Published in final edited form as:

J Exp Psychol Anim Learn Cogn. 2017 July ; 43(3): 231–242. doi:10.1037/xan0000146.

Feature predictiveness and selective attention in pigeons' categorization learning

Leyre Castro and

Department of Psychological and Brain Sciences, The University of Iowa

Edward A. Wasserman

Department of Psychological and Brain Sciences, The University of Iowa

Abstract

Prior categorization studies have shown that pigeons reliably track features that are perfect predictors of category membership (Castro & Wasserman, 2014; 2016a). One might further ask whether pigeons would also track features that are relevant, but imperfect predictors of category membership. In our present project, pigeons had to categorize multiple exemplars from two different artificial categories, in which the exemplars were composed of four different features, that were associated with one of two different report responses. Each exemplar contained one feature that perfectly predicted category membership; one feature that imperfectly predicted category membership; and, two irrelevant features that did not predict category membership. We monitored pigeons' choice accuracy as well as the location of their pecks to each of the four exemplar features to determine to which attributes the birds attended. As categorization accuracy rose, pecks to the perfect predictor of each category rose as well. Pigeons also showed evidence of attending more to the imperfect predictor than to the irrelevant features, but to a lesser degree. Overall, our results provide evidence of selective attention in pigeons' categorization behavior.

Keywords

categorization; attention; peck tracking; animal cognition; pigeons

Attention is typically defined as selectively focusing on one aspect of the environment while ignoring others (e.g., Styles, 2006). Yet, it is not easy to empirically determine whether an organism is focusing on one or another aspect of the environment. Attention cannot be directly measured, but must instead be inferred from an organism's behavior, thereby making it difficult to provide clear and compelling evidence of momentary increases and decreases in attention.

Despite these challenges, innovative methods have been developed to try to monitor attention while learning is taking place. Because of empirical evidence suggesting that eye movements and attention operate simultaneously (e.g., Deubel & Schneider, 1996), eyetracking technology has been used to study the deployment of attention in humans while

category learning is ongoing. Rehder and Hoffman (2005), for example, tracked human participants' eye movements while they were solving a categorization task in which some elements were relevant to correctly classifying the category exemplars, whereas other elements were not. Rehder and Hoffman found that, as learning progressed, participants' allocation of attention gradually shifted toward the relevant elements of the stimuli. Participants' eye gaze direction thus revealed that they indeed allocated attention in a way that enhanced their classification performance.

In animals, Dittrich, Rose, Buschmann, Bourdonnais, and Güntürkün (2010) have used touchscreen technology to track pigeons' peck locations while the birds were learning a go-no go task. In this study, pigeons were required to discriminate pictures in which human figures were present (always followed by food reinforcement) from pictures in which human figures were absent (never followed by food reinforcement). Dittrich et al. found that, as pigeons' discriminative performance improved, they increasingly focused their pecks on areas occupied by human figures, the distinctively diagnostic feature of the pictorial stimuli. Dittrich et al.'s (2010) study strongly suggests that peck tracking might prove to be a powerful, albeit indirect, means to measure pigeons' allocation of visual attention, in much the same way as eyetracking is considered to be an effective, albeit indirect, measure of human visual attention.

To further test this possibility, Castro and Wasserman (2014) taught pigeons to classify exemplars from two different artificial categories. Some of the features in each of the exemplars were relevant for the categorization task, whereas other features were irrelevant. When a category exemplar was presented on the screen, the pigeons had to peck it several times. However, pigeons did not have to peck any specific feature of the category exemplar; pecks to both the relevant and irrelevant features equally satisfied the observing response requirement. We found that, as categorization accuracy increased, pecks to the relevant features of the category exemplars progressively increased as well; conversely, pecks to the irrelevant features progressively fell. These results suggest that, as pigeons were learning to categorize the stimuli, they were also paying increasing attention to their relevant attributes. As hypothesized by several learning models, attention shifted to those cues that proved to be reliable predictors of the outcome (e.g., George & Pearce, 2012; Kruschke, 1992; LePelley, Mitchell, Beesley, George, & Wills, 2016; Mackintosh, 1975).

