Figure 5 displays the DRs as a function of duration for the two groups (for the six durations on which all birds were tested). All birds showed degraded performance as a function of decreasing duration. As shown in Figure 5, there was a tendency for birds in Group N-M to show better performance for the shorter durations (below 2 s). An ANOVA was conducted on the DRs for the six durations that all birds were tested on (4.00, 2.00, 1.00, 0.75, 0.50, and 0.35 s). There was a significant impairment in DR as a function of decreasing stimulus duration, $F(5, 25) = 3.6$, $p = .015$, $\eta_p^2 = .42$, power = .85, but there was no difference in performance between the two groups, $F(1, 5) = 1.8$, or any Group × Stimulus duration interaction, $F(5, 25) < 1$. Follow-up t tests on the stimulus duration main effect, using a Bonferroni correction, were conducted to determine the largest duration that differed from 4.00 s as an indicator of the point at which performance first dropped below the DR at 4.00 s. This revealed a significant difference at 0.50 s, $t(6) = 2.7$, $p = .037$, $d = 1.44$, 95% CI [0.01, 0.16].

One-sample t tests were conducted to assess whether performance on each of the durations was above chance (comparison to DR = 0.50), which included all birds that were tested on a given duration irrespective of group designation. This revealed that all durations resulted in above-chance performance, smallest $t(6) = 7.3$, $p < .001$.

Table 1
Sessions to Reach Performance Criterion of 0.80 Discrimination Ratio for Two Consecutive Sessions for Each Stimulus Duration During the Rapid Discrimination Training Phase for Each Pigeon

Stimulus duration (s)	Pigeon						
	Y3	Y4	Y9	Y10	Y2	Y6	Y8
4.00	2	9	2	2	2	2	2
2.00	2	3	2	2	2	2	2
1.00	8	2	2	2	2	6	2
0.75	2	5	2	2	2	2	2
0.50	18	3	4	20	3	10	2
0.35	. . .	24	2	5	2	. . .	2
0.20	X	. . .	. . .	. . .	18	X	3
0.10	X	X	X	X	. . .	X	. . .

Note. Stimulus durations that were not successfully mastered are marked with an ellipsis and durations that were not delivered are indicated with an X.

Discussion

Experiment 1 investigated whether pigeons, like humans, are able to rapidly categorize real-world scenes. Our investigation focused on fundamental aspects of scene gist recognition processes in human vision, namely the ability to discriminate between image categories, generalize to new exemplars, and categorize briefly flashed scene images as a function of their presentation duration. We found that pigeons successfully mastered a go/no-go discrimination with both a natural versus manmade (beach vs. street) and a natural versus natural (beach vs. mountain) discrimination. In addition, they successfully generalized their discrimination to new

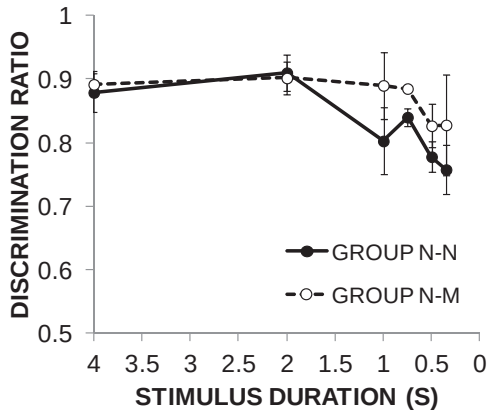


Figure 5. Mean (±*SEM*) discrimination ratios as a function of stimulus duration during the rapid categorization phase. N-N = natural–natural (beach vs. mountain); N-M = natural–manmade (beach vs. street).

exemplars of the categories. The generalization decrement was modest (5–10%), and performance to the novel images was well above chance. This indicates that the pigeons learned to categorize the images rather than simply memorizing individual exemplars; this is important because pigeons have been shown to be able to memorize large numbers of snapshots over long periods of time (Vaughan & Greene, 1984). The present results join a host of other studies that have successfully demonstrated categorization abilities in pigeons (e.g., Bhatt et al., 1988; Herrnstein et al., 1976) and extend those findings to categorizing whole scenes on the basis of both basic-level and superordinate-level scene category distinctions. Previous studies have focused on learning categories based on objects (e.g., people, cars, flowers) rather than the gist (or context) of an entire scene.

In addition to demonstrating scene categorization by pigeons, the present study also showed that pigeons could categorize a scene above chance with as little as 0.10 s of stimulus exposure. In examining recognition of simple geometric shapes and colors, Kramer (2010) found results similar to the present findings. Namely, the birds showed decrements in performance with durations of 1.00 s and below, but some birds were able to perform above chance with durations as short as 0.10 s. The birds in the present study did not begin to show noticeable deficits in performance until they reached stimulus durations below 0.50 s. It therefore seems that pigeons are able to discriminate the categories of scenes at least as quickly as they have been previously shown to discriminate much simpler shapes and colors, even though the scenes in the present study were visually more complex and contained richer visual information (e.g., Hansen & Hess, 2007). One factor that may have promoted performance in the present study is that we used blocked ascending presentation of the stimulus durations rather than randomly intermixed exposure. The intermixed exposure in Kramer's study may have interfered with the pigeons' ability to adapt to the stimulus exposure time, whereas our pigeons were able to do this and often showed improvements in performance after some experience on a particular duration.

Perhaps it should come as no surprise that pigeons can recognize the gist of complex real-world scenes. Pigeons presumably need to make decisions in flight and detecting gist would aid in the determination of possible food and water sources, roosting locations, and so forth. In addition, it has been shown in humans that scene gist aids in directing attention to task-relevant regions of a scene (Eckstein et al., 2006; Torralba et al., 2006) and may facilitate rapid object identification (Davenport & Potter, 2004; Palmer, 1975). Determining the gist of a scene would provide a survival advantage in terms of identifying predators or food items within a scene. Therefore, further studies should examine whether pigeons are able to use gist information to direct their attention and promote object identification. Other studies should also compare the time course of gist versus object identification in pigeons.

Although the pigeons' performance in the present study was fairly impressive, it remains the case that the birds could not successfully achieve scene gist categorization within the same time scale as humans. This difference in the minimum stimulus duration needed to recognize the gist of scenes by pigeons and humans shows an interesting parallel to the average durations of eye fixations between the two species. For humans, the minimum unmasked image duration needed for asymptotic scene gist recognition is 0.020–0.025 s (Bacon-Macé et al., 2005; Fei-Fei et al., 2007; Loschky et al., 2007). However, in natural viewing, humans achieve asymptotic scene gist recognition within a single eye fixation (Eckstein et al., 2006; Torralba et al., 2006), which averages 0.33 s (Rayner, 1998). Conversely, for pigeons, we have shown that the minimum unmasked image duration needed for asymptotic scene gist recognition is 0.20–0.50 s, which is an order of magnitude longer than for humans. The two studies we know of that have measured pigeons' eye movements showed average fixation durations ranging from 1.30 s (Nye, 1969) to 5.70–8.30 s (Bloch, Rivaud, & Martinoya, 1984), an average of between 3.50 and 4.80 s, which is also an order of magnitude longer than humans. For humans, the time courses of both scene gist recognition and eye movements are intimately related, as shown by the role of scene gist recognition in directing eye movements. If the same is true for pigeons, then the initial evidence at hand suggests that natural visual processing by pigeons, as reflected by scene gist processing and eye movements, takes roughly an order of magnitude longer than that for humans. Thus, the current findings suggest that although humans and pigeons may process scene gist on different absolute time scales, they both are able to categorize real-world scenes within the timeframe of a single eye fixation, which is particularly interesting given that the two species are highly divergent within the evolutionary tree.

