Taking pigeons to heart: Birds proficiently diagnose human cardiac disease

Victor M. Navarro¹ · Edward A. Wasserman¹ · Piotr Slomka²

Published online: 21 January 2020
© The Psychonomic Society, Inc. 2020

Abstract

In two experiments, we trained pigeons (*Columba livia*) to sort visual images (obtained by clinical myocardial perfusion imaging techniques) depicting different degrees of human cardiac disfunction (myocardial hypoperfusion of the left ventricle) into normal and abnormal categories by providing food reward only after correct choice responses. Pigeons proved to be highly proficient at categorizing pseudo-colored images as well as highly sensitive to the degree of the perfusion deficit depicted in the abnormal images. In later testing, the pigeons completely transferred discriminative responding to novel stimuli, demonstrating that they had fully learned the normal and abnormal categories. Yet, these pigeons failed to transfer discriminative responding to grayscale images containing no color information. We therefore trained a second cohort of pigeons to categorize grayscale image sets from the outset. These birds required substantially more training to achieve similar levels of performance. Yet, they too completely transferred discriminative responding to novel stimuli by relying on both global and local disparities in brightness between the normal and abnormal images. These results confirm that pseudo-colorization can enhance pigeons' categorization of human cardiac images, a result also found with human observers. Overall, our findings further document the potential of the pigeon as a useful aide in studies of medical image perception.

Keywords Concept learning · Categorization · Attention · Pigeon · Generalization · Discrimination

Introduction

Whatever the pigeon instinct-mind contains, it is safe to say that intelligence is hardly more than a grain hidden in bushels of instinct, and one may search more than a day and not find it (Whitman, 1919, p. 158).

Pigeons suffer from a bum rap. As the above quote attests, even after spending most of his long and distinguished scientific career studying pigeons, American ethologist Charles Otis Whitman could not muster much enthusiasm for the cognitive prowess of the species. Decades later, even B. F. Skinner, who famously pioneered the use of pigeons in operant conditioning research, gave these birds little credit for their adaptive flexibility. Indeed, in explaining the choice of

pigeons for his famous guided missile project, Skinner said: “We have used pigeons, not because the pigeon is an intelligent bird, but because it is a practical one and can be made into a machine, from all practical points of view” (1944, quoted in Capshew, 1993, p. 851).

Notwithstanding these initial pessimistic appraisals of pigeon smarts, these much-maligned birds have subsequently exhibited a noteworthy capacity for what is currently called “visual cognition.” Considerable research has divulged that pigeons are prodigious discriminators of extremely complex visual stimuli. Among other kinds of visual stimuli, pigeons can reliably discriminate: misshapen pharmaceutical capsules (Verhave, 1966); basic object categories such as cats, flowers, cars, and chairs (Wasserman, 2016); the identities and emotional expressions of human faces (Soto & Wasserman, 2011); oil paintings by Monet and Picasso (Watanabe, Sakamoto, & Wakita, 1995); letters of the alphabet (Blough, 1985; Morgan, Fitch, Holman, & Lea, 1976); as well as typewritten words from nonwords (Scarf et al., 2016).

We remain a long way from knowing precisely how pigeons so successfully discriminate such a wide variety of complex visual stimuli as those noted above; but, later assessments of feature control (Lea & Ryan, 1983) have revealed

✉ Edward A. Wasserman
ed-wasserman@uiowa.edu

¹ Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242, USA

² Cedars-Sinai Medical Center, University of California, Los Angeles, CA, USA

that color, size, shape, and their combined configural cues all seem to participate (Wasserman & Biederman, 2012). However, one thing is certain: entirely without verbal instructions, elementary operant conditioning procedures alone have proven to be highly effective in teaching these birds intricate visual discriminations.

Of course, many of the above visual discriminations may seem to us a bit mundane; we can plausibly identify and verbalize the relevant attributes of the stimuli involved. But, what if the visual stimuli we ask pigeons to discriminate are actually extraordinarily difficult for us to discriminate and describe? Precisely these conditions present themselves to novice medical diagnosticians trying to learn how to distinguish visual images of normal and abnormal tissues.

Because of the extreme challenges posed by medical diagnosis, our laboratory's first forays explored pigeons' discrimination of benign and cancerous human breast images (Levenson, Krupinski, Navarro, & Wasserman, 2015; Navarro & Wasserman, 2016). We gave the birds a simple two-alternative forced-choice task in which food reward followed correct classification responses, but non-reward followed incorrect classification responses. Using this method, we sought answers to three questions. First, can pigeons be taught to discriminate benign from malignant *pathology* images? Second, can pigeons be taught to discriminate benign from malignant *radiology* images? And, third, once learned, can such pathology and radiology discriminations reliably *transfer* to novel benign and malignant images?

Our pigeons proved remarkably able to distinguish benign from malignant images. Within a mere 2 weeks of training, one squad of birds achieved accuracy levels averaging 85–90% correct on the tissue histology images. Similar performance was obtained from a second squad of pigeons on a task requiring them to determine whether or not a mammogram contained cancerous microcalcifications. However, when faced with a diagnostic task involving localized masses in mammograms, a third squad of pigeons had much more difficulty discriminating benign from malignant images – just as do radiologists and trainees.

Of course, our pigeons' high levels of accuracy on the more tractable pair of diagnostic tasks could simply have involved image memorization. However, when the birds were tested with novel visual images, they performed at very high levels of accuracy, indicating that they had effectively acquired the benign and malignant *concepts* or *categories*; the birds learned the necessary features to discriminate the lesions and thus were able to accurately diagnose brand-new cases. These results strongly suggest that pigeons might be well suited to helping us better understand medical image perception in general; they might also prove to be useful in performance assessment and in the development of medical imaging analysis tools.

The present project

In the present project, we again chose pigeons to serve as our non-human diagnosticians. However, their task this time was to categorize healthy and diseased heart muscle. Coronary artery disease represents a major public health problem, causing nearly one of every seven deaths in the USA, with a total mortality of 363,452 cases in 2016 (Benjamin et al., 2019). Although effective preventive pharmacological and invasive therapies exist, their appropriate and cost-effective deployment requires a reliable and objective method for detecting the disease. One such method is Myocardial Perfusion SPECT (Single Photon Emission Computed Tomography) or MPS, which provides key information concerning myocardial perfusion and ventricular function. It is estimated that 15–20 million MPS scans are performed annually worldwide (Einstein et al., 2015).

One difficulty with MPS is its reliance on humans' subjective visual interpretation of perfusion abnormalities, resulting in high intra- and inter-observer variability. The best achievable individual observer accuracy is about 86% and inter-observer agreement by Board-certified cardiology experts is about 87% (Slomka et al., 2017).