Thus, Castro and Wasserman (2014; see also Castro & Wasserman, 2016a) demonstrated that pigeons do track features of exemplars that are relevant to solving a categorization task. In those previous studies, the relevant features for each category were present in all of the exemplars for that specific category; so, those relevant features were *perfect* predictors of each of the corresponding categories. One might wonder what would happen if a feature is a relevant, but *imperfect* predictor of a category. Pigeons might also track the imperfect predictor, but to a lesser extent, as a function of its capacity to predict the correct category. Or pigeons might ignore the imperfect predictor altogether if a perfect predictor is also available that allows them to master the task.

The effect of varying degrees of stimulus predictiveness on pigeons' discrimination learning was explored by Wasserman (1974) in an autoshaping procedure. Five groups of pigeons

were presented with two compound stimuli, AX and BX; X was a white-colored key that was irrelevant to the discrimination, whereas A and B were red- or green-colored keys that were relevant to varying degrees. Depending on the experimental group, A was paired with food reinforcement on 100%, 87.5%, 75%, 62.5%, or 50% of the trials, while B was respectively paired with food reinforcement on 0%, 12.5%, 25%, 37.5% and 50% of the trials. Therefore, A went from being a perfect predictor of food reinforcement to being uncorrelated with food reinforcement, whereas B went from being a perfect predictor of nonreinforcement to being uncorrelated with food reinforcement. The stimuli were presented on spatially separated response keys, so that the rate of responding to A, B, and X could be individually monitored during acquisition of the discrimination. When A and B were each paired with reinforcement 50% of the time, pigeons pecked them equally often. But, as the predictive values of A and B increased, the mean rate of pigeons' pecking A progressively increased and their mean rate of pecking B progressively decreased. Thus, the degree of predictiveness of elements A and B systematically influenced the pigeons' pecking responses.

In Wasserman (1974), the discriminative stimuli themselves were directly associated with reinforcement or nonreinforcement. We can also implement varying degrees of predictiveness within a categorization task, in which the stimuli serve an essentially informational function without themselves being directly associated with reinforcement or nonreinforcement and thereby exciting or inhibiting a behavioral response.

In our present project, pigeons had to learn to categorize stimuli from two different artificial categories, in which the exemplars were composed of four different features (cf. Castro & Wasserman, 2014). Here, there was one perfect predictor from each category; these relevant features predicted the correct category on 100% of the trials. There was also one imperfect predictor for each category that appeared along with the perfect predictor of a given category on 75% of the trials; on the other 25% of the trials, this imperfect predictor appeared along with the perfect predictor of the other, incorrect category. Thus, the imperfect predictor had a predictiveness value of 50%. The other two elements of the category exemplars were irrelevant features and appeared equally often in exemplars of both categories (see Figure 1 for a sampling of the category exemplars used in this study).

The critical issue was whether pigeons would track the category features based on their predictive value. Specifically, would most of the pecks be directed to the perfect predictor and, to a lesser degree, to the imperfect predictor, but not to either of the non-predictive features? Or would virtually all of the pecks be directed to the perfect predictor, because it was the only sufficient and unambiguous element to perfectly solve the category discrimination? Moreover, if the pigeons did not robustly peck the imperfect predictor, would they nevertheless learn about its task relevance?

It may seem obvious that organisms should focus their attention on the specific feature or features that are informative for solving a task. However, a category discrimination may also be learned by attending to the overall similarity or family resemblance of the exemplars in each category (e.g., Kemler Nelson, 1984; Nicholls, Ryan, Bryant, & Lea, 2011). Indeed, some authors have suggested that nonprimate animals, like pigeons, lack the capacity to

selectively attend to relevant features, instead relying on the perceptual resemblance of stimulus configurations when performing a categorization task (e.g., Smith et al., 2012). Although Castro and Wasserman's (2014; (2016a) studies cast doubt on this conclusion, further empirical evidence might help clarify this issue.

