The role of category density in pigeons' tracking of relevant information

Cassandra L. Sheridan¹ · Leyre Castro¹ · Sol Fonseca^{1,2} · Edward A. Wasserman¹

Published online: 4 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

Prior categorization studies have shown that pigeons reliably track features that are relevant to category discrimination. In these studies, category exemplars contained two relevant and two irrelevant features; therefore, category density (specifically, the relevant to irrelevant information ratio) was relatively high. Here, we manipulated category density both between and within subjects by keeping constant the amount of relevant information (one feature) and varying the amount of irrelevant information (one or three features). One group of pigeons started with low-density training, then proceeded to high-density training, and finally returned to low-density training (Low-High-Low); a second group of pigeons started with high-density training and then proceeded to low-density training (High-Low). The statistical density of the category exemplars had a large effect on pigeons' performance. Training with high-density exemplars greatly benefitted category learning. Accuracy rose faster and to a higher level with high-density training than with low-density training; the percentage of relevant pecks showed a very similar pattern. In addition, high-density training (in the Low-High-Low group) led to an increase in performance on the more difficult low-density task, an observation reminiscent of the *easy-to-hard* effect. These results illuminate factors affecting pigeons' accuracy and tracking of relevant information in visual categorization.

Keywords Categorization · Category structure · Statistical density · Selective attention · Easy-to-hard effect · Pigeons

Introduction

Category-relevant information refers to stimulus elements or dimensions (e.g., color, shape, size) that help establish membership in one or another category. Most theories of discrimination learning and categorization hypothesize that attention, in both animals and humans, must be allocated to the relevant features of training stimuli for learning to occur (Kruschke, 1992; Lawrence, 1949; Mackintosh, 1965, 1975).

There are many ways to understand the term “attention.” The psychologist William James (1890/1950) defined attention as “the taking possession of the mind, in clear and vivid form, of one out of what may seem several simultaneously possible objects or trains of thought... It implies withdrawal from some things in order to deal effectively with others (p. 403).” To simplify James' notion, most theorists now propose that attention refers to how we actively process some

information in our environment while ignoring other information. Defining attention is relatively easy compared to measuring attention, because attention cannot be measured directly – it must be inferred from an organism's behavior.

In human studies, eye movement has been suggested to represent a direct measure of attention, because there seems to be a clear connection between the direction of eye gaze and changes in attention (Rehder & Hoffman, 2005). That is, as one's attention switches among different kinds of stimuli, one's eye movements correspondingly follow. But, this way of measuring attention in humans cannot be applied to most animals, because the appropriate technology is not yet available. So, how can we more directly measure attention in animals – specifically, in pigeons?

In the past, many studies have recorded the rate of pecking assuming that, if a stimulus supports a high rate of pecking, more attention is allocated to it (Pearce, Esber, George, & Haselgrove, 2008). Even earlier, Wasserman (1974) presented pigeons with two compound stimuli, AX and BX, where X was a white-colored key, and A and B were red- or green-colored keys. AX was always paired with food reinforcement, whereas BX was never paired with food reinforcement, thereby making X an irrelevant stimulus for the discrimination. The stimuli were presented on spatially separated response keys, so that the rate of responding to A, B, and X could be

✉ Leyre Castro
leyre-castroruiz@uiowa.edu

¹ Department of Psychological and Brain Sciences, The University of Iowa, 301 East Jefferson Street, Iowa City, IA 52242, USA

² Present address: University of Puerto Rico, San Juan, Puerto Rico

separately monitored during the acquisition of the discrimination. Wasserman observed that, when A and X were presented together, the rates of pecking to the relevant stimulus A and to the common irrelevant stimulus X were similar at the beginning of training, but they quickly started to diverge as training continued. By the end of training, pigeons were directing all of their pecks to the relevant stimulus A (for related results see Allan, 1993; Bermejo & Zeigler, 1998; Jenkins & Sainsbury, 1970), suggesting a robust connection between pigeons' pecks and relevant stimulus information. More recently, Dittrich, Rose, Buschmann, Bourdonnais, and Güntürkün (2009) used touchscreen technology to track the location of pigeons' pecks to complex visual stimuli. Similar to Wasserman's report, as pigeons' performance improved, the birds increasingly directed their pecks at the diagnostic areas in the stimuli (see also Wills et al., 2009).

Based on these findings, we recently examined the validity of peck tracking as a way of measuring attention in pigeons. In Castro and Wasserman (2014), pigeons learned to discriminate among exemplars from two different visual categories. The category exemplars always contained four features: two relevant, presented only on exemplars of one given category, and two irrelevant, presented equally often in both categories. When one category exemplar was presented on the screen, the pigeons had to peck at it multiple times. The only "active" areas were those that contained either the relevant or the irrelevant features, but the pigeons were free to peck at any of the features in any given way. Castro and Wasserman found that the pigeons not only learned to categorize the complex visual stimuli, but, in doing so, their pecking at the relevant and irrelevant features also yielded meaningful results. As training proceeded, the pigeons increasingly directed their pecks at the relevant features of the stimuli, suggesting that they were tracking the relevant information to solve the task. Therefore, these findings indicate that peck location could be an appropriate measure of attention in pigeons, much as eye tracking is an appropriate measure of attention in humans (see also Castro & Wasserman, 2016, 2017).