There are several possible explanations for this discrepancy in the time needed for scene gist recognition between the two species. One possibility is that greater binocularity in humans may allow for faster and more accurate recognition of images presented in the frontal plane. In a related vein, it is possible that differences in hemispheric specialization could influence the speed of recognition. Pigeons do not possess ipsilateral fibers because their eyes are on either side of their head (Remy & Watanabe, 1993). The location of the eyes could itself present challenges as the pigeons may turn their head to obtain a more focal view for one of their eyes and this could have time costs. Finally, it is possible that the myopic frontal field of the pigeon visual system, which appears to be specialized for foraging, may not be as sensitive to scene gist information as the hyperopic lateral visual field, which is used for near-panoramic perception during flight (Nye, 1969; Remy & Watanabe, 1993; Roberts, Phelps, Macuda, Brodbeck, & Russ, 1996). Considering that the images were presented to the frontal field, we may not have capitalized fully on the pigeons' scene gist recognition capabilities, which may emerge more naturally from the lateral visual system. Future studies should assess these and other potential issues.

Lastly, regarding the natural–natural versus natural–manmade distinction, there was an apparent tendency for pigeons in the natural–natural group to take longer than the natural–manmade group to reach criterion during original learning, but this was not statistically significant (see Figure 3). Nevertheless, it is clear that there was considerably greater variability between pigeons in the natural–natural condition relative to the natural–manmade condition. Likewise, although pigeons in the natural–natural group showed a generalization decrement that was twice that of pigeons in the natural–manmade group, this difference between the groups during generalization testing (see Figure 4), which was conducted with the stimulus duration of 5 s, was not statistically significant. However, there was some indication of superior performance in the natural–manmade group under processing time challenges (see Figure 5). This mirrors the results with human participants, where

categorical distinctions between two members of the same superordinate category are more difficult than distinctions between members of two different superordinate categories, particularly with shorter processing times (Joubert et al., 2009; Loschky & Larson, 2010; Macé et al., 2009). However, given that we compared only a single pair of categories, and that the differences were not robust, further work will be needed to confirm the generality of this difference across a wider range of categories. In addition, Lazareva, Soto, and Wasserman (2010) showed that pigeons demonstrate a superordinate-level categorization advantage if the basic-level categories are similar, but the opposite pattern if the basic-level categories are very distinct. This suggests that the discriminability of different scene categories should be considered when making such comparisons.

Experiment 2

As mentioned in the introduction, the incredible speed of scene categorization by humans has led some to argue that scene category discrimination is based on critical low-level image statistical regularities. This raises the question of whether the same argument may also hold for pigeon scene categorization, which Experiment 2 investigated.

There is a long history of describing the low-level properties of complex images in the spatial domain (i.e., across the two-dimensional [*x*, *y*] coordinate space of an image) in terms of the statistical relationships between pixel luminance values (Julesz, Gilbert, Shepp, & Frisch, 1973; Julesz, Gilbert, & Victor, 1978; Klein & Tyler, 1986; Thomson, 1999). These statistical relationships are described broadly in terms of their order, with most research concerned with the second-order versus higher order statistics (e.g., third order and higher). Specifically, the pixel luminance histogram (i.e., the probability distribution of luminance) is referred to as the *first moment* (or *first-order statistic*) of the image. The degree of correlation between pairs of pixels as a function of all possible physical distances is referred to as the *second-order statistic*, which can be conceived of as the variance of the pixel luminance distribution, and constitutes the luminance contrast of the image. The relationship between more than two pixel luminance values (e.g., pixel triplets or quadruplets) is referred to as the *higher order statistics* of an image. These higher order statistics have been shown to carry information about the lines, edges, and luminance boundaries that form the image structures that provide the meaningful content of scenes (Thomson, 1999, 2001; Thomson & Foster, 1997).

One historical approach to quantifying the image statistics of two-dimensional (2-D) scene images comes from the global 2-D discrete Fourier transform (DFT). The 2-D DFT treats an image as a complex 2-D luminance contrast waveform, which can be represented as the sum of sinusoidal waveforms of different amplitudes (i.e., contrasts), frequencies, orientations, and phases (referenced in the Fourier domain, which is the linear transformation of the image in the 2-D spatial domain). The amplitude plotted as a function of spatial frequency and orientation is often referred to as the *amplitude spectrum*.[1] Importantly, the distribution of amplitude as a function of spatial frequency and orientation in the Fourier domain is a direct assessment of the degree of correlation between the luminance values of all possible pixel pairs in the spatial (i.e., 2-D image) domain. Therefore, the Fourier amplitude spectrum is a direct measurement of the second-order statistical relationships of the pixel luminance values in scene images.

Although much important information about scene content can be assessed by the second-order statistical relationships of scene images, recent work (e.g., Hansen & Loschky, 2013; Loschky, Hansen, et al., 2010) has shown that second-order image statistics by themselves cannot explain other crucial physical and perceptual attributes of real-world images. Specifically, such work has argued for considering the relative contribution of both second- and higher order image statistics. In the current experiment, we employed one useful way of quantifying the higher order image statistics of scenes, known as the *phase-only second spectrum* (which is directly calculated from manipulations of the DFT; Thomson, 1999). The edges and lines that make up scene images have been shown to arise from phase alignments across a wide range of spatial frequencies (Morrone & Burr, 1988; Morrone & Owens, 1987), and the phase-only second spectrum (Thomson, 2001) is very sensitive to such alignments. Here, we sought to explore the relative contribution of both second- and higher order image statistics in pigeon scene categorization.

In addition, as an avian species, pigeons in the wild spend considerable time in flight, in perching locations looking down on the ground, and on the ground foraging for food, and would presumably need to categorize scenes that afford different survival-relevant functions. This could conceivably have led them to evolve a preparedness to rapidly acquire the ability to recognize a range of views of natural scenes. Thus, if pigeons make use of second- and higher order image statistical regularities, they would likely do so across such a variety of viewpoints. In the current experiment, we used scene categories from three different viewpoints, namely *terrestrial* (used in Experiment 1; 0° from the ground plane), *bird's eye* view (45° from the ground plane, e.g., a view from an approaching flight path), and *zenith* view (90° from the ground plane, e.g., a view from high above when in flight). The pigeons' relative ability to recognize scenes from terrestrial, zenith, or perhaps most importantly, bird's eye views (45° from the ground plane) has not previously been assessed. To ensure that their previous visual experience would not be confounded with any possible species-specific predispositions, we used only lab-reared pigeons with no outdoor flight experience.

Method

Subjects. The subjects were eight mixed-sex, captive-bred homing pigeons (Double-T Farms) that were 1–4 years of age at the start of the experiment. Four of the birds had participated in Experiment 1 (Y2, Y4, Y6, and Y9), and four birds were experimentally naïve (Y1, Y13, Y16, and Y18). The housing and husbandry conditions were the same as those in Experiment 1.