To improve on this apparent human performance ceiling, Slomka and his colleagues have developed a machine-learning approach that they hope will surpass the accuracy of visual reading by cardiologists (e.g., Arsanjani et al., 2013). In pursuit of this aim, these researchers have collected a database of several thousand well-characterized normal and abnormal MPS images, already scored by multiple expert readers and validated by invasive angiography or follow-up for cardiovascular events. The majority of computer-assisted diagnosis algorithms developed in radiology and pathology have to date relied on asking clinicians which features they themselves scrutinize in order to detect and discriminate lesions. However, this is an inherently flawed approach, because there are visual features that clinicians may either be missing or misinterpreting. Using a different model organism – the pigeon – may help us identify visual features that can help improve both human and computer performance.

What cardiologists must discriminate can be seen in Fig. 1A, which depicts two rows of two-dimensional, normal and abnormal polar maps of the left ventricular stress myocardial perfusion. The images in the top and bottom rows were derived from three-dimensional images of the left ventricle from five normal and five abnormal patients, respectively. Inspecting these images reveals that no two polar maps are identical, thus posing the daunting challenge of determining to which category – normal or abnormal – the clinician should assign any given image. The defects in the abnormal images can be striking or subtle (prompting false negatives); and, the normal variants are not entirely uniform, occasionally containing regions suspiciously similar to abnormal images (prompting false positives).

Fig. 1 Pigeon diagnosis of human cardiac disease. **(A)** The computer software developed to visualize and diagnose coronary artery disease is often trained on 40–50 confirmed, normal polar maps (top row). Based on these data, it establishes the statistical probability that a particular location in other polar maps is abnormal (bottom row). Polar maps are characterized by their total perfusion deficit (TPD; the number over or below each polar map). **(B)** Pigeons were trained to categorize images as “normal” or “abnormal” images by pecking choice buttons (black and white patterns) on a touchscreen and receiving food reward for correct responses. **(C)** The

different polar maps shown to pigeons. In Experiment 1, we trained pigeons to categorize pseudo-colored and, later, grayscale versions of the polar maps (first and second columns, respectively). In Experiment 2, we initially trained pigeons to categorize grayscale versions of the polar maps. The extent to which local or global brightness cues controlled their categorization behavior was assessed by presenting the pigeons with images adjusted to match the training set’s overall brightness value (third column) and with images showing flat grayscale values (fourth column), respectively

Given their prodigious visual abilities, we expected that pigeons would be able to accurately classify these nuclear cardiology images as either normal or abnormal. Finding that they successfully did so, we proceeded to ask three important follow-up questions: (1) How sensitively would pigeons’ diagnoses of abnormal images correspond to the actual *degree* of damage depicted in the stimuli? (2) How well would pigeons’ diagnoses *transfer* from familiar training stimuli to novel testing stimuli? (3) How strongly might *color* and *brightness* cues have contributed to pigeons’ categorization behavior?

Experiment 1

In Experiment 1, we trained pigeons to categorize colored images depicting cardiac damage (Fig. 1A) using a two-alternative, forced-choice task. The images differed in the amount of damage they depicted, as denoted by their total perfusion deficit (TPD) scores (see *Methods*); thus, they allowed us to estimate the degree to which categorical choices (“normal” or “abnormal”) were controlled by this attribute. After pigeons

learned to categorize the stimuli in an accurate and reliable manner, we determined whether their performance would transfer to a novel set of stimuli. Finally, we assessed the degree to which differences in overall brightness between categories controlled categorization – normal images were, overall, brighter than abnormal images – by testing for generalization to a stimulus set that included no color information.

Method

Subjects

Five pigeons (*Columba livia*) were studied. The birds were food-deprived to 85% of their free-feeding weight, while given free access to water and grit. Housing conditions and training procedures were approved by the Institutional Animal Care and Use Committee at the University of Iowa. All animals had previously participated in unrelated experiments involving simple visual discrimination and categorization (Castro & Wasserman, 2017; Couto, Navarro, Smith, & Wasserman, 2016; García-Gallardo, Navarro, & Wasserman,

2017; Levenson et al., 2015; Navarro, Jani, & Wasserman, 2019), thereby avoiding the need for additional training to interact with the programmed contingencies in the present experimental setting.

Apparatus

We used five $36 \times 36 \times 41$ cm conditioning chambers (see Fig. 1B; detailed in Gibson, Wasserman, Frei, & Miller, 2004), located in a dark room with continuous white noise. Each chamber was equipped with a 15-in. LCD monitor ($1,024 \times 768$ resolution) behind a resistive touchscreen. The visible portion of the screen was 28.5×17 cm. The screen had one 6×6 cm area for the start stimulus, a 9×9 cm area for the target stimulus, and two 9×2.8 cm areas for the choice buttons. The start and target stimulus were located 9 and 6 cm, respectively, above the wire mesh floor and were centered within the horizontal axis. Choice buttons were shown to the left and right of the target stimulus, with 1.5 cm of separation. A controller outside the chamber processed pecks to the touchscreen. A rotary dispenser delivered 45-mg food pellets through a vinyl tube into a plastic cup in the center of the rear wall opposite the touchscreen. Illumination during experimental sessions was provided by a houselight mounted on the upper rear wall of the chamber. Both the pellet dispenser and the houselight were controlled by a serial I/O interface. A separate iMac computer controlled each chamber, using programs developed in MATLAB® with Psychtoolbox-3 extensions (Brainard, 1997; Pelli, 1997; <http://psychtoolbox.org/>).

Stimuli

A white square with a cross in its center served as a start stimulus. Additionally, two choice buttons (black and white patterns created by dots or lines) served as choice buttons. The full set of cardiac polar maps comprised 96 images (48 normal and 48 abnormal), obtained from previous studies without any identifiable patient information. Images were initially obtained as intensity maps. They were then pseudo-colored in accord with cardiologists' preferences in making difficult diagnoses, using standard clinical software for quantitative perfusion SPECT QPS (Cedars Sinai Medical Center, Los Angeles, CA, USA). Another set of the same stimuli in grayscale was used to assess the role of color in pigeons' categorization behavior (Fig. 1C). However, note that the cardiac polar maps depict the same information (degree of perfusion), regardless of the color table used to convey this information (pseudo-coloring or grayscale).

A normative database of normal patients (40 low-likelihood males and 40 low-likelihood females; Slomka et al., 2005) was initially used as a reference point to calculate total perfusion deficit (TPD) scores for each image. Normal images were randomly sampled from this database in the 0–

3% range. In contrast, abnormal images were selected to cover a wide range of perfusion deficits (from 4% to 50%), and had previously been corroborated by invasive coronary angiography. Based on these scores, the images were further split into two sets (Set A and Set B) for training and testing purposes. Each set contained 24 normal and 24 abnormal images (for a total of 48 images per set), and the TPD scores from each category in each set approximated normal distributions. Finally, in order to more closely reflect the range of perfusion deficit within each set, TPD scores were recalculated using the 24 normal images from each set as normative data. As a result, the mean TPD scores for the normal and abnormal categories in Set A were 1.06% ($SD = 1.32$) and 23.52% ($SD = 11.06$), respectively; for Set B they were 0.99% ($SD = 1.03$) and 23.55% ($SD = 11.01$), respectively. Of further interest, the overall brightness of the normal and abnormal images differed in the two sets, with normal images tending to be brighter than abnormal images (because hypoperfusion was depicted as low intensity values in the current application). In the 0–255 range, the average brightness for the normal and abnormal categories in Set A were 164 and 141, respectively; for Set B they were 164 and 145, respectively.