Method

Subjects

The subjects were 4 homing pigeons (*Columba livia*) maintained at 85% of their free-feeding weights by controlled daily feedings. The birds had served in unrelated studies prior to the present project. All procedures were approved by the Institutional Animal Care and Use Committee at the University of Iowa.

Apparatus

The experiment used four $36 \times 36 \times 41$ cm operant conditioning chambers (Gibson, Wasserman, Frei, & Miller, 2004) 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). The viewable portion of the screen was $28.5 \text{ cm} \times 17.0 \text{ cm}$. 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 mounted in the center of the rear wall opposite the touchscreen. Illumination during experimental sessions was provided by a houselight on the upper rear wall of the chamber. The pellet dispenser and houselight were controlled by a digital I/O interface board. Each chamber was controlled by its own Apple[®] iMac[®] computer. Programs to run this and the next experiment were developed in MatLab[®] with Psychtoolbox-3 extensions (Brainard 1997; Pelli 1997; <http://psychtoolbox.org/>).

Stimuli

A total of 7 multicolored 3×3 cm squares (features) were used to create the different category training exemplars (Figure 1, left). Additional 7 multicolored 3×3 cm squares were used to create the different category testing exemplars (Figure 1, right). Each of the category exemplars was created by placing one different feature in each of the four corners of an invisible 12×12 cm square; each of the features was located 6 cm (both vertically and horizontally) from the two adjacent features, and connected by a white line. Each category exemplar contained two relevant features and two irrelevant features. Critically, one of the relevant features was a perfect predictor of the presented category (A or B); it predicted the correct category on 100% of the trials. The other relevant feature was an imperfect predictor of the presented category (A or B). It appeared along with the perfect predictor on 75% of the trials; on the remaining 25% of the trials, this imperfect predictor was presented along with the perfect predictor of the opposite category, thereby putting its predictiveness value at 50%. Irrelevant features were common to both Categories A and B; their predictiveness value was thus 0%. One of the irrelevant features varied from trial to trial (variable irrelevant), whereas the other irrelevant feature was present on every trial (fixed irrelevant).

Each of the relevant and irrelevant features appeared equally often in each of the four corner locations: top-left, top-right, bottom-left, or bottom-right. So, critically, spatial location could not be used as a cue for where the relevant features would be presented. All possible combinations of features and locations were used to create the exemplars in each category. Thus, the total number of unique training exemplars was 192 for Category A and 192 for Category B. Daily training sessions comprised 96 trials, so that all of the training exemplars were presented across four daily sessions.

To create the testing exemplars, we replaced each of the training features with new features, one at a time (see Figure 1, right, for a sampling of testing exemplars). On NewPerfPred trials, the features for the perfect predictors of each category were replaced with new features; these trials still contained the trained imperfect predictors. Even when the perfect predictor was absent, the birds could nevertheless respond at above-chance levels on the trials with the imperfect predictor if they had learned that this feature was more likely to appear on the exemplars of one category than on the exemplars of the other category. On the rest of the testing trials, the perfect predictor of each category was always present, so the discrimination should be easy to solve if the birds had learned to attend to them. On NewImpPred trials, the imperfect predictors of each category were replaced with new features; if the birds were also using the imperfect predictor to help them solve the discrimination, then its absence could result in some decrement in accuracy. On NewVarIrrel and NewFixedIrrel trials, the variable and fixed irrelevant features were replaced, respectively, by new features; as the perfect and imperfect predictors were the same as in training, no drop in accuracy might be expected for these trials.

For the testing stimuli, not all possible combinations of features and locations were used (due to the very large number of total combinations); but, as for the training exemplars, all of the different types of features appeared equally often in each of the four corner locations. A total of 128 unique testing stimuli were presented, distributed across eight testing sessions.