Apparatus. The apparatus was the same as in Experiment 1.

Stimuli. The stimuli consisted of beach and mountain images that were created to the same specifications as in Experiment 1, except that there were three viewpoints for each category. The terrestrial viewpoints were from the same set as Experiment 1. The bird's eye images for the beach and mountain categories were gathered from Bing maps using a 45° bird's eye image tool. The bird's eye

[1] Alternatively, it is plotted as the power spectrum (i.e., the amplitude spectrum squared).

beach category images were representative of both sandy and rocky shorelines. The bird's eye and zenith mountain images were captured in the same manner, but the altitude was increased to allow for mountains to be recognizable and account for ground-level differences between the ranges. Zenith beach and mountain images were collected using Google Earth's Satellite imagery. The terrestrial category images were from the set of images used in Experiment 1; these images were presented to four of the pigeons previously. Given the limitations of the Website and software used, the exact altitudes of the images could not be determined. Sample images from each viewpoint are displayed in Figure 6. There were 20 images from each viewpoint for each category (S+ and S–) used during the training and intermixed phases.

Procedure.

Pretraining. Pretraining was conducted in the same fashion as in Experiment 1 for the four experimentally naïve birds. The remaining birds were transferred directly into the discrimination training phase.

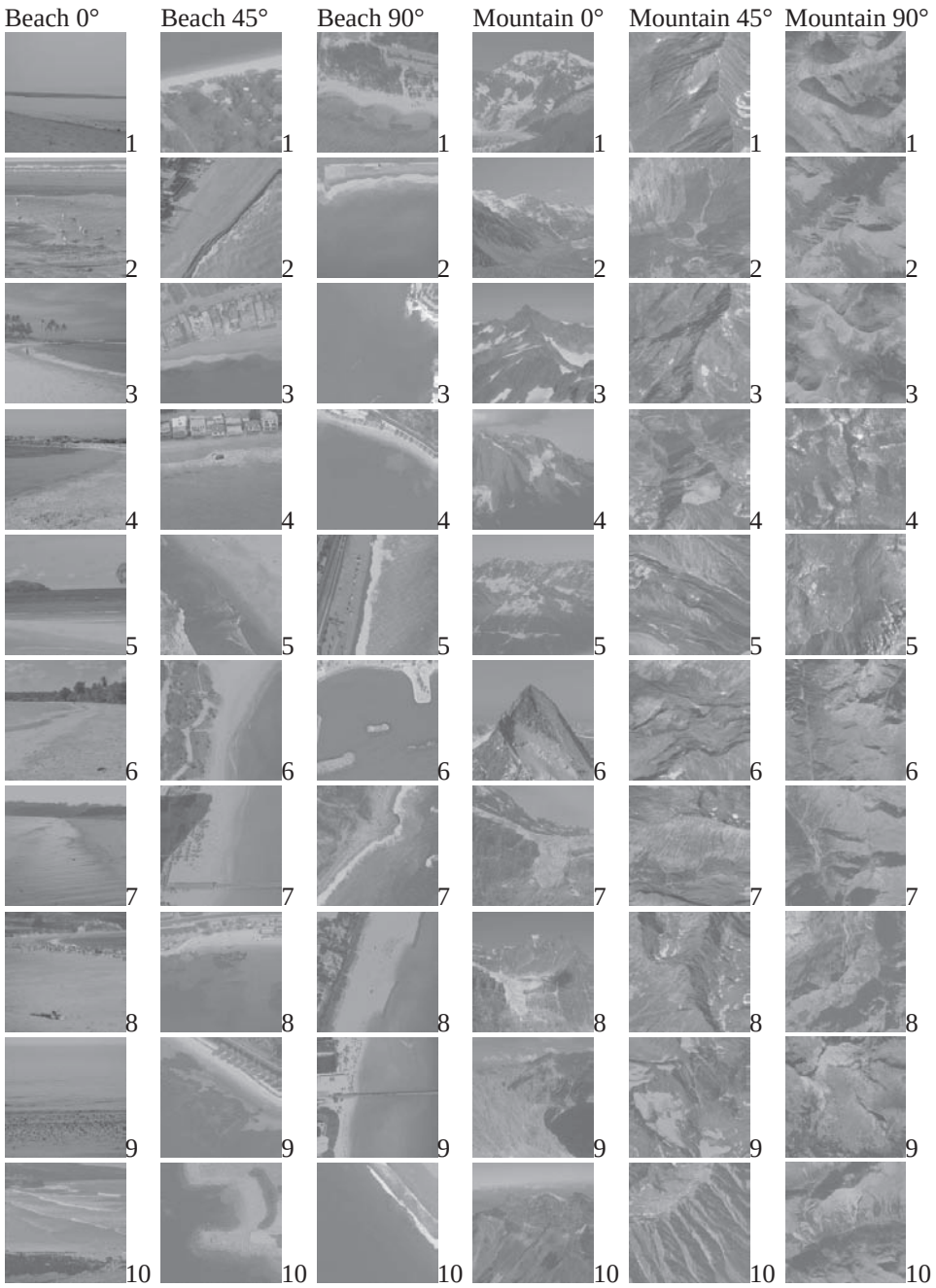


Figure 6. A representative subset of the training images from the terrestrial (0°), bird's eye (45°), and zenith (90°) viewpoints.

Discrimination training. The pigeons were trained with the go/no-go procedure to discriminate between beaches versus mountains. Discrimination training was conducted in the same fashion as in Experiment 1 (see Figure 2), except that the stimulus duration was 2 s and the go/no-go response duration was 3 s. Each bird received training with each of the three viewpoints (terrestrial, bird's eye, and zenith) in a counterbalanced order. The training orders are presented for each bird in Table 2. There were 20 exemplars from each category used in the training phase. As in Experiment 1, there were a total of 200 trials per session, 100 S+ and 100 S–, presented in five 40-trial blocks. The pigeons were trained until producing a DR of at least 0.80 for two consecutive sessions, at which point they were transferred to the new viewpoint until all three viewpoints had been trained.

Intermixed training. Following training on each of the three viewpoints individually, the birds received training with the viewpoints intermixed in the same session. This training phase allowed for a more direct comparison of the performance for the three viewpoints. The same go/no-go contingency was in place as in the discrimination training phase. The birds experienced all 120 images from the previous phase within each session (20 S+ and 20 S– for each of three viewpoints), with each image delivered twice, for a total of 240 trials per session, 120 S+ and 120 S– per category. There were 20 total sessions of intermixed training.

Image analysis. All scene images were subjected to Fourier filtering algorithms in MATLAB (Version R2011b), using functions contained within its Signal Processing and Image Processing Toolboxes (Versions 6.16 and 7.3, respectively). To avoid creating image artifacts ("edge effects"), prior to filtering, we fit each image with a circular edge-ramped window (ramped to mean luminance, window diameter = 512 pixels) and subjected it to a DFT. The resulting power and phase spectra were shifted into polar coordinates.