One way of describing the structure of different categories relies on attending to what Kloos and Sloutsky (2008) called statistical density, that is, the proportion of category-relevant information to the total amount of information provided. Statistical density is thus an easily quantifiable concept that can capture graded differences between categories. In our prior studies, the category exemplars contained two relevant features out of a total of four features, so that category density was relatively high. Categories that are dense have various intercorrelated features that are relevant for category learning; categories that are sparse have only one or a few relevant features, whereas the rest of the features are irrelevant (Kloos & Sloutsky, 2008). Therefore, it should be easier to learn categories that are statistically dense than categories that are statistically sparse.

In the current study, we explored how the density of category information affects learning and relevant feature tracking. We manipulated density by varying the number of irrelevant features in the category exemplars, holding constant the number of relevant features, only one in the current study. Two groups of pigeons had to between two different categories, with exemplars similar to those in Castro and Wasserman (2014, 2016, 2017). Here, one group was first presented with category exemplars containing one relevant feature and three irrelevant features, so that the density of the categories was relatively low; density was later increased, and finally reduced again (group Low-High-Low). A second group was first presented with category exemplars containing one relevant and one irrelevant feature, so that the density of the categories was relatively high, and then the density of the categories was reduced (group High-Low; see low- and high-density category exemplars in Figs. 1 and 2). We hypothesized that pigeons would learn the category discrimination and track the relevant features that help them solve the task with both low- and high-density categories. However, we hypothesized that learning and tracking should be easier with high-density categories than with low-density categories.

Method

Subjects

The subjects were eight homing pigeons (*Columba livia*) maintained at 85% of their free-feeding weight by controlled daily feedings. The eight pigeons were randomly distributed into two groups of four pigeons each. Group Low-High-Low was trained with low-density exemplars for 100 days, high-density exemplars for 50 days, and back again to low-density exemplars for 20 days. Group High-Low was trained with high-density exemplars for 50 days and low-density exemplars for 20 days. The pigeons had served in unrelated studies¹ prior to the present project. All procedures were approved by the Institutional Animal Care and Use Committee at the University of Iowa.

¹ All of the birds had taken part in an experiment involving a category discrimination of stimuli composed of different geometrical features in black and white stimuli (unpublished study). In addition, one of the birds in the Low-High-Low group, bird 52B, had also taken part in a project involving matching colored stimuli (Castro, L., Wasserman, E. A., & Lauffer, M. C., Unsupervised learning of complex associations in an animal model. *Cognition*, 178, 28–33, 2018). Furthermore, one of the birds in the High-Low group, bird 81B, had also taken part in a project involving a discrimination of black and white 3-D cubes (Navarro, V., Zhao, X., & Wasserman, E. A. Integrability and separability of visual dimensions in the pigeon. Manuscript in preparation). All of these experiments were run in experimental boxes with touchscreen technology.

Fig. 1 Examples of Category A and Category B low-density exemplars. All of the low-density exemplars contained only one relevant feature (the rainbow for Category A and the green spiral for Category B) and three irrelevant features, so the ratio of relevant to irrelevant features was low

(1:3) and, therefore, their statistical density was relatively low. On half of the exemplars, the features were presented in the corners of the square display, whereas on the other half of the exemplars, the features were presented in the center of the lines forming the square

Apparatus

The experiment used four $36 \times 36 \times 41$ cm operant conditioning chambers as detailed by Gibson, Wasserman, Frei, and Miller (2004). The chambers were located in a dark room with continuous white noise. Each chamber was equipped with a 15-in. LCD monitor located behind an AccuTouch® resistive touchscreen (Elo TouchSystems, Fremont, CA, USA). The portion of the screen that was viewable by the pigeons was 28.5×17.0 cm (970×640 pixels). Pecks to the touchscreen were processed by a serial controller board outside the box. A rotary dispenser delivered 45-mg pigeon pellets through a vinyl tube into a food cup located in the center of the rear wall opposite the touchscreen. Illumination during the experimental sessions was provided by a house light mounted on the upper rear wall of the chamber. The pellet dispenser and house light were controlled by a digital I/O interface board. Each chamber was controlled by its own Apple® iMac® computer. Programs to run this experiment were developed in MATLAB® with Psychtoolbox-3 extensions (Brainard, 1997; Pelli, 1997; <http://psychtoolbox.org/>).

Stimuli

A total of eight multicolored 3×3 cm squares (features) were used to create the different category training exemplars. Based

on statistical density, there were two different types of stimuli: low-density exemplars and high-density exemplars. Each type of exemplar used the same features, but with different numbers of irrelevant features shown at a given time. The low-density exemplars contained a total of four features: one that was relevant to the categorization task and three more that were irrelevant. So, the ratio of relevant to irrelevant features was low (1:3) and, therefore, the statistical density of these exemplars was relatively low. The high-density exemplars contained only two features: one that was relevant to the task and the other that was irrelevant. So, the ratio of relevant to irrelevant features was high (1:1) and, therefore, the statistical density of these exemplars was relatively high.

For both low- and high-density stimuli, each of the exemplars was created by placing one different feature into either the two or four possible locations of each of two possible spatial configurations. For the low-density exemplars, each feature was randomly placed into each of the four corners (corners configuration) or into the center of the lines (cross configuration) of an invisible 12×12 cm square (see Fig. 1); each of the features was 6 cm apart (both vertically and horizontally) from the two adjacent features, and they were connected by a white line. The high-density exemplars had the same two configurations as the low-density exemplars, but only two locations were used at the same time (see Fig. 2).