Procedure

Training—Daily training sessions comprised 96 trials; half presented Category A exemplars and half presented Category B exemplars, in randomized order. At the start of a trial, the pigeons were presented with an orienting 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 response requirement (this requirement gradually increased from 1 peck to 25 pecks over the course of 40 days, and stayed at 25 pecks thereafter) to any of the features—relevant or irrelevant—in the display. Only pecks within any of the four features' areas 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 buttons were 2.3×6 cm rectangles filled with distinctive black-and-white patterns. From trial to trial, the buttons were randomly located, to the left or right of the category exemplar, 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 an 8–12 s intertrial interval (ITI) ensued. If the choice response was incorrect, then food was not delivered, the houselight darkened, and a correction trial was given. Correction trials were given until the correct response was made. No data were analyzed from correction trials. Daily training sessions were given for a total of 120 days; then, we started the testing phase.

Testing—Each testing session comprised 96 training trials and 16 testing trials, randomly presented, for a total of 112 trials. A total of eight testing sessions were given in order to present all 128 testing stimuli; thus, each particular testing exemplar was presented only once. In each of the testing exemplars, either the perfect predictor, or the imperfect predictor, or the irrelevant features were replaced by novel features, as explained above.

On training trials, only the correct response was reinforced; incorrect responses were followed by correction trials (differential reinforcement). On testing trials, any choice response was reinforced (nondifferential reinforcement); food was given regardless of the pigeons' choice responses, so that testing could proceed without teaching the birds the correct responses to the testing exemplars. No correction trials were given on testing trials. Testing sessions and training sessions were given on alternate days to reduce any carryover effects of prior testing.

Data analysis

We calculated the birds' percentage of pecks to each of the relevant and irrelevant features over the total number of daily pecks. To have a parallel measure of accuracy, we calculated the birds' percentage of correct choice responses over the total number of daily choice responses. In addition, we grouped the 120 training days into 12 blocks of 10 days each. Also, because the percentage of correct responses across training typically violate the assumption of homoscedasticity—that is, variance tends not to be constant across training sessions, but is normally smaller at the beginning and end compared to the middle of training—these data were empirically logit-transformed for the purposes of analysis. To enhance clarity and understanding, percent scores are reported in the text and the figures as well.

The data were subjected to linear mixed effects (i.e., multilevel or random coefficients) analysis, using restricted maximum likelihood (REML). This type of analysis is especially well suited to analyzing training data, because it appropriately treats training session or training block as a continuous rather than as a categorical variable. Mixed-effects models extend the standard linear framework by adding random effects (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 a random slope. When we had sufficient evidence that a within-subject variable had effects that did indeed differ, it was included in the model to ensure the best estimates of the fixed effects. In the analyses, main effects and interactions were evaluated with *F* tests (type III Wald *F* tests with degrees of freedom calculated using the Kenward-Roger first-order approximation). Pairwise

comparisons were analyzed using Tukey HSD with the Tukey-Kramer adjustment. For the comparisons of interest, we report Cohen's d as the measure of standardized effect size. In order to compute the 95% confidence interval for d (CI_d), we followed procedures described by Smithson (2003) (<https://dl.dropboxusercontent.com/u/1857674/CIstuff/CI.html>). For comparisons against chance level, we report the 95% CI for the mean estimate (CI_{mean}). Finally, for differences of interest that were not statistically significant, we report the Bayes Factor (BF) in favor of the null hypothesis.

Results

Training

Accuracy—The birds' choice accuracy started near chance level at 53% on the first block of training, rose to over 75% by the sixth block, and reached 81% on the last block. However, as illustrated in Figure 2, accuracy differed depending on whether the exemplars contained the imperfect predictor that appeared on 75% of the trials or on 25% of the trials of a given category. So, although the perfect predictor was always present and was by itself sufficient to successfully solve the task, it seems that either the presence of the imperfect predictor of the same category facilitated the birds' choice of the correct response or the presence of the imperfect predictor of the opposite category impaired the birds' choice of the correct response.