Procedure

Training with pseudo-colored images Pigeons received daily training sessions until they attained a performance criterion (two consecutive sessions with a .75 proportion of correct choices in each category). In each training session, each stimulus from the training set (A or B, counterbalanced across birds) was randomly presented on four trials, for a total of 96 trials.

A trial started with the display of a start stimulus. Once pecked, the start stimulus disappeared and the stimulus to be categorized was displayed in the center of the screen. Pigeons then had to complete an observing response requirement by pecking the stimulus multiple times (starting at two in the first training session and increasing across sessions on the basis of each pigeon's performance; final values ranged from 26 to 48 pecks depending on the pigeon). Once the observing response requirement was satisfied, the two choice buttons appeared randomly to the left and right sides of the stimulus (in order to minimize position biases). The button assignment for each category was counterbalanced across birds (i.e., for three birds, the dots and lines patterns were associated with normal and abnormal images, respectively, but the reverse was true for the other two birds). A final peck to either of the buttons terminated the trial. If the chosen button corresponded with the category of the exemplar being displayed, then two to three food pellets were randomly delivered and a 6- to 10-s intertrial interval (ITI) ensued; however, if the chosen button did not correspond with the category of the exemplar being displayed, then no food was delivered and a correction trial

with the same stimulus and button positions was given after the ITI. Correction trials were given indefinitely until the correct choice button was chosen. Data from correction trials were excluded from all statistical analyses.

Transfer to novel, pseudo-colored images Each pigeon received two blocks of generalization testing. Each block contained eight training sessions and eight testing sessions (interleaved). Training sessions were identical to the training sessions that were given during the earlier phase; these sessions aimed to sustain criterion discrimination performance. Testing sessions contained all 96 trials of a training session plus 12 probe trials (six per category) with novel testing stimuli (i.e., stimuli from the stimulus set that had *not* been presented during training). Each of the 48 novel testing stimuli was presented twice per testing block – once in the first half and once in the second half – for a total of four presentations, and with the location of the two response buttons balanced across blocks. Finally, to prevent the pigeons from learning the true category of the novel testing stimuli, both nominally “correct” and “incorrect” choice responses produced food.

Transfer to grayscale training images Images were originally created as perfusion intensity maps; pseudo-colorizing was used when displaying these images to human observers to highlight critical diagnostic features. In order to determine whether the pigeons were relying on these color cues in their categorization behavior, we gave them an additional block of testing, including probe trials with grayscale versions of the same stimuli in the training set. The testing regime with these stimuli was identical to that used in the previous phase, except for the fact that each grayscale stimulus was only presented twice.

Training with grayscale images After finishing testing with the grayscale images, three of the five pigeons received training with the grayscale versions of the training stimuli for 50 sessions, in order to see whether differential reinforcement for correct and incorrect choices would lead to accurate categorization performance; the other two pigeons lagged too far behind to be tested given other laboratory contingencies.

Results

All analyses were performed using R version 3.3.2, with the lme4 (Bates, Mächler, Bolker, & Walker, 2015, p. 4), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2016), and car (Fox et al., 2016) packages. The random-effects structures of the mixed-effects models was selected in steps. Starting with a model that included no random effects, we compared models that had random-effects structures of increasing complexity,

using χ^2 tests. The data and scripts used to perform the following analyses are available on request.

Training with pseudo-colored images Individual pigeons differed in the number of sessions necessary to reach the training criterion, the fastest and slowest pigeons taking 28 and 73 sessions, respectively ($M = 47.0$, $SD = 18.32$). Because of this variability, we vincentized the proportion of correct choices over 10 *relative* blocks (Vincent, 1912). We then assessed categorization performance using a linear mixed-effect model on the empirical logit transformation of the proportion of correct choices. The final model included category (normal and abnormal, contrast coded as -0.5 and +0.5, respectively) and relative block (1–10, centered) as fixed effects, and a pigeon intercept and slope for category as random effects.

Figure 2A depicts the proportion of correct choices across relative blocks of training. The pigeons learned to categorize the cardiac stimuli, reliably increasing their proportion of correct choices as training ensued, $B = 0.11$, $SE = 0.01$, 95% CI [0.10, 0.12], $t(88) = 16.35$, $p < .001$. Although their performance in both categories was close to chance in relative block 1 (.51), by relative block 10, the pigeons had reached .77 ($SD = .02$) and .74 ($SD = .03$) proportion of correct choices for normal and abnormal categories, respectively. Furthermore, neither the relative speed of learning nor the overall proportion of correct choices differed between categories (both $ps > .05$).

We thus decided to focus on end-state categorization performance. Figure 2B shows the proportion of abnormal choices as a function of TPD, for each pigeon during the last 20 sessions of training. During this period, the overall proportions of correct responses for normal and abnormal images were .70 ($SD = .05$) and .71 ($SD = .02$), respectively. We assessed these data using a logistic mixed-effects model, including normalized TPD (0 to 1) as a fixed effect, and a pigeon intercept and slope for normalized TPD as random effects. As illustrated by the model fit in Fig. 2B, our pigeons' choices were reliably controlled by the degree of perfusion deficit depicted in the images, $B = 3.75$, $SE = 0.36$, 95% CI [3.05, 4.45], $Z = 10.52$, $p < .001$; the likelihood with which the pigeons made an “abnormal” choice progressively increased with the TPD of the training images.

Transfer to novel, pseudo-colored images Our pigeons transferred their performance to novel images remarkably well. Their overall proportions of correct responses with familiar images were .81 ($SD = .05$) and .79 ($SD = .04$) for normal and abnormal categories, respectively. Similarly, their overall proportions of correct responses with novel images were .76 ($SD = .07$) and .76 ($SD = .05$) for normal and abnormal categories, respectively.