Fig. 2 Examples of Category A and Category B high-density exemplars. All of the high-density exemplars contained one relevant feature (the rainbow for Category A and the green spiral for Category B) and one irrelevant feature, so the ratio of relevant to irrelevant features was high

All possible combinations of features and locations were used to create the exemplars in each category (see Figs. 1 and 2 for a representative sampling of the many possible exemplars).

Thus, for both low- and high-density exemplars, there was one relevant feature that defined Category A and a different relevant feature that defined Category B. In addition, there were a total of six irrelevant features that were common to Categories A and B; they varied from trial to trial, and they appeared equally often within each of the category exemplars. Each of the relevant and irrelevant features appeared equally often in each of the locations. So, critically, spatial location could not be used as a cue for where the relevant features would be shown.

Procedure

Low-density training Group Low-High-Low was initially trained with low-density exemplars for a total of 100 days. Daily training sessions comprised 96 trials: half presented Category A exemplars and half presented Category B exemplars, in a random fashion, with no constraints on the trial sequence. At the beginning of a trial, the pigeons were presented with a start stimulus, a white square (3×3 cm) in the middle of the computer screen. After one peck anywhere on this white square, one category exemplar was displayed in the center of the screen. The pigeons had to satisfy an observing

(1:1) and, therefore, their statistical density was relatively high. On half of the exemplars, the features were presented in the corners of the square display, whereas on the other half of the exemplars, the features were presented in the center of the lines forming the square

response requirement (gradually increased from 1 to 15 during the first days of training) to any of the features – relevant or irrelevant – in the display. Only pecks within any of the four features' area were deemed valid. We recorded the location of these pecks in order to determine whether or not the pigeons selectively directed their pecks to the relevant features of the category exemplars.

On completion of the observing response requirement, two report buttons appeared 4.5 cm to the left and right of the category exemplar. The report buttons were 2.3×6 cm rectangles filled with two distinctive black and white patterns. From trial to trial, the buttons were randomly located, to the left or right of the category exemplar, in order to reduce any bias to peck the features adjacent to the report buttons. The pigeons had to select one of the two report buttons, depending on the category presented. If the choice response was correct, then food reinforcement was delivered and the intertrial interval (ITI) ensued; the ITI randomly ranged from 6 to 10 s. If the choice response was incorrect, then food was not delivered, the house light darkened, and a correction trial was given. Correction trials were given until the correct response was made. No data were analyzed from correction trials.

High-density training Group Low-High-Low was given high-density exemplars in their second phase of training for a total of 50 days. Group High-Low started training at

this point, also for a total of 50 days. Daily training sessions comprised 120 trials: half presented Category A exemplars and half presented Category B exemplars, in a random fashion. The procedure was the same as in Phase 1 except for the density of the stimuli presented.

Low-density training In this phase, Group Low-High-Low returned to initial training, in which low-density exemplars were presented. Group High-Low was presented with low-density exemplars as well, but this was their first encounter with these stimuli. The procedure was the same as in the initial low-density phase. This phase lasted 20 days for both groups.

Data analysis

Data are available at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XG4Y5Y>

The data were subjected to linear mixed-effects (i.e., multilevel or random coefficients) analyses, using restricted maximum likelihood (REML). This type of analysis is especially well suited to analyze training data because it appropriately treats “training session” as a continuous rather than as a categorical variable. Mixed-effects models extend the standard linear framework by adding random effects (random intercepts and random slopes specific to the subjects taking part in an experiment) to the fixed effects (the independent variables familiar in traditional analyses). Thus, mixed-effects models allow one to take into account each subject’s variability by computing a random intercept and/or a random slope for each subject and thereby ensure the best estimates of the fixed effects. To select an appropriate random-effects structure (only random intercepts or random slopes as well), we compared models with the same fixed-effects structure and varying complexity in their random-effects structure using the log likelihood ratio test (see Wagenmakers & Farrell, 2004). All analyses were conducted using the lme4, version 1.1-17 (Bates et al., 2018) and lmerTest, version 3.0-1 (Kuznetsova, Brockhoff, & Christensen, 2018), packages of R, version 3.3.2 (R Core Team, 2016).

Results

Accuracy

Figure 3 shows mean accuracy in all training phases for both groups. Accuracy in group Low-High-Low started at chance level (50%), it slowly rose until it reached about 75% after 50 days of low-density training, and then started to level off and stayed at about 75% for the remainder of this first phase. When group Low-High-Low was moved to the high-density phase, in which exemplars contained one relevant and one irrelevant feature, its

accuracy quickly rose to over 90%. Group High-Low, which began on the high-density phase, started at chance level, and its accuracy quickly increased until it also reached levels over 90% at the end of the high-density training phase.

After 50 days of high-density training, both groups were moved to low-density training. Their accuracy levels dropped slightly after the decrease in the density of the exemplars, but they remained just a bit below 90% correct. So, high-density training helped promote greater accuracy than would have been expected from continued training with the low-density exemplars (see accuracy for group Low-High-Low at the end of their first phase and in the last training phase). Moreover, both groups ended at virtually the same accuracy level, suggesting that the extra initial 100 days of training with the low-density exemplars did not provide group Low-High-Low with any additional performance advantage.