We analyzed the birds' choice accuracy in training with a linear mixed-effects model, in which type of trial [those that included the perfect predictor and the imperfect predictor of the same category (Perf+ImpS) vs. those that included the perfect predictor and the imperfect predictor of the opposite category (Perf+ImpO)] and training block (1–12) were the fixed effects. The random effects structure included bird intercept and random slope for training block. The analysis yielded a main effect of block, $F(1,3) = 12.08$, $p = .04$, confirming that accuracy increased throughout training. There was also a main effect of type of trial, $F(1,950) = 449.50$, $p < .001$, due to overall accuracy being higher on trials that contained the imperfect predictor of the same category ($M = 77\%$) than on trials that contained the imperfect predictor of the opposite category ($M = 65\%$), as well as an interaction between training block and type of trial, $F(1,950) = 87.02$, $p < .001$, due to accuracy rising faster and to a higher level on trials that contained the imperfect predictor of the same category ($M = 88\%$ in the last training block) than on trials that contained the imperfect predictor of the opposite category ($M = 73\%$ in the last training block).

Feature tracking—Next, we calculated the birds' percentage of daily pecks to each of the four different features. A score of 25% would mean that the birds were indiscriminately pecking the features in the display, regardless of their predictive value. Clearly, this was not the case. Figure 3 shows that, as categorization accuracy increased, pecks to the perfect predictor rose as well, whereas pecks to the rest of the features stayed near or dropped below chance level.

We analyzed the birds' percentage of pecks with a linear mixed-effects model, in which type of trial (Perf+ImpS vs. Perf+ImpO), training block (1–12), and feature pecked (perfect predictor vs. imperfect predictor vs. variable irrelevant vs. fixed irrelevant) were the fixed

effects. The random effects structure included bird intercept. This analysis yielded a main effect of exemplar feature, $F(3,3744) = 994.44$, $p < .001$. Tukey HSD comparisons disclosed that, overall, more pecks ($M = 43\%$) were directed to the perfect predictor than to the imperfect predictor ($M = 22\%$), or to the fixed irrelevant feature ($M = 20\%$). The percentage of pecks to the variable irrelevant feature ($M = 15\%$) was significantly lower than to any of the other features.

The main effect of trial type was not significant, $F < 1$, $BF = 8.85$. So, whether the imperfect predictor of the same category was present or the imperfect predictor of the opposite category was present had no reliable effect on where pecks were directed; pigeons predominately pecked the perfect predictor regardless of the predictive value of the imperfect predictor.

There was no main effect of training block, but the Block X Feature Pecked interaction was significant, $F(3,3744) = 117.57$, $p < .001$, confirming a change in the percentage of pecks to the exemplar features throughout training. Follow-up analyses confirmed that pecks to the perfect predictor increased from the first ($M = 28\%$) to the last block of training ($M = 48\%$), $t(158) = 7.51$, $p < .001$, $d = 1.19$, 95% CI_d [0.85, 1.52]. We compared the percentage of pecks to the perfect predictor in the last training block against a hypothesized mean of 25% (chance level). This difference was significant on both type of trials: 95% CI_{mean} [42.1, 55.7], $t(3) = 7.05$, $p < .001$, and 95% CI_{mean} [39.5, 54.2], $t(3) = 6.02$, $p < .001$, on trials that contained the imperfect predictor of the same and the opposite category, respectively. So, although not required to do so, the birds learned to predominately peck the most predictive feature in the category exemplars.

Pecks to all of the other features decreased from the first to the last block of training: from 25% to 21%, $t(158) = -2.75$, $p < .01$, $d = 0.43$, 95% CI_d [0.12, 0.75], for the imperfect predictor; from 23% to 18%, $t(158) = -2.72$, $p < .01$, $d = 0.43$, 95% CI_d [0.11, 0.74], for the fixed irrelevant feature; and from 24% to 12%, $t(158) = -8.10$, $p < .001$, $d = 1.28$, 95% CI_d [0.94, 1.61], for the variable irrelevant feature.