We assessed the choice data from this transfer phase using a logistic mixed-effects model, including normalized TPD and stimulus type (familiar or novel, coded as -0.5 and +0.5,

Fig. 2 Performance of pigeons across the different stages of Experiment 1. (A) Proportion of correct responses across relative blocks of training with pseudo-colored images, for all five pigeons. Error bars represent the standard error of the mean. (B) Proportion of “abnormal” choices per stimulus as a function of its TPD score during the last 20 sessions of training. The solid line represents a logistic fit of the pigeons’ responses. (C) Proportion of “abnormal” choices per stimulus as a function of its TPD score during transfer to pseudo-colored, novel stimuli. (D)

Proportion of “abnormal” choices per stimulus as a function of its TPD score during testing with grayscale versions of the training stimuli. (E) Proportion of correct responses across five-session blocks of training with grayscale versions of the training stimuli for the three remaining pigeons. (F) Proportion of “abnormal” choices per stimulus as a function of its TPD score, during the last 20 sessions of training with grayscale versions of the training stimuli

respectively) as fixed effects, and a pigeon intercept and slope for normalized TPD as random effects. Figure 2C portrays the proportion of abnormal choices for familiar (black) and novel (red) stimuli as a function of the TPD scores, averaged across the testing sessions in this phase. As illustrated by the striking overlap between the fitted curves, the pigeons evidenced remarkable success in transferring their categorization performance to novel stimuli. Importantly, the pigeons remained highly sensitive to the TPD scores of the images, $B = 6.30$, $SE = 0.83$, 95% CI [4.67, 7.93], $Z = 7.56$, $p < .001$; neither this sensitivity nor the overall probability of reporting an image as abnormal differed between categories (both $ps > .05$). These results show that our pigeons’ categorization performance did not rely on rote memorization of the individual training images.

Transfer to grayscale training images Our pigeons failed to transfer their categorization performance to grayscale images. Whereas their proportions of correct responses with familiar colorized images were .83 ($SD = .38$) and .80 ($SD = .40$) for both normal and abnormal categories, respectively, their proportions of correct responses with grayscale images were only

.37 ($SD = .63$) and .71 ($SD = .45$) for normal and abnormal categories, respectively.

We analyzed the choice data from this transfer phase using a logistic mixed-effects model, including normalized TPD and stimulus type (pseudo-colored or gray, coded as -0.5 and +0.5, respectively) as fixed effects, and a pigeon intercept and slopes for normalized TPD and stimulus type as random effects. Figure 2D shows the proportions of abnormal choices for pseudo-colored (black) and grayscale (blue) stimuli as a function of TPD score, averaged across the testing sessions of this transfer phase. The model revealed a significant Stimulus Type $\times$ TPD interaction, $B = -6.14$, $SE = 0.52$, 95% CI [-7.15, -5.13], $Z = -11.90$, $p < .001$. A follow-up analysis of this interaction revealed that the pigeons’ proportions of “abnormal” choices for normal images was below chance for pseudo-colored images, but not for grayscale images ($Intercept_{pseudo-colored} = -1.57$, $SE = 0.11$, 95% CI [-1.79, -1.35], $Z = -13.98$, $p < .001$, and $Intercept_{Grayscale} = 0.70$, $SE = 0.63$, 95% CI [-0.54, 1.94], $Z = 1.11$, $p > .10$). Furthermore, although the pigeons were sensitive to the TPD scores of both types of stimuli

($B = 7.45$, $SE = 0.87$, 95% CI [5.74, 9.15], $Z = 8.55$, $p < .001$ and $B = 1.12$, $SE = 0.52$, 95% CI [0.09, 2.15], $Z = 2.13$, $p < .05$ for pseudo-colored and grayscale stimuli, respectively), they were decidedly less sensitive to the TPD scores of the grayscale images than the pseudo-colored images.

Training with grayscale images Figure 2E and F depict the proportion of correct choices as a function of training blocks and the proportion of “abnormal” choices as a function of TPD, during the last 20 sessions of grayscale training, respectively. The models used to assess these data were identical to those used to assess the training data with pseudo-colored images.

As illustrated in Fig. 2E, the three remaining pigeons learned to correctly categorize the grayscale images when given differential reward for correct and incorrect choices, $B = 0.04$, $SE = 0.01$, 95% CI [0.02, 0.06], $t(52) = 4.16$, $p < .001$. Furthermore, neither the speed of learning nor the overall proportion of correct choices reliably differed between categories (both $ps > .05$), although their categorization of normal images tended to be more accurate than their categorization of abnormal images.

The analysis of abnormal choices during the last 20 sessions of training revealed that the pigeons had become increasingly sensitive to the TPD scores of the grayscale stimuli, $B = 4.65$, $SE = 1.39$, 95% CI [1.93, 7.37], $Z = 3.35$, $p < .01$. Importantly, the pigeons’ sensitivity to the TPD of the grayscale images was now similar to their sensitivity to pseudo-colored images (cf. Fig. 2B and F).

Discussion

Experiment 1 documented that pigeons learned to accurately categorize pseudo-colored polar plots of human heart muscle perfusion abnormalities. In doing so, the pigeons’ responding was tightly controlled by the degree of abnormality depicted in each image (TPD score; Fig. 2B). Furthermore, pigeons transferred their accuracy and sensitivity to images that they had never seen before (Fig. 2C). Nevertheless, the pigeons failed to spontaneously transfer this categorization performance to familiar images that had their color information removed (grayscale images; Fig. 2D). Although their categorization of grayscale images improved after we gave the pigeons explicit training (i.e., differential reward for correct and incorrect responses; Fig. 2E and F), we decided to study a different cohort of pigeons given grayscale images from the outset of training in order to learn more about the importance of color to pigeons’ categorization behavior.

Experiment 2

In Experiment 2, we further examined the role of color and brightness cues in pigeons’ categorization of cardiac images. As in Experiment 1, we began by training different pigeons to categorize the, now grayscale, stimuli accurately and reliably. After they reached criterion, we tested whether they would transfer their categorization behavior to a novel set of stimuli. In a final test, we assessed the control that was exerted by local and global brightness cues, using brightness-equalized images and flat images depicting different grayscale values, respectively.

Subjects

Four different pigeons (*Columba livia*) were studied. The animals were housed in identical conditions and had past experiences similar to the animals in Experiment 1.

Apparatus

The same operant chambers in Experiment 1 were used in this investigation.

Stimuli

The grayscale images presented to the pigeons during the “transfer to grayscales” phase of Experiment 1 were used in this experiment (96 images). Again, two sets were created for training and testing purposes, with each set containing 24 normal and 24 abnormal images. A subset of 48 of these images (24 normal and 24 abnormal images, see *Procedure*) was used to create a set of stimuli that had the same overall brightness, by adding the difference between the average pixel value across all images, regardless of their category and the average pixel value of the image to each of the pixels in the image. The average brightness value across all images was 152 (within the 0–255 range). For example, if an abnormal image had an average pixel value of 140, then a new, brightness-equalized image was created by adding 12 to each of its pixels. Values exceeding the range after transformation were truncated. Effectively, all of the images in this set had the same overall brightness (152), but they retained the contrasts between local pixels (Fig. 1C, third column), thereby allowing us to assess the control exerted by local brightness cues. Finally, in order to assess control by global brightness cues, we created five images depicting various flat grayscale values (Fig. 1C, fourth column): 133 (a value even lower than the average of the abnormal images), 143 (the average value of the abnormal images), 153 (the average of both the abnormal and normal images), 163 (the average value of the normal images), and 173 (a value even higher than the average of the normal images).