In order to unpack the differences attributable to low- and high-density training, we first examined accuracy during the first 50 training days in both groups, so we could compare initial training with either low-density (group Low-High-Low) or high-density (group High-Low) exemplars. This comparison is shown in Fig. 4. We used a linear mixed-effects model, in which group (treatment coded, group Low-High-Low = 0) and session (1–50) were the fixed effects. In this and all subsequent analyses with session as a fixed effect, session was rescaled by subtracting 1, so that the intercept was moved to the first session. The maximal random effects structure supported by the data included random intercepts for bird; random slopes were not necessary. Overall, accuracy improved over these first 50 days of training ($b = 0.49$, 95% CI [0.42, 0.55], $t(790) = 17.06$, $p < .001$), and accuracy in group High-Low was higher overall than in group Low-High-Low ($b = 19.64$, 95% CI [6.41, 32.88], $t(6.3) = 2.86$, $p = .02$).² The Session $\times$ Group interaction was statistically significant as well ($b = 0.16$, 95% CI [0.07, 0.26], $t(790) = 2.86$, $p < .001$). Clearly, training with high-density exemplars made the categorization task easier than training with low-density exemplars. At the end of the 50-day period, group High-Low reached 97.26% correct compared to 73.38% correct in group Low-High-Low.

Group Low-High-Low was trained for 50 days more with the low-density exemplars, but its accuracy barely improved over those days. During the last 20 days of initial low-density training, its overall accuracy was 76.41% correct. A linear mixed-effects model on accuracy over these final 20 days with session as a fixed effect and random intercepts for bird showed

² b coefficients represent the model estimates for each fixed effect or interaction. Because our fixed effects are treatment coded (Session 1 = 0 and group Low-High-Low = 0), here $b = 0.49$ for Session indicates that accuracy increases, overall, by 0.49 points per day; and $b = 19.64$ indicates that overall accuracy is 19.64 points higher in the High-Low than in the Low-High-Low group.

Fig. 3 Mean percentage of correct responses for both group Low-High-Low and group High-Low throughout all the training phases. Note that group High-Low started training in the high-density phase. The dashed

line, at 50%, represents the chance level for correct responses. Error bars indicate the standard errors of the means (± 1 SEM)

that the effect of session was not statistically significant ($b = 0.12$, 95% CI $[-0.02, 0.27]$, $t(75) = 1.73$, $p = .09$), that is, the birds' accuracy had ceased to improve.

Thus, after 100 days of low-density training, we moved group Low-High-Low to high-density training, a change that very quickly resulted in large increases in accuracy (see the transition between the first and second training phases in Fig. 3). Overall accuracy during the first 20 days of high-density training reached 85.84%. We used a linear mixed-effects model to compare accuracy during the last 20 days of low-density training to accuracy during the first 20 days of high-density

training, in which phase (treatment coded, low-density phase = 0) and session were the fixed effects, and random intercepts for bird were included. The analysis confirmed that accuracy during the high-density phase was higher than accuracy during the low-density phase ($b = 7.49$, 95% CI $[4.85, 10.14]$, $t(153) = 5.53$, $p < .001$). No other effects were statistically significant. Thus, moving the birds in group Low-High-Low from low- to high-density training greatly benefitted their categorization performance.

At the end of the high-density phase, accuracy in both group Low-High-Low and group High-Low was very high ($M = 92.46\%$ and $M = 91.98\%$, for group Low-High-Low and group High-Low, respectively). A linear mixed-effects model with group (treatment coded, group Low-High-Low = 0) and session as fixed effects, and random intercepts for bird, confirmed that there were no differences in accuracy between the groups ($b = -3.78$, 95% CI $[-13.69, 6.11]$, $t(6.24) = -0.73$, $p = .49$). The Session $\times$ Group interaction was statistically significant ($b = 0.34$, 95% CI $[0.20, 0.49]$, $t(150) = 4.67$, $p < .001$), because of a very small advantage of the Low-High-Low group at the beginning of these last 20 days that at the end had turned into a very small advantage for the High-Low density group. Thus, both groups ended the high-density phase at essentially the same accuracy level.

Next, we compared both groups in the last low-density phase. Low-density exemplars presented in this phase were novel for group High-Low, but familiar for group Low-High-Low. Regardless of that difference in prior training with low-density exemplars, overall accuracy was very similar in both groups ($M = 88.94\%$ and $M = 87.68\%$, for group Low-High-Low and group High-Low, respectively). A linear mixed-effects model with group (treatment coded, group Low-High-Low = 0) and session as fixed effects, and random

Fig. 4 Mean percentage of correct responses for both groups on their respective first 50 training sessions: low-density training for group Low-High-Low and high-density training for group High-Low. The dashed line, at 50%, represents the chance level for correct responses. Error bars indicate the standard errors of the means (± 1 SEM)

intercepts and slopes for bird, yielded no statistically significant effects. So, the initial 100 sessions of low-density training for group Low-High-Low did not provide this group with any performance advantage.

Finally, for group Low-High-Low, we compared the last 20 sessions of the initial low-density phase (during which, as indicated above, accuracy seemed to have reached asymptote) to the last 20 sessions of low-density training *after* high-density training. A linear mixed-effects model with phase (treatment coded, low-density phase = 0) and session as fixed effects, and a random intercept for bird, showed that accuracy in the last phase was higher ($M = 88.94\%$) than accuracy in the last 20 sessions of the initial phase ($M = 76.41\%$) ($b = 13.17$, 95% CI [8.98, 17.37], $t(153) = 6.12$, $p < .001$). Thus, it is fair to conclude that training with high-density exemplars, rather than merely extended training, helped group Low-High-Low attain higher levels of performance with the low-density exemplars.