For the analysis of second-order statistics, we first generated 60 one-octave wide (full width at half height) log-Gabor filters in the Fourier domain (see Hansen & Hess, 2007, for details). Specifically, five sets of log-Gabors were constructed, with each set centered on a different spatial frequency. The central spatial frequencies (which had a peak-to-peak difference of 1 octave) were as follows: 0.25, 0.50, 1.00, 2.00, and 4.00 cycles per degree of visual angle. The particular central spatial frequencies were selected such that they evenly tiled the entire contrast sensitivity function of the pigeon visual system, which is a 4-octave-wide band-pass function centered on 1.0 cycles per degree (e.g., Hodos, Ghim, Potocki, Fields, & Storm, 2002). Each set of log-Gabors consisted of 12 filters (all centered on a specific spatial frequency), each centered on a different orientation, ranging from 0° to 165° in steps of 15°. Thus, our bank of 60 log-Gabor filters evenly tiled the entire visible range of spatial frequencies across both spatial frequency and orientation of the pigeon visual system. Next, each scene image's power spectrum was filtered with each one of the 60 different log-Gabor filters in the Fourier domain. After filtering with a given log-Gabor, the filtered spectrum was summed to yield the contrast energy at that particular spatial frequency and orientation, with the resulting value stored in a 5×12 matrix with spatial frequency represented on the *y*-axis and orientation on the *x*-axis. The resulting matrix therefore corresponds to a second-order feature spectrum (Torralba & Oliva, 1999, 2003), and can be considered to approximate the response of different spatial frequency and orientation-tuned neurons in the early visual system of the pigeon. We created a second-order feature spectrum for each image in our stimulus set.

To create a metric that would reflect the relative second-order feature spectra differences, both *within* each scene category as well as *between* scene categories, we calculated a between-to-within category difference ratio (BWDR) for each scene category within each one of our three viewpoints. For simplicity, we describe the procedure for the two scene categories within one viewpoint (the procedure was carried out for each viewpoint). First, a between-categories difference metric was calculated for each beach image relative to each image in the mountain category. This was done by taking the sum of the squared differences between a given beach image's second-order feature spectrum and that of another image from the mountain category (with all second-order feature spectra converted to log values prior to this operation). The process was repeated for that particular beach image and every image in the mountain category. The between-categories difference metric for that particular beach image was then created by averaging across all summed and squared differences relative to every image in the mountain category. The process was then repeated for every image in the beach category. Finally, the same procedure was carried out for the images in the mountain category relative to the images in the beach category. Next, a within-category difference metric was calculated in exactly the same manner, except that each beach image's second-order feature spectrum was differenced with each of the other images' spectra within the beach category. The same was then conducted for the mountain category. Finally, the BWDR was calculated for each image by dividing its between-categories difference metric by its within-category difference metric. Thus, the BWDR reflects an estimate of the dissimilarity between images from different categories while factoring out the within-category variability. If pigeons rely on second-order features to discriminate between images at a particular viewpoint, we would expect greater discriminability for images with larger second-order BWDRs.

For the analysis of higher order statistics associated with edges and boundaries, we used the DFT to generate a phase-only second spectrum for each image in our stimulus set (see Loschky, Hansen, et al., 2010, Appendix for details). The phase-only second spectrum offers a global assessment of "edge strength" in terms of higher order statistical image structure, as measured by the strength of sinusoidal fluctuations (i.e., signal variance) as a function of different spatial frequency offsets (Thomson, 2001). Edges and boundaries are a critically important source of information for

Table 2
Order of Training for the Three Viewpoints, Terrestrial (0°), Bird's Eye (45°), and Zenith (90°) for Each Pigeon in Experiment 2

Pigeon	Phase 1	Phase 2	Phase 3
Y2	90°	45°	0°
Y3	90°	45°	0°
Y4	90°	0°	45°
Y6	0°	90°	45°
Y9	0°	90°	45°
Y13	0°	45°	90°
Y16	45°	0°	90°
Y18	45°	90°	0°

discriminating natural images, and they are created when a wide range of spatial frequencies at a given orientation are in phase (i.e., aligned) at a given spatial location. Such phase alignment is measured by the phase-only second spectrum because, for example, a large value shows the presence of a significant interaction among a number of sinusoidal modulations by a particular offset. Thus, the phase-only second spectrum assesses the degree of edge strength in terms of phase alignment across all spatial frequencies in an image. To create different higher order feature spectra that would possess "edge-strength" information for each image, we filtered each image's phase-only second spectrum with a bank of 60 log-Gabor filters (see above) to create a higher order feature spectrum for each image in our stimulus set. Lastly, higher order BWDRs were calculated in an identical manner to that described in the second-order analysis section. Thus, if pigeons rely on higher order edge or boundary features to discriminate between images at a particular viewpoint, we would expect greater discriminability for images with larger higher order BWDRs.

Results

Intermixed training performance. The data analysis focused on the intermixed training phase in which all three viewpoints were experienced concurrently. A DR was calculated for each bird's performance on each viewpoint, and these were analyzed over the first half versus second half of training to assess any training effects on performance. Figure 7 (top panel) displays these DRs as a function of viewpoint. In examining the figure, it appears that the 0° (terrestrial) viewpoint was discriminated less well than the 45° and 90° viewpoints, but that the discrimination improved over training to become more similar across viewpoints. Because half of the birds were naïve and half were experienced with terrestrial views from Experiment 1, the bottom panel of the figure presents the results as a function of experience. As seen in the figure, both naïve and experienced birds demonstrated the same general DR patterns when responding to the different viewpoints, but the experienced birds had higher overall DRs. An ANOVA was conducted with the variables of training (first vs. second half), viewpoint (0°, 45°, or 90°), and experience (naïve vs. experienced birds). This revealed a main effect of training, $F(1, 6) = 6.1$, $p = .049$, $\eta_p^2 = .50$, power = .54, 95% CI [.01, .07], and a Training × Viewpoint interaction, $F(2, 12) = 9.3$, $p = .004$, $\eta_p^2 = .61$, power = .93. There was no main effect of viewpoint, $F(2, 14) = 2.5$, $p = .128$, no main effect of experience, $F(1, 6)$ 1.3, $p = .292$, or any interactions with experience: Training × Experience, $F(1, 6) = 1.0$, $p = .348$; Viewpoint × Experience, $F(2, 12) = 1.2$, $p = .344$; Training × Viewpoint × Experience, $F(2, 12) < 1$. Post hoc tests using a Bonferroni correction procedure on the Training × Viewpoint interaction indicated that the 45° and 90° viewpoints were discriminated significantly better during the first half of training than the 0° viewpoint, 0° versus 45°: $t(7) = 2.9$, $p = .024$, $d = 1.00$, 95% CI = [0.02, 0.16]; 0° versus 90°: $t(7) = 2.8$, $p = .028$, $d = 1.01$, 95% CI = [0.01, 0.16]; 45° versus 90°: $t(7) = 0.3$, $p = .809$. There were no differences in DRs in the second half of training, largest $t(7) = 0.9$, $p = .386$.