Procedure

Training with grayscale images Training with grayscale images was carried out as in the “*Training with pseudo-colored images*” phase of Experiment 1. Pigeons received daily training sessions until they performed two consecutive sessions with a .70 proportion of correct choices in each category (this slightly lower criterion was established given the asymptotic performance of the pigeons in Experiment 1; Fig. 2E).

Transfer to novel, grayscale images Only two of the initial four pigeons participated in this and subsequent phases; the other two pigeons lagged too far behind to be tested given other laboratory contingencies. Testing with novel, grayscale images was carried out here as in the “testing with novel, pseudo-colored images” phase of Experiment 1.

Training with an easier set of grayscale images After extensive training, the pigeons’ overall proportion of correct choices was far from perfect – pigeons reached an asymptote of just .70 proportion of correct choices on the training images (see Fig. 3A). Hence, in preparation for future testing, both pigeons received training with an “easy” subset of the images that they had seen earlier. This image set contained no images in the 2–11 TPD range (where most errors were located). The average TPD values for the normal and abnormal categories were thus 0.13 ($SD = 0.33$) and 20.67 ($SD = 4.36$), respectively. Given the limited number of familiar images in the 12–50 range that each pigeon had seen, we decided to create a common image set for all of the pigeons. In order to do so, the images in this set were selected so that nearly half of them had been part of the transfer set given to each pigeon in the previous transfer phase. Pigeons received daily sessions with this set until they reached a proportion of .85 correct choices in both categories for a single session (12 and 23 sessions, for 36R and 13Y, respectively; see Fig. 3D). No statistical analyses were performed on these data.

Testing with brightness-equalized images and grayscale values After finishing training with the easier sets of grayscale images, pigeons received 35 testing sessions. Each testing session contained 76 training trials (38 per category) in which the image to be categorized was sampled at random, with replacement, from the set of grayscale images contained in the easier set (see above). Additionally, each testing session contained 12 probe trials (six per category) in which the image to be categorized was sampled at random, without replacement, from the set of images with equalized brightness (Fig. 1C, third column). Finally, each testing session also contained ten probe trials depicting five flat grayscale values (two trials per value). Pigeons received differential reward on training trials but non-differential reward on trials depicting images with equalized-brightness images or flat grayscale values.

Results

Training with grayscale images Compared to the pigeons in Experiment 1, the pigeons in this experiment took more training sessions to reach the training criterion ($M = 96.75$, $SD = 66.08$), indicating that the grayscale images were harder to categorize. The fastest of the four pigeons took 22 sessions to reach criterion, whereas the slowest pigeon took 180 sessions to do so. We assessed categorization performance using a linear mixed-effects model identical to the one used to analyze Experiment 1.

Figure 3A depicts the proportion of correct choices across relative blocks of training. Again, the pigeons successfully learned to categorize the cardiac stimuli, reliably increasing their proportion of correct button choices as training ensued, $B = 0.10$, $SE = 0.01$, 95% CI [0.08, 0.12], $t(73) = 10.55$, $p < .001$. However, in contrast to categorization performance with pseudo-colored images, the overall proportion of correct responses did differ significantly between categories; pigeons were more accurate at categorizing the normal images, $B = -0.15$, $SE = 0.05$, 95% CI [-0.25, -0.04], $t(73) = -2.73$, $p < .01$ (cf. Figs. 2A and 3A). Note, however, that the pigeons’ learning rates did not significantly differ between categories, $B = -0.01$, $SE = 0.02$, 95% CI [-0.04, 0.03], $t(73) = -0.50$, $p > .10$. This difference in overall accuracy disappeared as the pigeons approached the learning criterion; the proportions of correct responses in the last training session were .74 ($SD = 0.02$) and .79 ($SD = .07$) for normal and abnormal categories, respectively.

Figure 3B shows the proportion of abnormal choices as a function of TPD, across all four pigeons during the last 20 sessions of training with grayscale stimuli. Here, the fixed- and random-effects structures included only TPD and pigeon intercepts, respectively. Despite an overall difference in performance between categories, the pigeons were still highly sensitive to the TPD of the images, $B = 3.32$, $SE = 0.11$, 95% CI [3.11, 3.52], $Z = 31.47$, $p < .001$. Indeed, neither the overall probability of reporting an image as abnormal nor the TPD sensitivity of these pigeons differed significantly from those of the pigeons in Experiment 1 (both $ps > .10$, as assessed by a logistic model including experiment as a factor).

Transfer to novel, grayscale images The two pigeons that underwent testing with novel grayscale images transferred their categorization performance remarkably well (Fig. 3C). Their overall proportions of correct responses with familiar images were .81 ($SD = .02$) and .75 ($SD = .02$) for normal and abnormal categories, respectively. Similarly, their overall proportions of correct choices with novel images were .76 ($SD = .02$) and .79 ($SD = .04$) for normal and abnormal categories, respectively.

The individual differences between the pigeons did not justify any kind of random-effects structure. Thus, a logistic model disclosed that the pigeons remained highly sensitive to

Fig. 3 Performance of pigeons across the different stages of Experiment 2. (A) Proportion of correct responses across relative blocks of training with grayscale images for all four pigeons. Error bars represent the standard error of the mean. (B) Proportion of “abnormal” choices per stimulus as a function of its TPD score during the last 20 sessions of training. The solid line represents a logistic fit of the pigeons’ responses. (C) Proportion of “abnormal” choices per stimulus as a function of its TPD score during transfer to novel grayscale stimuli for

the two remaining pigeons. (D) Proportion of correct responses across relative blocks of training with an easier set of grayscale images. (E) Proportion of “abnormal” choices per stimulus as a function of its TPD during transfer to brightness-equalized images. (F) Proportion of “abnormal” choices as a function of brightness (0–255) for the two remaining pigeons. The average brightness values for abnormal and normal images are annotated

the TPD of the images, $B = 5.83$, $SE = 0.35$, 95% CI [5.15, 6.50], $Z = 16.89$, $p < .001$, and neither this sensitivity nor the overall probability of reporting an image as abnormal differed between categories (both $ps > .05$). These results show that the categorization performance of the pigeons in Experiment 2 did not rely on memorization of individual training images: the pigeons had indeed learned the normal/abnormal categories or concepts, even in the absence of color cues.

Testing with brightness-equalized images and grayscale values

Pigeons’ accuracy was rather well preserved even in the absence of global differences in brightness between the image categories. Indeed, the proportions of correct choices for familiar and brightness-equalized images were .81 ($SD = .01$) and .77 ($SD = .03$), respectively. Figure 3E depicts the proportion of “abnormal” choices as a function of TPD for familiar and equalized-brightness images. A logistic model including normalized TPD and stimulus type (familiar or equalized-brightness, coded as -0.5 and 0.5, respectively) revealed that the pigeons’ overall choices remained highly sensitive to the TPD of the images, $B = 3.65$, $SE = 0.12$, 95% CI [3.41, 3.90], $Z = 29.41$, $p < .001$, but that they were slightly

more likely to make abnormal choices when presented with brightness-equalized images, $B = 0.28$, $SE = 0.13$, 95% CI [0.02, 0.54], $Z = 2.11$, $p < .05$. More importantly, as the crossing between the functions depicted in Fig. 3E suggests, the pigeons were less sensitive to the TPD of the brightness-equalized images than to that of the familiar images, $B = -0.69$, $SE = 0.25$, 95% CI [-1.18, -0.20], $Z = -2.77$, $p < .01$. These results suggest that the difference in overall brightness between the categories partly controlled our pigeons’ categorization performance.