Relevant pecks

Next, we looked at the birds' pecks at the relevant features. We calculated the birds' percentage of pecks at the relevant features (*relevant pecks*) over the total number of daily pecks. Low-density stimuli contained four features, of which only one was relevant; so, the chance level of relevant pecking was 25%. High-density stimuli contained two features, of which one was relevant; so, the chance level of relevant pecking was 50%. Figure 5 shows the mean percentage of relevant pecks in all training phases for the two groups. Please note that the birds were *not* differentially reinforced depending on which features they pecked; they could peck at any of the relevant or irrelevant features to complete the

observing requirement before making their category choice response. Nonetheless, as training proceeded, birds in both groups gradually increased their pecks at the relevant features. Relevant pecks in group Low-High-Low started at chance level (25%), gradually rose until they reached about 50% after 50 days of low-density training, and then started to level off and stayed at about 55% for the remainder of this first phase. When group Low-High-Low was moved to the high-density phase, in which exemplars contained one relevant and one irrelevant feature, its relevant pecks quickly rose to approximately 80%. Group High-Low started at chance level (in this case, 50%, because they started with high-density training), and their accuracy increased quickly until they reached approximately 90% at the end of the high-density training phase. When both groups were finally moved to low-density training, their percentage of relevant pecks dropped slightly, but still ended up at about 70% for group Low-High-Low and at about 80% for group High-Low. Just as with accuracy, high-density training for group Low-High-Low seemed to promote greater tracking of the relevant cues than would have been expected from continued training with the low-density exemplars (see the end of the initial low-density phase compared to the last low-density phase).

As we did with accuracy, we first examined relevant pecks during the first 50 training days in both groups (depicted in Fig. 6), so we could compare initial training with either low-density (group Low-High-Low) or high-density (group High-Low) exemplars. A direct comparison of the percentage of relevant pecks was inappropriate because of the difference in chance level between low- and high-density training. So, we transformed the percentage of relevant pecks to the signal detection measure d' (Algorithm 1; Smith, 1982), which has

Fig. 5 Mean percentage of relevant pecks for both group Low-High-Low and group High-Low throughout all the training phases. Note that group High-Low started training in the high-density phase. The dashed line at 25% represents the chance level for training with low-density exemplars,

whereas the dashed line at 50% represents the chance level for training with high-density exemplars. Error bars indicate the standard error of the means (± 1 SEM)

Fig. 6 Mean percentage of relevant pecks for both groups on their respective first 50 training sessions: low-density training for group Low-High-Low and high-density training for group High-Low. The dashed line at 25% represents the chance level for group Low-High-Low, whereas the dashed line at 50% represents the chance level for group High-Low. Error bars indicate the standard errors of the means (± 1 SEM)

been used with a large variety of tasks and procedures (Green & Swets, 1966). After the transformation, the chance level corresponded to a d' of 0.00 in both conditions.

Now, we analyzed the differences in d' between groups for relevant pecks. A linear mixed-effects model with group (treatment coded, group Low-High-Low = 0) and session as fixed effects, and random intercepts for bird showed a main effect of session ($b = 0.027$, 95% CI [0.022, 0.033], $t(383) = 10.71$, $p < .001$) and, most importantly, a Session $\times$ Group interaction, ($b = 0.028$, 95% CI [0.021, 0.035], $t(383) = 7.63$, $p < .001$), due to the two groups starting at chance level, but group High-Low rising faster and to a higher point than group Low-High-Low.

We also examined relevant pecks in group Low-High-Low during the last 20 sessions of their initial low-density training. A linear mixed-effects model on relevant pecks with session as a fixed effect and a random intercept for bird revealed that the effect of session was not statistically significant ($b = 0.04$, 95% CI [-0.11, 0.20], $t(75) = 0.59$, $p = .55$); that is, the birds' relevant pecks ($M = 54.93\%$) seemed to have ceased to increase during the final sessions of initial low-density training.

When we moved group Low-High-Low to high-density training, their percentage of relevant pecks greatly increased ($M = 75.91\%$ in the first 20 sessions). Of course, this increase could merely be due to the difference in chance level: from 25% in the low-density phase to 50% in the high-density phase. However, a linear mixed-effects model with session as a fixed effect and random intercepts for bird confirmed the effect of session during the first 20 sessions of high-

density training ($b = 0.39$, 95% CI [0.26, 0.53], $t(79) = 5.62$, $p < .001$). So, the percentage of relevant pecks that had ceased to improve at the end of the low-density phase ($b = 0.04$, see above) started to rise again when the birds in group Low-High-Low were moved from low-density to high-density training. Indeed, by the end of the high-density phase, relevant pecks in group Low-High-Low were quite high ($M = 81.51\%$ in the last 20 sessions), as were relevant pecks in group High-Low ($M = 91.92\%$ in the last 20 sessions). In order to analyze the differences in relevant pecks at the end (last 20 sessions) of high-density training, we conducted a linear mixed-effects model with group (treatment coded, group Low-High-Low = 0) and session as fixed effects, and random intercepts and slopes for bird. The analysis showed no effects of group or session, but the Session $\times$ Group interaction was statistically significant ($b = 0.65$, 95% CI [0.43, 0.88], $t(6) = 2.61$, $p = .04$), because at the end of the high-density training, relevant pecks, which had reached a similar high point in the middle of the phase, increased slightly for group High-Low but decreased slightly for group Low-High-Low.