In addition, we analyzed the birds' percentage of correct responses depending on which of the four features of the category exemplars the birds actually pecked (Figure 3). When the birds pecked the perfect predictor, accuracy was extremely high on both Perf+ImpS and Perf+ImpO trials (over 95% on all training blocks after Block 7). When the birds pecked the imperfect predictor on Perf+ImpS trials (Figure 3, left), accuracy was also high (over 80% on all training blocks after Block 7). However, when the birds pecked the imperfect predictor on Perf+ImpO trials (Figure 3, right), accuracy fell below chance (37% across blocks); this fall follows from the fact that the imperfect predictor on Perf+ImpO trials was the imperfect predictor for the opposite category.

To confirm and further explore these observations, we analyzed the birds' percentage of correct responses depending on which feature the birds had pecked with a linear mixed-effects model, separately for Perf+ImpS and Perf+ImpO trials. Feature pecked (perfect predictor vs. imperfect predictor vs. variable irrelevant vs. fixed irrelevant) and training

block (1–12) were the fixed effects; the random effects structure included bird intercept and slope for training block.

On Perf+ImpS trials, the analysis yielded a main effect of block, $F(1,1854) = 538.23, p < .001$, due to an overall increase in accuracy. There was also a main effect of feature pecked, $F(1,1854.1) = 284.21, p < .001$, confirming that accuracy differed depending on the feature that the birds had pecked, as well as an interaction between training block and feature pecked, $F(1,1854.1) = 18.27, p < .001$. Tukey HSD comparisons showed that overall accuracy after pecking the perfect predictor was higher ($M = 89\%$) than after pecking the imperfect predictor ($M = 78\%$), or after pecking the irrelevant features ($M = 61\%$).

On Perf+ImpO trials, the analysis yielded a main effect of block, $F(1,1789.3) = 62.11, p < .001$, due to an overall increase in accuracy. There was also a main effect of feature pecked, $F(1,1789.6) = 314.34, p < .001$, confirming that accuracy differed depending on the feature the birds had pecked, as well as an interaction between training block and feature pecked, $F(1,1789.3) = 22.80, p < .001$. Tukey HSD comparisons showed that overall accuracy after pecking the perfect predictor was higher ($M = 86\%$) than after pecking the imperfect predictor ($M = 37\%$), or after pecking the irrelevant features ($M = 54\%$).

In summary, both the percentage of correct responses and the percentage of pecks to the perfect predictor progressively increased over training; conversely, the percentage of pecks to the irrelevant cues progressively decreased over training. Accuracy was higher on trials in which the imperfect predictor of the same category was present than on trials in which the imperfect predictor of the opposite category was present; this higher accuracy was not due to increased tracking of the imperfect predictor on trials in which it predicted the same category as the perfect predictor, most likely because, given that the perfect predictor was always present, there was no incentive to track other features. Still, it is clear that the birds did learn the predictive value of the imperfect predictors, despite those features having been trained in the presence of perfectly predictive features, as shown by the higher accuracy on trials with the imperfect predictor of the same category than on trials with the imperfect predictor of the opposite category.

Testing

Accuracy—The percentage of correct responses in testing scores are displayed in Figure 4. Accuracy for training trials remained as high as at the end of training: 88% for Perf+ImpS trials and 71% for Perf+ImpO trials. On testing trials, however, accuracy varied greatly depending on which of the features of the category exemplars were replaced by a novel feature.

Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality, so we used the percentage of correct responses and the percentage of pecks to the stimulus features to analyze the birds' performance in testing. A linear mixed-effects model with type of trial as a fixed effect and bird intercept as a random effect yielded a main effect of type of trial, $F(8,264) = 295.44, p < .001$, confirming overall differences in accuracy depending on the type of testing trial; that is, pigeons' choice accuracy depended on which features were replaced by novel features in the testing

exemplars. To further analyze these differences, we performed strategic comparisons between testing and training trials during testing, as well as comparisons against the chance accuracy level of 50%.