Image analysis. To investigate the image statistical differences between beach and mountain images, based on our model of early pigeon visual areas' response to the images, we ran independent *t* tests separately for second- and higher order BWDRs between beaches and mountains within each viewpoint. All were found to be significantly different, smallest $t(38) = 4.4$, $p < .001$, $d = 1.38$, 95% CI [0.59, 1.59]. Thus, for each type of low-level image statistic, there were sufficient differences in the BWDRs between the beach and mountain categories at each viewpoint to potentially allow pigeons to use such image statistical regularities.

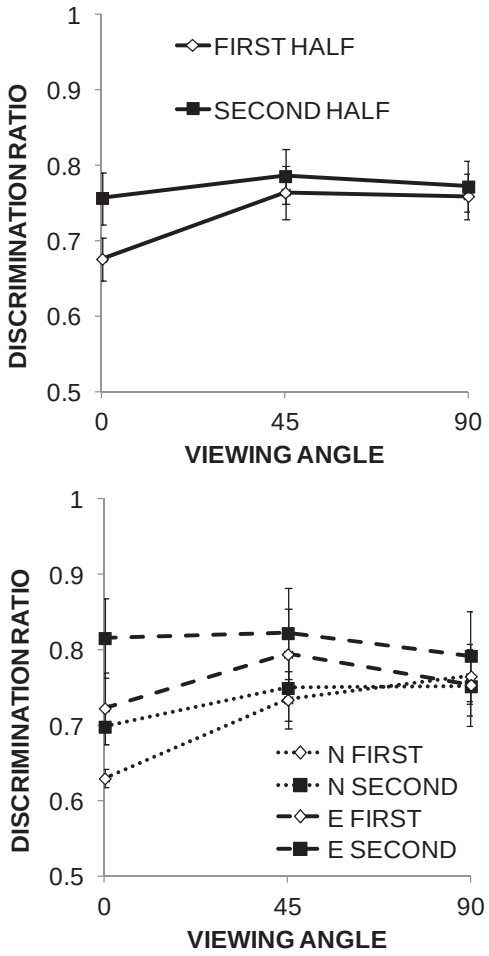


Figure 7. Top panel: Mean (±*SEM*) discrimination ratios as a function of stimulus viewing angle during the first versus second half of the intermixed viewpoint training phase. Bottom panel: Mean (±*SEM*) discrimination ratios as a function of stimulus viewing angle for naive (N) and experienced (E) birds during the first and second half of training.

However, it is important to note that statistically significant differences between physical measures of image luminance characteristics do not necessarily translate to perceivable differences. For example, although a set of sinusoidal luminance gratings with Michelson contrasts of 80% versus 82% would certainly register as a statistically significant physical difference in contrast, the two sets would be perceptually identical. Thus, we ran a stepwise multiple regression analysis to test whether either or both of the second- and higher order BWDRs could account for overall pigeon categorization performance. Pigeon performance was measured by the DR (described in Experiment 1), which served as an index of image discriminability from the pigeons' perspective.

As a first global analysis, we carried out a standard regression analysis of averaged pigeon DR, across the three views, against

both second- and higher order BWDRs for each category, beach and mountain.[2] Figure 8 shows pigeon DRs versus predicted DRs on the basis of the conjoint image statistical BWDRs. Figure 8 shows that, overall, simple image statistical discriminability based on both second- and higher order global features accounted for ~27% of the variance in pigeons' DRs for the beach category and ~47% of the variance in DRs for the mountain category. Thus, it is plausible that the pigeons' difference in performance across the three views, as shown in Figure 7, is at least partly explained by image statistics, which show larger differences between the scene categories for the zenith and bird's eye views than for the terrestrial views.

Given the positive results reported in Figure 8, we conducted a more detailed multiple regression analysis in which the DRs for each beach and mountain image of each viewpoint were entered into a stepwise multiple regression analysis (alpha for entry = .05), using the second-order, higher order, or both BWDRs as predictors. The results, shown in Figure 9 panels a–c, indicated that second- and higher order BWDRs accounted for sizable portions of pigeon DR variance in five of the six conditions tested. Specifically, (a) for zenith views of beaches, both second- and higher order BWDRs were significant predictors ($R^2 = 0.73$, higher order $p < .001$, second-order $p = .012$); for zenith views of mountains, only second-order BWDRs were entered as predictors ($R^2 = 0.35$, $p = .006$). (b) For bird's eye views of beaches, only higher order BWDRs were entered as predictors ($R^2 = 0.27$, $p = .011$); for terrestrial views of beaches, only higher order BWDRs were entered ($R^2 = 0.28$, $p = .017$). (c) For terrestrial views of mountains, only second-order BWDRs were entered ($R^2 = 0.45$, $p = .001$); the bird's eye views of mountains could not be explained by either second- or higher order BWDRs.

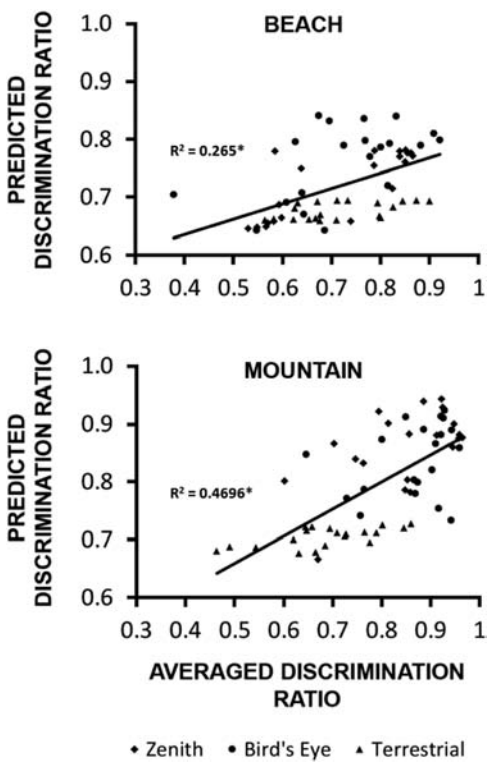


Figure 8. Scatterplots from Experiment 2. Diamonds = zenith view; circles = bird's eye view; triangles = terrestrial view. The top panel shows scatterplot of predicted versus observed averaged pigeon discrimination ratios across the three views for beach images. Predictions are from a stepwise multiple regression analysis after factoring in either second-order, higher order, or both between-to-within category difference ratios. The bottom panel shows the same for mountain images.

To provide further detail regarding the above relationships, we also calculated separate Pearson correlation coefficients for each of the six conditions of the current experiment between pigeon discrimination ratios and second-order (see Figure 9 panels d–f) and higher order (see Figure 9 panels g–i) BWDRs. As shown in Figure 9 panels d–f, for second-order BWDRs, significant correlations were found for zenith beaches ($r = .50$, $p = .024$), zenith mountains ($r = .59$, $p = .006$), and terrestrial mountains ($r = .67$, $p = .001$). Figure 9 panels g–i show that for higher order BWDRs, significant correlations were found for zenith beaches ($r = .78$, $p < .001$), bird's eye beaches ($r = .56$, $p = .011$), and terrestrial beaches ($r = .53$, $p = .017$). Thus, overall, it seems that the second-order BWDRs were more useful for the mountain category and the higher order BWDRs were more useful for the beach category.