Next, we compared relevant pecks in both groups in the last low-density phase. Regardless of group Low-High-Low having initially been trained with low-density exemplars for 100 sessions, their percentage of relevant pecks was lower ($M = 71.53\%$) than that in group High-Low ($M = 81.08\%$), which was never previously presented with low-density exemplars. Despite this numerical difference, a linear mixed-effects model with group (treatment coded, group Low-High-Low = 0) and session as fixed effects, and random intercepts and slopes for bird, yielded no statistically significant effects.

Finally, for group Low-High-Low, we compared the last 20 sessions of the initial low-density phase (during which, as indicated above, relevant pecks seemed to have ceased to increase) to the 20 sessions of low-density training after high-density training. A linear mixed-effects model with phase (treatment coded, low-density phase = 0) and session as fixed effects, and random intercepts for bird, showed that the percentage of relevant pecks during the last phase was higher ($M = 71.53\%$) than during the last sessions of the initial phase ($M = 54.93\%$) ($b = 16.81$, 95% CI [11.62, 21.99], $t(153) = 6.33$, $p < .001$). Thus, it is fair to conclude that, just as with accuracy, training with high-density exemplars, rather than merely extended training, enhanced group Low-High-Low's tracking of relevant features with low-density exemplars.

Discussion

Pigeons' categorization accuracy reliably increased throughout the course of training, regardless of their being trained with low-density or high-density exemplars (Figs. 3 and 4). Moreover, pigeons not only learned to categorize the training stimuli, but they also learned which features were relevant and

which features were irrelevant to solve the task, in both groups given initial low-density or high-density training (Figs. 5 and 6). These results replicate, with different levels of density, the results of our prior studies in which we also reported that, as training progressed, pigeons' accuracy increased as did their pecks at the relevant features of the category exemplars (Castro & Wasserman, 2014, 2016, 2017). All of these data thus suggest that the birds were tracking the relevant information to solve the categorization task, as predicted by attentional theories of learning and categorization (e.g., George & Pearce, 2012; Kruschke, 1992; Le Pelley, Mitchell, Beesley, George, & Wills, 2016; Mackintosh, 1975).

Category density

The statistical density of the category exemplars proved to have a large effect on the pigeons' performance. Training with high-density exemplars greatly benefitted category learning. Not only did accuracy in the first 50 sessions rise faster and to a higher level with high-density training than with low-density training, but the percentage of relevant pecks did as well, in a parallel way.

Statistical density computes the ratio of information that is relevant for category membership to the total amount of information available (see Kloos & Sloutsky, 2008 for a detailed explanation). In the past, Garner (1962) considered that learning about a set of stimuli is a function of the internal structure of the set, in particular feature redundancy. In a similar vein, Trabasso and Bower (1968) argued that the proportion of category-relevant to category-irrelevant information determines the efficiency of category learning.

It may not be surprising that high-density categories are learned better than low-density categories. Because in high-density stimuli the amount of category-relevant information is large, we could also argue that learning highly dense categories puts small demands on selective attention. In contrast, learning low-density categories requires the organism to ignore a large amount of category-irrelevant information while focusing at the same time on category-relevant information, thereby increasing the demands on selective attention. Thus, the larger the proportion of irrelevant information, the more difficult it would be to ascertain what should be ignored and, as a consequence, the more difficult learning should be.

However, it is possible to argue that our low-density exemplars, which contained four features, were visually more complex stimuli than our high-density exemplars, which contained only two features; it may have been this difference in visual complexity that made low-density training more difficult. The high-density exemplars had one relevant and one irrelevant feature, so that 50% of the information presented in each exemplar was category relevant; we could have presented high-density exemplars with four features, two of them relevant and two of them

irrelevant, maintaining the proportion of relevant information at 50%. Indeed, this is just what we did in Castro and Wasserman (2014, 2016), where several groups of pigeons (two in 2014, and four in 2016) were trained with high-density exemplars containing two relevant and two irrelevant features. Despite the limitations involved in making comparisons among different experiments, we consider it instructive to examine our pigeons' learning rates in those prior studies. Previously, the number of sessions to reach 75% accuracy ranged from 8 to 12, and to reach 85% accuracy, the number of sessions ranged from 10 to 24. In the current experiment, the High-Low group initially took eight sessions to reach 75% accuracy and 12 sessions to reach 85% accuracy, a range of sessions within that previously reported. As for relevant pecks, the number of sessions to reach 70% relevant pecks ranged from 7 to 18, and the number of sessions to reach 75% relevant pecks ranged from 15 to 20 in prior studies. In the current experiment, the High-Low group initially took 11 sessions to reach 70% relevant pecks, and 15 sessions to reach 75% relevant pecks; again, this number of sessions is within the range of sessions previously reported. Thus, we do not believe that it is visual complexity per se, but statistical density that is mainly responsible for the initial differences in difficulty between our low- and high-density categories.