This conclusion was further supported by our pigeons’ choices on trials with grayscale values. Figure 3F depicts the proportion of “abnormal” choices as a function of the brightness value of the patch for each pigeon. As annotated in the figure, the average brightness value for abnormal images was darker than the average brightness value of normal images (145 vs. 163, respectively, calculated from the set of familiar images). And, although our pigeons showed a bias towards making more “normal” choices when presented with these stimuli, their proportion of “abnormal” choices decreased reliably and monotonically as the patches were made brighter, $B = -0.04$, $SE = 0.01$, 95% CI [-0.05, -0.03], $Z = -7.04$, $p < .001$.

Discussion

Experiment 2 showed that inexperienced pigeons too can learn to categorize grayscale images of human cardiac damage and spontaneously transfer their b to novel images. Again, the pigeons' categorization was tightly controlled by the degree of abnormality depicted in each image (Fig. 3B and C). Subsequent tests with images equalized in overall brightness and with grayscale patches of varying brightness further revealed that the categorization of grayscale stimuli was primarily controlled by local differences in brightness (Fig. 3E), but was also controlled by the overall brightness of the images (Fig. 3F). Notably, the pigeons in Experiment 2 were slower than the pigeons in Experiment 1 to reach a lower learning criterion. Additionally, the pigeons in Experiment 2 were reliably more accurate in categorizing normal over abnormal images, a difference that did not reach significance in Experiment 1. Consider that normal images are more similar to each other, in terms of their TPD scores; so, the normal category is more homogeneous than the abnormal category. This fact alone might lead to an advantage in categorization performance (Nosofsky, 1988). So, it is surprising that this difference only appeared when we removed the color cues (cf. Figs. 2A and 3A). Hence, it is possible that color cues increased the homogeneity of abnormal images, making the categorization task easier. These observations, coupled with the fact that pigeons in Experiment 1 failed to spontaneously transfer their categorization behavior to grayscale versions of familiar colorized stimuli, strongly suggests that color plays an important role in pigeons' categorization of these images.

General discussion

In Experiment 1, pigeons successfully learned to categorize polar maps representing cardiac perfusion deficits into normal and abnormal categories (Fig. 2A); in doing so, they also became highly sensitive to the degree of perfusion deficit present in each polar plot (Fig. 2B). Most notably, the pigeons successfully transferred this damage-based categorization behavior to novel images with virtually no decrement (Fig. 2C), thereby demonstrating that they had learned true normal/abnormal categories. However, the birds failed to categorize the same images when all of the color cues were removed (Fig. 2D), only learning to do so when given differential reward for correct and incorrect responses (Fig. 2E). After this differential reward training, the pigeons were once again as sensitive to the degree of perfusion abnormality as they were with the pseudo-colored images (cf. Fig. 2B and F).

Experiment 2 further explored the categorization of grayscale images using a different cohort of birds. Pigeons in this experiment needed more sessions to reach the training criterion; but, once they did so, they transferred their categorization

performance to novel images with little decrement in performance (Fig. 2C). Finally, subsequent tests using images with equalized brightness and patches of grayscale values (Fig. 1C) revealed that these pigeons' categorization performance was controlled by both local and global brightness cues (Fig. 3E and F).

Both of our experiments suggested that our pigeons' choices were better described with a logistic function than with a step function. In other words, the odds of an "abnormal" choice were tied to the perfusion deficit being depicted in the images, not to their category labels (normal/abnormal). This might be considered a flaw in our pigeons' categorization performance, because a binary, unequivocal diagnosis is the clinical ideal. Without a doubt, our pigeons could do better if we trained them with stimuli closer to the normal/abnormal boundary, thereby forcing their responses to be more deterministic (Alfonso-Reese, Ashby, & Brainard, 2002; Kalish & Kruschke, 1997). However, the subtle transition from "normal" to "abnormal" choices reflects exactly the relation between normal hearts and hearts with barely abnormal degrees of hypoperfusion. In our view, our pigeons' choices, and the uncertainty they carry when the stimuli are close to the boundary, faithfully capture the inherent fuzziness of natural variation.

Together, these experiments also indicate that color can play a commanding role in categorizing images of human heart muscle. Pigeons ignored brightness cues if color cues were available (Experiment 1); indeed, even though they were able to categorize cardiac images on the basis of brightness cues alone, they required substantially more training to achieve similar degrees of accuracy and sensitivity to detect abnormalities (Experiment 2). It is not uncommon to find mixed control by all or several stimulus dimensions, especially when using natural, uncontrolled stimuli as in the present experiment (Lea & Ryan, 1983; Lea et al., 2018; Lea, Lohmann, & Ryan, 1993).

Note, however, that both color and grayscale images both afford the same diagnostic information. The color of the polar plots depicted in Fig. 1C was not inherent to the images, but was instead added using a color map that accorded with the degree of perfusion in the myocardium (Slomka et al., 2005). Given this added colorization, the categorization space for the color polar plots is more complex than that for the grayscale polar plots. This key difference might explain why the pigeons in Experiment 1 were faster to learn than those in Experiment 2. If both color and brightness information are valid sources of information about category membership, then the pigeons in Experiment 1 could have relied on either color or brightness, or both; on the other hand, the pigeons in Experiment 2 were forced to rely on brightness alone.

Multiple studies have documented that the learning rate in a categorization problem is directly related to the number of stimuli (or stimulus dimensions) that convey information

concerning category membership (Bourne & Restle, 1959; Bourne & Haygood, 1959; Restle, 1959; Trabasso, 1960), possibly because they present subjects with a variety of equally valid sources of information that might satisfy their perceptual aptitudes (Sutherland & Mackintosh, 1964). In the case of the pigeon, control by color information is often stronger than is control by other stimulus dimensions (Farthing & Hearst, 1970; Lea & Wills, 2008; Wasserman, Bhatt, Chatlosh, & Kiedinger, 1987); so, the availability of color information may have given the pigeons in Experiment 1 an edge over the pigeons in Experiment 2. Note again, however, that the pigeons in both experiments were fundamentally controlled by the same source of information: namely, local differences in perfusion. These differences were represented by differences in brightness in the grayscale stimuli, but were additionally represented by abrupt transitions between hues in the pseudo-colored stimuli (so-called pseudocontouring; e.g. Hansen, 2006).