Discussion

Experiment 2 showed that, just as with human rapid scene categorization, a considerable amount the variance in our pigeons' DRs could be explained in terms of the discriminability of the scene categories in relation to their lower level image statistics. We found that second- and higher order statistical regularities could account for a sizable portion of pigeon scene categorization performance in five of the six conditions in the current experiment. This is consistent with recent work (e.g., Loschky, Hansen, et al., 2010) reporting that humans make use of both second-order and higher order scene statistical regularities during rapid scene categorization. However, whereas that previous work showed that higher order statistics tend to contribute more to rapid scene categorization in humans, here we find that the relationship seemed to depend on scene category. Specifically, regardless of viewpoint, second-order scene regularities accounted for pigeon categorization performance more for the mountain category, perhaps due to oriented global contrast differences playing a larger role (e.g., the oblique angles of mountains being very different from the horizontal orientation of beaches). Conversely, higher order scene regularities accounted for pigeon categorization performance more for the beach category, perhaps due to differences in oriented edges and boundary structure playing a larger role. However, it is important to note that the previous work exploring the use of image statistics in rapid scene categorization in humans used many more scene categories. Thus, the relative contribution of second-order and higher order scene statistics may depend on the specific image category.

[2] Averaged DRs were calculated for each image by determining the mean number of pecks to each stimulus when it served as an S+ and when it served as an S–, and this served as an index of image discriminability from the pigeons' perspective. Note that this is the same equation as the DRs reported elsewhere, except that the DR in this case is bidirectional across S+ and S– categories.

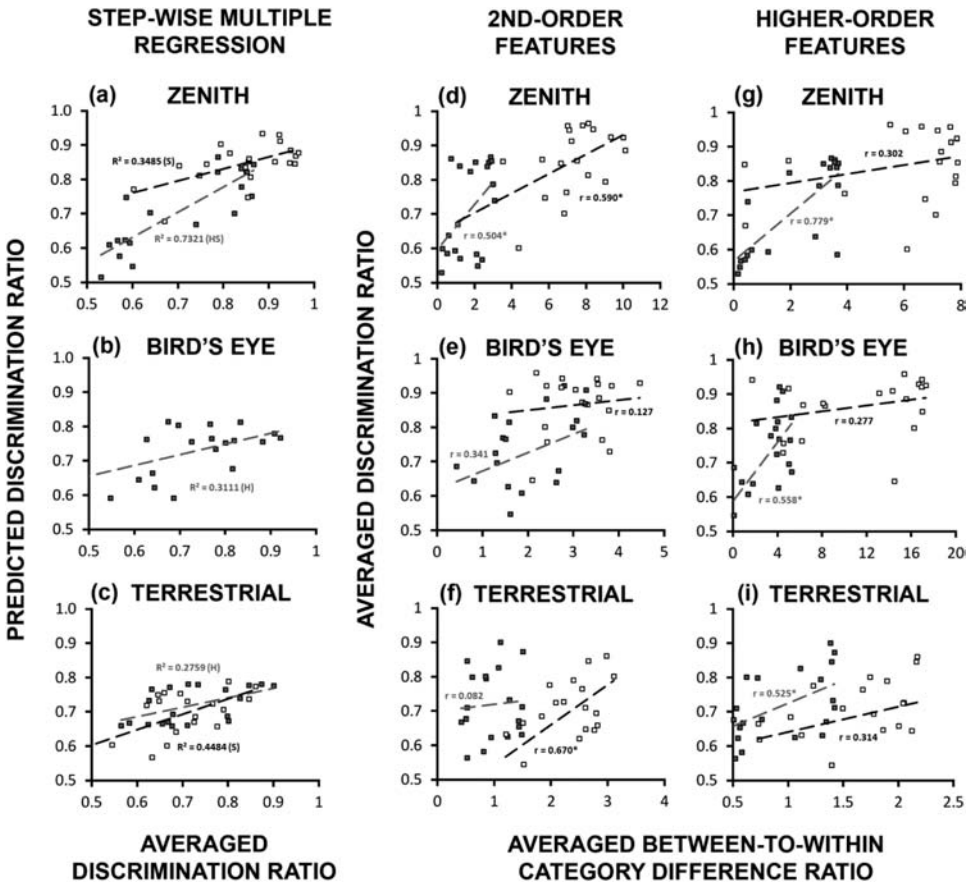


Figure 9. Scatterplots from Experiment 2. Gray squares designate the beach category, and white squares the mountain category. Panels a–c show scatterplots from the stepwise multiple regression analysis. On the abscissa is the averaged pigeon discrimination ratio for each image, with the stepwise multiple regression predicted discrimination ratio after factoring in either second-order, higher order, or both between-to-within category difference ratios (BWDRs). The R^2 for each analysis is shown in each panel, where (S) = only second-order BWDRs entered, (H) = only higher order BWDRs entered, and (HS) = both second- and higher order BWDRs entered. Panels d–f show scatterplots between averaged pigeon discrimination ratio for each image (ordinate) and averaged second-order BWDR for each image (abscissa). Panels g–i show scatterplots between averaged pigeon discrimination ratio for each image (ordinate) and averaged higher order BWDR for each image (abscissa).

Furthermore, the observed viewpoint effect, wherein pigeons were better at discriminating the bird's eye (45°) and zenith (90°) viewpoints than the terrestrial (0°) viewpoint, in the first half of the intermixed discrimination training, was consistent with the image statistical discriminability of the categories from those views. Thus, the current results show that the use of image statistics to explain scene categorization applies to pigeons as well as humans.

Overall, such second- and higher order image statistical regularities explained ~27% of the pigeons' variance in discriminating beach images and ~47% of their variance in discriminating mountain images (see Figure 8). Although impressive, particularly for the mountain images, this also means that ~73% of the pigeons' variance in their discrimination of beach images and ~53% of the variance in their discrimination of mountain images could not be captured by our image statistical analyses. This raises the following question: What might explain the remaining variance? One possibility is that our image statistical analyses were simply not sophisticated enough to capture the richness of the information contained in the natural images and used by the pigeons to discriminate them. However, with regard to the observed viewpoint effect, another possibility is that beyond the information contained in the images themselves, pigeons may have evolved a highly conserved predisposition to more easily learn to discriminate zenith and birds' eye views than terrestrial views. This might explain the difference between our pigeons' slower learning to discriminate terrestrial scenes versus the fact that humans are far better at discriminating terrestrial (0°) views than zenith (90°) views (e.g., Hansen & Loschky, 2013; Loschky, Ellis, Sears, Ringer, & Davis, 2010; Loschky, Ringer, Ellis, & Hansen, 2013). This is consistent with the role of evolved behavioral predispositions interacting with the environment such that some stimulus category discriminations are more easily learned than others. For example, M. Cook and Mineka (1989) found that lab-reared monkeys showed rapid observational learning of fear of evolutionarily fear-relevant stimuli, a toy snake and toy crocodile, but not evolutionarily fear-irrelevant stimuli, a flower and a toy rabbit. Our pigeons, too, were lab-

reared, and thus their more rapid discrimination learning of aerial (45° and 90°) views than terrestrial (0°) views of scenes cannot be explained in terms of their prior experience. However, the pigeons did learn to improve their performance of the terrestrial views over the course of training. This indicates that there is a role for experiential effects in addition to any possible evolutionary predisposition, an issue that would be interesting to explore in human populations as well.