When the perfect predictor was replaced by a new feature and the imperfect predictor of the same category was present (NewPerfPred+ImpS), accuracy dropped compared to training, but still remained above chance ($M = 61\%$), 95% $CI_{mean} [49.6, 73.2]$, $t(27) = 1.78$, $p = .04$; but, when the imperfect predictor of the opposite category was present (NewPerfPred+ImpO), accuracy dropped below chance ($M = 29\%$), 95% $CI_{mean} [16.14, 43.23]$, $t(27) = -3.06$, $p < .01$. These results were to be expected if the birds had learned that each of the imperfect predictors was predominately, albeit not perfectly, associated with one category. So, although the birds had not directed their pecks to the imperfect predictor more than to any of the irrelevant features during training, they had learned about its predictive value.

On all of the other testing trials, the perfect predictor was present; so, the birds ought to have been able to exhibit high accuracy levels. When the imperfect predictor was replaced by a new feature (NewImpPred), accuracy was high ($M = 74\%$); this accuracy score was lower than on Perf+ImpS training trials, $t(27) = 1.89$, $p = .06$, $d = .33$, 95% $CI_d [-0.02, .69]$, but similar to accuracy on Perf+ImpO training trials, $t(27) = 0.36$, $p = .71$, $d = .06$, 95% $CI_d [-.28, .41]$. So, the absence of the imperfect predictor had a negative effect on accuracy, again suggesting that the birds had learned the predictive value of the imperfect predictor.

When the variable and fixed irrelevant features were replaced by novel stimuli and both the perfect predictor and the imperfect predictor of the same category were present, accuracy was very high ($M = 86\%$ in both cases) and not significantly different from accuracy on Perf+ImpS training trials, $t(27) = 0.34$, $p = .73$, $d = .06$, 95% $CI_d [-.29, .41]$, $BF = 4.73$, and $t(27) = 0.17$, $p = .86$, $d = .03$, 95% $CI_d [-.32, .37]$, $BF = 4.92$, for NewVarIrrel+ImpS and NewFixedIrrel+ImpS trials, respectively. However, accuracy was much lower when the variable and fixed irrelevant features were replaced by novel stimuli and the imperfect predictor of the opposite category was present ($M = 53\%$ and $M = 65\%$, respectively). The drop in accuracy compared to Perf+ImpO training trials was significant on NewVarIrrel+ImpO trials, $t(27) = 2.52$, $p = .01$, $d = .45$, 95% $CI_d [.08, .81]$, but not on NewFixedIrrel+ImpO trials, $t(27) = 0.68$, $p = .49$, $d = .12$, 95% $CI_d [-.23, .47]$. This pattern of performance is somewhat peculiar. Note that the perfect predictor was always present on these trials; only irrelevant features were replaced by novel features. Still, accuracy did suffer on some of these trials (especially when the imperfect predictor of the opposite category was present on NewVarIrrel trials). As we shall see later, the birds' pattern of peck tracking sheds some light on this result.

Feature tracking—Next, we calculated the birds' percentage of pecks to each of the different features on each of the different types of trials in testing (Figure 5). On training trials, the distribution of pecks was the same as at the end of training: most of the pecks on both Perf+ImpS and Perf+ImpO trials were directed to the perfect predictor ($M = 48\%$ and $M = 47\%$, respectively); on both types of trials, pecks to the perfect predictor were above the 25% chance level, 95% $CI_{mean} [40.6, 55.9]$, $t(31) = 6.20$, $p < .01$, and 95% $CI_{mean} [39.5,$

56.1], $t(31) = 5.60$, $p < .01$, for Perf+ImpS and Perf+ImpO trials, respectively. Pecks to the remaining features, including the imperfect predictors, were at or below chance level.