Nonetheless, we did find a noteworthy difference between our prior reports and the performance of the High-Low group in the current experiment. With the high-density exemplars containing only two features, the percentage of relevant pecks exceeded 90% in the High-Low group at the end of high-density training compared to asymptotic percentages between 75% and 80% in prior studies (Castro & Wasserman, 2014, 2016). The reason for this large increase in tracking of relevant features could have been that, when the features appear in only two out of the four possible locations, there is a considerable percentage of the trials (approximately 16%) in which there are *no* features available in the bottom locations (see Fig. 2), whereas there were always two features available in the bottom locations in prior studies. We have frequently observed that some birds have a strong preference for pecking at the bottom locations of the stimulus configuration, a preference that conflicts with pecking at the relevant features when they appear in the top locations. Being forced, on some trials, to peck at the top locations because there are no features available in the bottom locations may have lessened the bottom bias, thereby helping improve the present birds' tracking of the relevant features.

Easy-to-hard effect

Importantly, not only did high-density training encourage faster and higher accuracy and relevant tracking, but high-

density training proved to help promote later performance on low-density trials as well. Group Low-High-Low had apparently reached asymptotic performance at the end of its 100 sessions of low-density training, with accuracy at approximately 75% and relevant pecks at 55%. However, accuracy and relevant pecks greatly improved when this group was moved to high-density training and, most critically, when they were returned to low-density training; now, accuracy and relevant pecks stayed high, far higher than at the end of initial low-density training (see Figs. 3 and 5). Thus, decreasing the difficulty of the task led to an increase in performance on the more difficult task that did not seem was going to happen with extended training alone.

Our results are reminiscent of the easy-to-hard effect, originally reported by Pavlov (1927). Using different visual and auditory stimuli, Pavlov reported that a discrimination involving stimulus values that are nearby along a dimension proved to be easier to learn if prior training had been given involving stimuli with values that are farther apart along the same dimension. For example, one of Pavlov's dogs failed to discriminate a circle from a very round ellipse; however, when the dog was trained with more elongated ellipses and the shape of the ellipse was gradually changed, in four steps, to the original very round ellipse, the dog could now successfully discriminate the shapes. Thus, initial training with an easy version of the discrimination task facilitated subsequent learning of the more difficult version of the task.

In the first formal experimental exploration of the easy-to-hard phenomenon, Lawrence (1952) trained different groups of rats on a brightness discrimination task. Rats initially trained with distant brightness values before being trained with adjacent values learned faster and made fewer errors than rats trained only with the adjacent brightness values for the same total amount of time. Especially noteworthy for our current project, Lawrence argued that, in discrimination learning, it is critical “for the animal to isolate functionally the relevant stimulus dimension from all the other background and irrelevant cues” (p. 516), something that it is more achievable when the discrimination is easy, be that because of a larger perceptual difference in the stimuli or because of a higher proportion of relevant information, as in our high-density exemplars.

Demonstrations of this easy-to-hard phenomenon ensued, with different species and different stimulus modalities (see Walker, Lee, & Bitterman, 1990). However, all these explorations of the easy-to-hard effect focused on perceptual discriminations involving stimuli varying along the same dimension. To the best of our knowledge, our current results are the first demonstration of the easy-to-hard effect in a complex categorization task. The practical implications are significant for many areas of animal cognition and comparative psychology. When we examine and compare cognitive performance in humans and other animals, we tend to ignore the long process

of scaffolding that humans go through, starting in their very early infancy. Providing animals with initial and gradually challenging experiences – as long as it is feasible – may permit us to observe cognitive achievements that we currently deem to be beyond their reach (see Smirnova, Zorina, Obozova, & Wasserman, 2015, for a possible example of such scaffolding in relational concept learning).

A final note about peck tracking

It could be argued that pigeons tend to peck at the relevant cues because they are strong predictors of the outcome and, consequently, they have acquired high associative strength. Indeed, relevant cues are said to be relevant precisely because they are reliable and strong predictors of the correct outcome; so, it is difficult to disentangle attention from associative strength. To do so requires complex experimental designs. Uniquely, George and Pearce (1999) pursued this dissociation and demonstrated that the pigeons' allocation of attention to a particular stimulus was determined by its relevance to the solution of their visual discrimination, rather than by its correlation with the outcome. Although we cannot provide such an empirical distinction in our experiment, it is reasonable to suspect that the pigeons' processing mechanisms in George and Pearce's study are similar to those of the pigeons in our present study.

We began this line of experimental investigation to test the utility of peck tracking as a reliable proxy for attention in the pigeon (Castro & Wasserman, 2014, 2016, 2017). Most tests of selective attention are administered long after differential attention has been deployed during discrimination learning (e.g., Mackintosh & Little, 1969; Reynolds, 1961) or category learning (e.g., Shepard, Hovland, & Jenkins, 1961). However, to fully understand the dynamics of attention demands much more.

What we have found in the present experiment is that there are conditions in which relevant peck-tracking scores in excess of 90% are attainable. Indeed, these tracking scores are numerically comparable to pigeons' categorization accuracy scores themselves. We believe that obtaining such comparably high tracking and accuracy scores testifies to the validity of peck tracking as a proxy for attention (see Castro & Wasserman, 2014, 2016, 2017, for further discussion of the relationship between peck tracking and attention), thus opening the door to further exploring the role of selective attention in category learning.

Author Note Sol Fonseca is now at the University of Puerto Rico. This research was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the NIH under Award Number P01HD080679.