General Discussion

The results of Experiments 1 and 2 demonstrate that pigeons, like humans, can categorize real-world scenes after presentation durations of less than a single eye fixation. Given that pigeons and humans diverged in the evolutionary tree prior to the age of the dinosaurs (Jarvis et al., 2005), this suggests that the ability to recognize scene gist is either based on homology due to shared evolutionarily ancient brain structures or that shared selective pressures in the environment have pushed both species to a homoplastic shared ability using different brain structures. In either case, these results suggest that the ability to recognize scene gist may be widely shared across species. Further studies with other divergent species could test this hypothesis.

With regard to scene gist recognition by pigeons, we must note that the current study showed that pigeons could learn to accurately discriminate between real-world scene categories, not that they necessarily interpreted those scenes at a semantic level. The degree to which pigeons' successful picture categorization implies that they understand the real-world referents of those pictures is a much-debated topic (e.g., Aust & Huber, 2006; Fagot, 2000). Furthermore, a strong argument could be made that because our pigeons were lab-reared and had never been exposed to real-world beaches, mountains, and streets, their picture discrimination was almost certainly devoid of semantic content. However, based on the current study, pigeons do appear able to make the sorts of perceptual discriminations necessary to rapidly categorize real-world scenes. For pigeons in the wild, having the ability to rapidly discriminate different real-world scene categories, combined with actual experience with behavioral outcomes associated with those scene categories, would likely result in learning to map between scenes' perceptual and semantic discriminations.

The results of Experiment 1 also showed that pigeons differ from humans in that they required stimulus durations roughly an order of magnitude longer to rapidly categorize scenes. The cause of this difference is as yet unknown, but it is likely to be due to the different structure of pigeons' visual systems from that of humans (Kirkpatrick, 2001; Soto & Wasserman, 2012). Of potential importance, however, is the fact that pigeons not only require longer stimulus durations than humans to rapidly discriminate scene images, but they also appear to make much longer eye fixations than humans (Bloch et al., 1984; Nye, 1969). Given that scene gist recognition occurs within a single eye fixation (Eckstein et al., 2006; Torralba et al., 2006) and that fixation durations are generally tied to the processing times required for various visual recognition tasks (Nuthmann, Smith, Engbert, & Henderson, 2010; Rayner, 1998), it seems a reasonable assumption that pigeons may both require longer stimulus durations to recognize gist and tend to make longer fixation durations than humans, both by roughly an order of magnitude.

Experiment 2 showed that pigeons, like humans, appear to make use of complex second- and higher order image statistical regularities. Interestingly, because pigeons need to dissociate scene categories from a number of different viewpoints, the results from Experiment 2 suggest that they may indeed make use of the same image statistical regularities across those variable viewpoints. Furthermore, it appears that pigeons may find certain scene statistical regularities more informative depending on the category. Of course, such an observation is limited by the fact that only two scene categories were used in Experiment 2, and future work should expand on this by including more image categories.

Nevertheless, the image statistical analyses left a considerable amount of unexplained variance. An interesting hypothesis to test in further research is that pigeons are prepared to more easily categorize aerial views of scenes, namely bird's eye or zenith views, rather than terrestrial views of scenes, whereas the opposite is true for humans. Our results of Experiment 2 provide some suggestive evidence that this might be the case, but we should be cautious before accepting it. First, the image statistical analyses explained much of the variance between views, albeit leaving a great deal of unexplained variance. Second, the limited number of scene categories may have played a role in producing this effect. Nevertheless, the viewpoint effect found in Experiment 2 cannot be explained by the pigeons' experiences because all of the pigeons were lab-raised and had no outdoor flight experience.

Together, the results suggest that two highly divergent species, pigeons and humans, show a shared functional capacity to rapidly categorize real-world scenes, which may well be shared across a wide range of species, but that their divergent evolutionary histories have led to adaptive specializations producing important predispositions and constraints on their ability to recognize scene gist. To our knowledge, this is the first study to have shown such a pattern of both shared and divergent capacities across species to recognize the gist of real-world scenes, and it points to a rich direction for further research in this area of comparative cognition.

References

Aust, U., & Huber, L. (2006). Picture-object recognition in pigeons: Evidence of representational insight in a visual categorization task using a complementary information procedure. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 190–195. doi:10.1037/0097-7403.32.2.190

Bacon-Macé, N., Macé, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research, 45,* 1459–1469. doi:10.1016/j.visres.2005.01.004

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F. J., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorisation of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 219–234. doi:10.1037/0097-7403.14.3.219

Bloch, S., Rivaud, S., & Martinoya, C. (1984). Comparing frontal and lateral viewing in the pigeon: III. Different patterns of eye movements for binocular and monocular fixation. *Behavioural Brain Research, 13,* 173–182. doi:10.1016/0166-4328(84)90147-5

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 207–230. doi:10.1016/0010-0285(81)90008-6

Cook, M., & Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology, 98,* 448–459. doi:10.1037/0021-843X.98.4.448

Cook, R. G. (2001). *Avian visual cognition*. Medford, MA: Comparative Cognition Press. Retrieved form http://www.pigeon.psy.tufts.edu/avc/

Cook, R. G., & Blaisdell, A. P. (2006). Short-term item memory in successive same-different discriminations. *Behavioural Processes, 72,* 255–264. doi:10.1016/j.beproc.2006.03.013

Cook, R. G., Cavoto, B. R., Katz, J. S., & Cavoto, K. K. (1997). Pigeon perception and discrimination of rapidly changing texture stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 23,* 390–400. doi:10.1037/0097-7403.23.4.390

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15,* 559–564. doi:10.1111/j.0956-7976.2004.00719.x

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17,* 973–980. doi:10.1111/j.1467-9280.2006.01815.x

Fagot, J. (2000). *Picture perception in animals*. Hove, UK: Psychology Press.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7,* 1–29. doi:10.1167/7.1.10

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12,* 893–924. doi:10.1080/13506280444000571

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience, 8,* 1643–1646. doi:10.1038/nn1608

Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biology, 3,* e342. doi:10.1371/journal.pbio.0030342

Hansen, B. C., & Hess, R. F. (2007). Structural sparseness and spatial phase alignment in natural scenes. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 24,* 1873–1885. doi:10.1364/JOSAA.24.001873

Hansen, B. C., Jacques, T., Johnson, A. P., & Ellemberg, D. (2011). From spatial frequency contrast to edge preponderance: The differential modulation of early visual evoked potentials by natural scene stimuli. *Visual Neuroscience, 28,* 221–237. doi:10.1017/S095252381100006X

Hansen, B. C., Johnson, A. P., & Ellemberg, D. (2012). Different spatial frequency bands selectively signal for natural image statistics in the early visual system. *Journal of Neurophysiology, 108,* 2160–2172. doi:10.1152/jn.00288.2012

Hansen, B. C., & Loschky, L. C. (2013). The contribution of amplitude and phase spectra defined scene statistics to the masking of rapid scene categorization. *Journal of Vision, 13,* 1–21. doi:10.1167/13.13.21

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 285–302. doi:10.1037/0097-7403.2.4.285

Hodos, W., Ghim, M. M., Potocki, A., Fields, J. A., & Storm, T. (2002). Contrast sensitivity in pigeons: A comparison of behavioral and pattern ERG methods. *Documenta Ophthalmologica, 104,* 107–118.