Whether the pseudo-colorization of grayscale images helps human observers assess medical images seems to be domain specific. For example, the use of pseudo-colorization increases accuracy, diagnostic confidence, and inter-observer agreement of computed tomography (CT) scans depicting carotid artery dissection (Saba et al., 2014). However, pseudo-colorization does not improve the detection of interproximal caries (Booshehry, Davari, Ardakani, & Nejad, 2010; Takeshita et al., 2013). To complicate things further, the specific color gamut used for pseudo-colorization is also an important factor. For example, Zabala-Travers, Choi, Cheng, and Badano (2015, but see Li & Burgess, 1997) found that participants were more accurate when presented images pseudo-colored with a rainbow colormap (i.e., mapping intensity values to the full spectra of light perceived by humans) than when presented grayscale images, or images pseudo-colored with a hot colormap. Although some experts advocate against the use of rainbow coloring without considering the specific needs of the data being portrayed (Rogowitz & Treinish, 1998), rainbow coloring remains ubiquitous in the depiction of medical images and can sometimes be preferred by expert observers over other color palettes (Borkin et al., 2011; Borland & Li, 2007).

In all, the full effects of pseudo-colorization on humans' categorization of medical images remain an unknown yet exciting realm of future research (Krupinski, 2010; Wolfe, 2016). Thus, as the field of medical imaging moves to identify the practices that most facilitate the diagnoses made by medical professionals (Badano et al., 2015), we believe that the pigeon may be a promising surrogate for medical image perception studies. This bird may have no particular knack for medical diagnosis, yet its eye and brain endow it with sufficient perceptual and cognitive equipment to provide researchers with practical methods for assessing human and machine performance.

Author notes This research was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the NIH under Award Number P01HD080679.

References

- Alfonso-Reese, L. A., Ashby, F. G., & Brainard, D. H. (2002). What makes a categorization task difficult? *Perception & Psychophysics*, 64, 570–583. <https://doi.org/10.3758/BF03194727>
- Arsanjani, R., Xu, Y., Dey, D., Fish, M., Dorbala, S., Hayes, S., ... Slomka, P. (2013). Improved accuracy of myocardial perfusion SPECT for the detection of coronary artery Disease using a support vector machine algorithm. *Journal of Nuclear Medicine*, 54, 549–555. <https://doi.org/10.2967/jnumed.112.111542>
- Badano, A., Revie, C., Casertano, A., Cheng, W.-C., Green, P., Kimpe, T., ... Yokoi, H. (2015). Consistency and standardization of color in medical imaging: A consensus report. *Journal of Digital Imaging*, 28, 41–52. <https://doi.org/10.1007/s10278-014-9721-0>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67. <https://doi.org/10.18637/jss.v067.i01>
- Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., ... American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2019). Heart disease and stroke statistics-2019 update: A report from the american heart association. *Circulation*, 139, e56–e66. <https://doi.org/10.1161/CIR.0000000000000659>
- Blough, D. S. (1985). Discrimination of letters and random dot patterns by pigeons and humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 261–280. <https://doi.org/10.1037/0097-7403.11.2.261>
- Booshehry, M. Z., Davari, A., Ardakani, F. E., & Nejad, M. R. R. (2010). Efficacy of application of pseudocolor filters in the detection of interproximal caries, 4, 4.
- Borkin, M., Gajos, K., Peters, A., Mitsouras, D., Melchionna, S., Rybicki, F., ... Pfister, H. (2011). Evaluation of artery visualizations for heart disease diagnosis. *IEEE Transactions on Visualization and Computer Graphics*, 17, 2479–2488. <https://doi.org/10.1109/TVCG.2011.192>
- Borland, D., & Li, R. M. T. (2007). Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27, 14–17. <https://doi.org/10.1109/MCG.2007.323435>
- Bourne, L. E., & Haygood, R. C. (1959). The role of stimulus redundancy in concept identification. *Journal of Experimental Psychology*, 58, 232–238. <https://doi.org/10.1037/h0040047>
- Bourne, L. E., & Restle, F. (1959). Mathematical theory of concept identification. *Psychological Review*, 66, 278–296. <https://doi.org/10.1037/h0041365>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. <https://doi.org/10.1163/156856897X00357>
- Capshaw, J. H. (1993). Engineering behavior: Project Pigeon, World War II, and the conditioning of B. F. Skinner. *Technology and Culture*, 34, 835. <https://doi.org/10.2307/3106417>
- Castro, L., & Wasserman, E. A. (2017). Feature predictiveness and selective attention in pigeons' categorization learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, 43, 231–242. <https://doi.org/10.1037/xan0000146>
- Couto, K. C., Navarro, V. M., Smith, T. R., & Wasserman, E. A. (2016). Concept learning without differential reinforcement in pigeons by