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

References

- Allan, R. W. (1993). Control of pecking response topography by stimulus-reinforcer and response-reinforcer contingencies. In H. P. Zeigler, & H.-J. Bischof (Eds.), *Vision, brain, and behavior in birds* (pp. 285–300). Cambridge, MA: MIT Press.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., & Green, P. (2018). lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1-17. Retrieved from <https://CRAN.R-project.org/package=lme4>
- Bermejo, R., & Zeigler, H. P. (1998). Conditioned "prehension" in the pigeon: Kinematics, coordination and stimulus control of the pecking response. *Behavioural Brain Research*, 91, 173–184. [https://doi.org/10.1016/S0166-4328\(97\)00121-6](https://doi.org/10.1016/S0166-4328(97)00121-6)
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Castro, L., & Wasserman, E. A. (2014). Pigeons' tracking of relevant attributes in categorization learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, 40, 195–211. doi:<https://doi.org/10.1037/xan0000022>
- Castro, L., & Wasserman, E. A. (2016). Attentional shifts in categorization learning: Perseveration but not learned irrelevance. *Behavioural Processes*, 123, 63–73. doi:<https://doi.org/10.1016/j.beproc.2015.11.001>
- 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. doi:<https://doi.org/10.1037/xan0000146>
- Dittrich, L., Rose, J., Buschmann, J. F., Bourdonnais, M., & Güntürkün, O. (2009). Peck tracking: A method for localizing critical features within complex pictures for pigeons. *Animal Cognition*, 13, 133–143. doi:<https://doi.org/10.1007/s10071-009-0252-x>
- Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- George, D. N., & Pearce, J. M. (1999). Acquired distinctiveness is controlled by stimulus relevance not correlation with reward. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 363–373. doi: <https://doi.org/10.1037/0097-7403.25.3.363>
- George, D. N., & Pearce, J. M. (2012). A configural theory of attention and associative learning. *Learning & Behavior*, 40, 241–254. doi: <https://doi.org/10.3758/s13420-012-0078-2>
- Gibson, B. M., Wasserman, E. A., Frei, L., & Miller, K. (2004). Recent advances in operant conditioning technology: A versatile and affordable computerized touch screen system. *Behavior Research Methods, Instruments*
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- James, W. (1890/1950). *The principles of psychology*. New York: Dover Publications.
- Jenkins H. M., & Sainsbury R. S. (1970). Discrimination learning with the distinctive feature on positive or negative trials. In D. I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis*. (pp. 239–273). New York: Appleton-Century-Crofts.
- Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, 137, 52–72. doi:<https://doi.org/10.1037/0096-3445.137.1.52>
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Journal of Experimental Psychology*, 99, 22–44. doi:<https://doi.org/10.1037/0033-295X.99.1.22>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). lmerTest: Tests in Linear Mixed Effects Models. R package version 3.0-1. Retrieved from <https://CRAN.R-project.org/package=lmerTest>
- Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, 39, 770–784. doi:<https://doi.org/10.1037/h0058097>
- Lawrence, D. H. (1952). The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology*, 45, 511–516. doi:<https://doi.org/10.1037/h0057135>
- Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*, 142, 1111–1140. <https://doi.org/10.1037/bul0000064>
- Mackintosh, N. J. (1965). Selective attention in animal discrimination learning. *Psychological Bulletin*, 64, 124–150. doi:<https://doi.org/10.1037/h0022347>
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298. doi:<https://doi.org/10.1037/h0076778>
- Mackintosh, N. J., & Little, L. (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science*, 14, 5–6. doi:<https://doi.org/10.3758/bf03336395>
- Pavlov, I. P. (1927). *Conditioned reflexes* (translated by G.V. Anrep). Oxford, UK: Oxford University Press.
- Pearce, J. M., Esber, G. R., George, D. N., & Haselgrove, M. (2008). The nature of discrimination learning in pigeons. *Learning & Behavior*, 36, 188–199. doi:<https://doi.org/10.3758/lb.36.3.188>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. doi:<https://doi.org/10.1163/156856897x00366>
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org>
- Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 51, 1–41. doi:<https://doi.org/10.1016/j.cogpsych.2004.11.001>
- Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior*, 4, 203–208. doi:<https://doi.org/10.1901/jeab.1961.4.203>
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75, 1–42. doi:<https://doi.org/10.1037/h0093825>
- Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology*, 25, 256–260. doi:<https://doi.org/10.1016/j.cub.2014.11.063>
- Smith, J. E. K. (1982). Simple algorithms for M-alternative forced choice calculations. *Perception & Psychophysics*, 31, 95–96. doi:<https://doi.org/10.3758/bf03206208>
- Trabasso, T. R., & Bower, G. H. (1968). *Attention in learning: Theory and research*. New York: Wiley.
- Walker, M. M., Lee, Y., & Bitterman, M. E. (1990). Transfer along a continuum in the discriminative learning of honeybees (*Apis mellifera*). *Journal of Comparative Psychology*, 104, 66–70. doi: <https://doi.org/10.1037/0735-7036.104.1.66>
- Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, 103, 284–297. doi:<https://doi.org/10.1037/h0036872>
- Wills, A. J., Lea, S. E. G., Leaver, L. A., Osthaus, B., Ryan, C. M. E., Suret, M. B., Bryant, C. M. L., Chapman, S. J. A., & Millar, L. (2009). A comparative analysis of the categorization of multidimensional stimuli: I. Unidimensional classification does not necessarily imply analytic processing; evidence from pigeons (*Columba livia*), squirrels (*Sciurus carolinensis*), and humans (*Homo sapiens*). *Journal of Comparative Psychology*, 123, 391–405. doi:<https://doi.org/10.1037/a0016216>