Huber, L., Roje, N. F. T., Loidolt, M., Aust, U., & Grass, D. (2000). Natural categorization through multiple feature learning in pigeons. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 53(B),* 341–357. doi:10.1080/027249900750001347

Jarvis, E. D., Gunturkun, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., . . . Butler, A. B. (2005). Avian brains and a new understanding of vertebrate evolution. *Nature Reviews Neuroscience, 6,* 151–159. doi:10.1038/nrn1606

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision, 9,* 1–16.

Julesz, B., Gilbert, E. N., Shepp, L. A., & Frisch, H. L. (1973). Inability of humans to discriminate between visual textures that agree in second-order statistics—revisited. *Perception, 2,* 391–405. doi:10.1068/pp.020391

Julesz, B., Gilbert, E. N., & Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics, 31,* 137–140. doi:10.1007/BF00336998

Kirkpatrick, K. (2001). Object recognition. In R. G. Cook (Ed.), *Avian visual cognition* [Online]. Available from http://www.pigeon.psy.tufts.edu/avc/kirkpatrick/

Kirkpatrick, K., Wilkinson, A., & Johnston, S. (2007). Pigeons discriminate continuous versus discontinuous line segments. *Journal of Experimental Psychology: Animal Behavior Processes, 33,* 273–286. doi:10.1037/0097-7403.33.3.273

Kirkpatrick-Steger, K., Wasserman, E. A., & Biederman, I. (1996). Effects of spatial rearrangement of object components on picture recognition in pigeons. *Journal of the Experimental Analysis of Behavior, 65,* 465–475. doi:10.1901/jeab.1996.65-465

Kirkpatrick-Steger, K., Wasserman, E. A., & Biederman, I. (1998). Effects of geon deletion, scrambling, and movement on picture recognition in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 24,* 34–46. doi:10.1037/0097-7403.24.1.34

Klein, S. A., & Tyler, C. W. (1986). Phase discrimination of compound gratings: Generalized autocorrelation analysis. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 3,* 868–879. doi:10.1364/JOSAA.3.000868

Kramer, K. (2010). *Rapid discrimination and categorisation of briefly flashed stimuli in the pigeon* (Columba livia) (Unpublished doctoral dissertation). Universitat Wein, Vienna, Austria.

Lazareva, O. F., Soto, F. A., & Wasserman, E. A. (2010). Effect of between-category similarity on basic-level superiority in pigeons. *Behavioural Processes, 85,* 236–245. doi:10.1016/j.beproc.2010.06.014

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, USA, 99,* 9596–9601. doi:10.1073/pnas.092277599

Loschky, L. C., Ellis, K., Sears, T., Ringer, R., & Davis, J. (2010). Broadening the horizons of scene gist recognition: Aerial and ground-based views. *Journal of Vision, 10,* 1238. doi:10.1167/10.7.1238

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimari, T. (2010). The role of higher order image statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics, 72,* 427–444. doi:10.3758/APP.72.2.427

Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision, 8,* 4, 1–9.

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made prior to basic-level distinctions in scene gist processing. *Visual Cognition, 18,* 513–536. doi:10.1080/13506280902937606

Loschky, L. C., Ringer, R., Ellis, K., & Hansen, B. C. (2013). *Comparing rapid scene categorization of aerial and terrestrial views: A new perspective on scene gist*. Manuscript in preparation.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1431–1450. doi:10.1037/0096-1523.33.6.1431

Macé, M. J. M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS One, 4,* e5927. doi:10.1371/journal.pone.0005927

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society B: Biological Sciences, 235,* 221–245. doi:10.1098/rspb.1988.0073

Morrone, M. C., & Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognition Letters, 6,* 303–313. doi:10.1016/0167–8655(87)90013–4

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117,* 382–405. doi:10.1037/a0018924

Nye, P. W. (1969). The monocular eye movements of the pigeon. *Vision Research, 9,* 133–144. doi:10.1016/0042-6989(69)90036-4

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175. doi:10.1023/A:1011139631724

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519–526. doi:10.3758/BF03197524

Pavel, M., Sperling, G., Riedl, T., & Vanderbeek, A. (1987). Limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America A: Optics, Image Science, and Vision,* 2355–2365. doi:10.1364/JOSAA.4.002355

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 587–595. doi:10.1037/0278-7393.15.4.587

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

Remy, M., & Watanabe, S. (1993). Two eyes and one world: Studies of interocular and intraocular transfer in birds. In H. P. Zeigler & H.-J. Bischof (Eds.), *Vision, brain, and behavior in birds* (pp. 333–350). Cambridge, MA: MIT Press.

Roberts, W. A., Phelps, M. T., Macuda, T., Brodbeck, D. R., & Russ, T. (1996). Intraocular transfer and simultaneous processing of stimuli presented in different visual fields of the pigeon. *Behavioral Neuroscience, 110,* 290–299. doi:10.1037/0735-7044.110.2.290

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630.

Schyns, P., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200. doi:10.1111/j.1467-9280.1994.tb00500.x

Soto, F., & Wasserman, E. (2012). Visual object categorization in birds and primates: Integrating behavioral, neurobiological, and computational evidence within a "general process" framework. *Cognitive, Affective, & Behavioral Neuroscience, 12,* 220–240. doi:10.3758/s13415-011-0070-x

Thomson, M. G. A. (1999). Visual coding and the phase structure of natural scenes. *Network, 10,* 123–132. doi:10.1088/0954-898X/10/2/302

Thomson, M. G. A. (2001). Sensory coding and the second spectra of natural signals. *Physical Review Letters, 86,* 2901–2904. doi:10.1103/PhysRevLett.86.2901

Thomson, M. G. A., & Foster, D. H. (1997). Role of second- and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 14,* 2081–2090. doi:10.1364/JOSAA.14.002081

Thorpe, S. J. (2002). Ultra-rapid scene categorization with a wave of spikes In H. H. Bulthoff (Ed.), *Biologically motivated computer vision* (pp. 1–15). Berlin, Germany: Springer. doi:10.1007/3-540-36181-2_1

Torralba, A., & Oliva, A. (1999). Semantic organization of scenes using discriminant structural templates. *Proceedings of International Conference on Computer Vision, 99,* 1253–1258.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems, 14,* 391–412. doi:10.1088/0954-898X/14/3/302

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786. doi:10.1037/0033-295X.113.4.766

Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology, 15,* 121–149. doi:10.1016/0010-0285(83)90006-3

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience, 15,* 209–217. doi:10.1162/089892903321208141

Vaughan, W. J., & Greene, S. L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 256–271. doi:10.1037/0097-7403.10.2.256

Von Fersen, L., & Lea, S. E. G. (1990). Category discrimination by pigeons using five polymorphous features. *Journal of the Experimental Analysis of Behavior, 54,* 69–84. doi:10.1901/jeab.1990.54-69

Received January 17, 2013
Revision received October 25, 2013
Accepted October 28, 2013 ■