- means of contextual cueing. *Journal of Experimental Psychology: Animal Learning and Cognition*, 42, 221–227. <https://doi.org/10.1037/xan0000099>
- Einstein, A. J., Pascual, T. N. B., Mercuri, M., Karthikeyan, G., Vitola, J. V., Mahmarian, J. J., ... INCAPS Investigators Group. (2015). Current worldwide nuclear cardiology practices and radiation exposure: results from the 65 country IAEA Nuclear Cardiology Protocols Cross-Sectional Study (INCAPS). *European Heart Journal*, 36, 1689–1696. <https://doi.org/10.1093/eurheartj/ehv117>
- Farthing, G. W., & Hearst, E. (1970). Attention in the pigeon: Testing with compounds or elements. *Learning and Motivation*, 1, 65–78. [https://doi.org/10.1016/0023-9690\(70\)90129-3](https://doi.org/10.1016/0023-9690(70)90129-3)
- Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., ... R-Core. (2016). car: Companion to applied regression (Version 2.1-4). Retrieved from <https://cran.r-project.org/web/packages/car/index.html>
- García-Gallardo, D., Navarro, V. M., & Wasserman, E. A. (2017). Assessing the acquisition of anticipatory responding in the pigeon using reaction time. *Journal of Experimental Psychology: Animal Learning and Cognition*, 43, 197–203. <https://doi.org/10.1037/xan0000136>
- Gibson, B. M., Wasserman, E. A., Frei, L., & Miller, K. (2004). Recent advances in operant conditioning technology: A versatile and affordable computerized touchscreen system. *Behavior Research Methods, Instruments, & Computers*, 36, 355–362. <https://doi.org/10.3758/BF03195582>
- Hansen, C. L. (2006). The role of the translation table in cardiac image display. *Journal of Nuclear Cardiology*, 13, 571–575.
- Kalish, M. L., & Kruschke, J. K. (1997). Decision boundaries in one-dimensional categorization. *Journal of Experimental Psychology: Learning Memory and Cognition*, 23, 1362–1377. <https://doi.org/10.1037/0278-7393.23.6.1362>
- Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception & Psychophysics*, 72. <https://doi.org/10.3758/APP.72.5.1205>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models (Version 2.0-33). Retrieved from <https://cran.r-project.org/web/packages/lmerTest/index.html>
- Lea, S. E. G., Lohmann, A., & Ryan, C. M. E. (1993). Discrimination of five-dimensional stimuli by pigeons: Limitations of feature analysis. *The Quarterly Journal of Experimental Psychology Section B*, 46, 19–42. <https://doi.org/10.1080/14640749308401093>
- Lea, S. E. G., Pothos, E. M., Wills, A. J., Leaver, L. A., Ryan, C. M. E., & Meier, C. (2018). Multiple feature use in pigeons' category discrimination: The influence of stimulus set structure and the salience of stimulus differences. *Journal of Experimental Psychology: Animal Learning and Cognition*, 44, 114–127. <https://doi.org/10.1037/xan0000169>
- Lea, S. E. G., & Ryan, C. M. E. (1983). Feature analysis of pigeons' acquisition of concept discrimination. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior* (Vol. 4, pp. 239–253). Cambridge, MA: Ballinger.
- Lea, S. E. G., & Wills, A. J. (2008). Use of multiple dimensions in learned discriminations. *Comparative Cognition & Behavior Reviews*, 3. <https://doi.org/10.3819/ccbr.2008.30007>
- Levenson, R. M., Krupinski, E. A., Navarro, V. M., & Wasserman, E. A. (2015). Pigeons (Columba livia) as trainable observers of pathology and radiology breast cancer images. *PLOS ONE*, 10, e0141357. <https://doi.org/10.1371/journal.pone.0141357>
- Li, H., & Burgess, A. E. (1997). Evaluation of signal detection performance with pseudocolor display and lumpy backgrounds. In *Proc.SPIE* (Vol. 3036). Retrieved from <https://doi.org/10.1117/12.271286>
- Morgan, M. J., Fitch, M. D., Holman, J. G., & Lea, S. E. G. (1976). Pigeons learn the concept of an 'A.' *Perception*, 5, 57–66. <https://doi.org/10.1068/p050057>
- Navarro, V. M., Jani, R., & Wasserman, E. A. (2019). Pigeon category learning: Revisiting the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: Animal Learning and Cognition*, 45, 174–184. <https://doi.org/10.1037/xan0000198>
- Navarro, V. M., & Wasserman, E. A. (2016). Stepwise conceptualization in pigeons. *Journal of Experimental Psychology: Animal Learning and Cognition*, 42, 44–50. <https://doi.org/10.1037/xan0000083>
- Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 54–65. <https://doi.org/10.1037/0278-7393.14.1.54>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. <https://doi.org/10.1163/156856897X00366>
- Restle, F. (1959). Additivity of cues and transfer in discrimination of consonant clusters. *Journal of Experimental Psychology*, 57, 9–14. <https://doi.org/10.1037/h0041241>
- Rogowitz, B. E., & Treinish, L. A. (1998). Data visualization: the end of the rainbow. *IEEE Spectrum*, 35, 52–59. <https://doi.org/10.1109/6.736450>
- Saba, L., Argiolas, G. M., Raz, E., Sannia, S., Suri, J. S., Siotto, P., ... Wintermark, M. (2014). Carotid artery dissection on non-contrast CT: Does color improve the diagnostic confidence? *European Journal of Radiology*, 83, 2288–2293. <https://doi.org/10.1016/j.ejrad.2014.09.001>
- Scarf, D., Boy, K., Reinert, A. U., Devine, J., Güntürkün, O., & Colombo, M. (2016). Orthographic processing in pigeons (Columba livia). *Proceedings of the National Academy of Sciences*, 113, 11272–11276. <https://doi.org/10.1073/pnas.1607870113>
- Slomka, P. J., Dey, D., Sitek, A., Motwani, M., Berman, D. S., & Germano, G. (2017). Cardiac imaging: working towards fully-automated machine analysis & interpretation. *Expert Review of Medical Devices*, 14, 197–212. <https://doi.org/10.1080/17434440.2017.1300057>
- Slomka, P. J., Nishina, H., Berman, D. S., Akincioglu, C., Abidov, A., Friedman, J. D., ... Germano, G. (2005). Automated quantification of myocardial perfusion SPECT using simplified normal limits. *Journal of Nuclear Cardiology*, 12, 66–77. <https://doi.org/10.1016/j.nuclcard.2004.10.006>
- Soto, F. A., & Wasserman, E. A. (2011). Asymmetrical interactions in the perception of face identity and emotional expression are not unique to the primate visual system. *Journal of Vision*, 11, 1–18. <https://doi.org/10.1167/11.3.24>
- Sutherland, N. S., & Mackintosh, J. (1964). Discrimination learning: Non-additivity of cues. *Nature*, 201, 528. <https://doi.org/10.1038/201528b0>
- Takeshita, W., Iwaki, L. V., Da Silva, M., Filho, L., Queiroz, A. F., & Geron, L. G. (2013). Comparison of the diagnostic accuracy of direct digital radiography system, filtered images, and subtraction radiography. *Contemporary Clinical Dentistry*, 4, 338. <https://doi.org/10.4103/0976-237X.118391>
- Trabasso, T. R. (1960). Additivity of cues in discrimination learning of letter patterns. *Journal of Experimental Psychology*, 60, 83–88. <https://doi.org/10.1037/h0042922>
- Verhave, T. (1966). The pigeon as a quality-control inspector. *American Psychologist*, 21, 109–115. <https://doi.org/10.1037/h0023273>
- Vincent, S. B. (1912). The function of the vibrissae in the behavior of the white rat. *Animal Behavior Monographs*, 1, 5, 84–84.
- Wasserman, E. A. (2016). Conceptualization in pigeons: The evolution of a paradigm. *Behavioural Processes*, 123, 4–14. <https://doi.org/10.1016/j.beproc.2015.09.010>
- Wasserman, E. A., Bhatt, R. S., Chatlosh, D. L., & Kiedinger, R. E. (1987). Discrimination of and memory for dimension and value

- information by pigeons. *Learning and Motivation*, 18, 34–56. [https://doi.org/10.1016/0023-9690\(87\)90022-1](https://doi.org/10.1016/0023-9690(87)90022-1)
- Wasserman, E. A., & Biederman, I. (2012). Recognition-by-components: A bird's eye view. In *How animals see the world: Comparative behavior, biology, and evolution of vision* (pp. 191–215). New York, NY, US: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195334654.003.0012>
- Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, 63, 165–174. <https://doi.org/10.1901/jeab.1995.63-165>
- Whitman, C. O. (1919). *The behavior of pigeons*. (H. A. Carr, Ed.) (Vol. 3). Carnegie Institution of Washington.
- Wolfe, J. M. (2016). Use-inspired basic research in medical image perception. *Cognitive Research: Principles and Implications*, 1. <https://doi.org/10.1186/s41235-016-0019-2>
- Zabala-Travers, S., Choi, M., Cheng, W.-C., & Badano, A. (2015). Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images. *Medical Physics*, 42, 2942–2954. <https://doi.org/10.1118/1.4921125>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.