Curiously, when the perfect predictor was replaced by a new feature, but the imperfect predictor of the same category was present (NewPerfPred+ImpS trials), pecks were not directed to the next best feature, the imperfect predictor, but predominately to the novel feature ($M = 38\%$), and clearly above chance level, 95% CI_{mean} [28.5, 55.8], $t(31) = 2.58$, $p = .01$. A similar performance pattern was observed when the perfect predictor was replaced by a new feature, but the imperfect predictor of the opposite category was present (NewPerfPred+ImpO trials); most of the pecks were directed to the novel feature as well ($M = 46\%$), above chance level, 95% CI_{mean} [29.5, 58.4], $t(31) = 2.67$, $p = .01$.

When the imperfect predictor was replaced by a new feature (PerfPred+NewImpPred trials), pecks to the perfect predictor were above chance ($M = 39\%$), 95% CI_{mean} [28.4, 51.1], $t(31) = 2.65$, $p = .01$, but so were pecks to the new feature ($M = 35\%$), 95% CI_{mean} [26.3, 45.8], $t(31) = 2.30$, $p = .02$. This pattern of performance was repeated for the testing trials in which the irrelevant features were replaced (Figure 5, bottom panel); pecks to the perfect predictor were above chance overall ($M = 34\%$), 95% CI_{mean} [28.1, 40.9], $t(118) = 2.26$, $p = .003$, but so too were pecks to the new features ($M = 35\%$), 95% CI_{mean} [26.8, 39.3], $t(118) = 2.56$, $p = .01$. Thus, the pigeons were strongly attracted to the new features of the testing exemplars, even when these features had no prior association with any of the report responses and, therefore, could not help the birds choose the correct response.

To further explore the relationship between feature tracking and accuracy, we next looked at the pattern of correct responses depending on peck location, as we had in the training phase. Just as in training, accuracy was the highest when the perfect predictor was presented and the pigeons pecked it, regardless of whether the imperfect predictor of the same or the opposite category was present (overall, $M = 89\%$). When the perfect predictor was present, but pigeons pecked the *new features*, accuracy was still high when the imperfect predictor of the same category was present ($M = 81\%$); however, accuracy dropped to near chance level when the imperfect predictor of the opposite category was present ($M = 42\%$). Note that on these latter trials, the perfect predictor was *actually present*; yet, the pigeons pecked the novel feature rather than the perfect predictor. Finally, when the perfect predictor was absent and the pigeons pecked the new features, accuracy dropped to chance level when the imperfect predictor of the same category was present ($M = 52\%$), and fell below chance level when the imperfect predictor of the opposite category was present ($M = 22\%$). So, it seems that the drop in accuracy on testing trials was mostly due to the pigeons being attracted to the new features in the category exemplars.

Discussion

As in our two previous studies (Castro & Wasserman, 2014; 2016a), we found that, as categorization accuracy increased, pecks to the perfect predictor of each category increased as well (Figure 2); these results further strengthen the idea that, as pigeons learn to categorize the training stimuli, they also pay increasing attention to the relevant attributes of those stimuli (see also Dittrich et al., 2010; George & Pearce, 2012; Rehder & Hoffman,

2005). In addition, pigeons' categorization accuracy was much higher on trials in which they pecked the perfect predictor than on trials in which they pecked the irrelevant features of the category exemplars (Figure 3). Pigeons' categorization accuracy was also much higher after pecking the imperfect predictor of the same category as the perfect predictor than after pecking the irrelevant features (Figure 3); clearly, the birds learned the predictive value of the imperfect predictor as well, even when there was a perfect predictor available (see also Wills, Inkster, & Milton, 2015).

Nonetheless, the likelihood of pigeons being more accurate after pecking the imperfect predictor did not lead them to peck it appreciably more often than the irrelevant features (Figure 2). If this had been the case, then removing the perfect predictor while the imperfect predictor was still present should have occasioned an increase in responding to the imperfect predictor. Instead, pecks that could no longer be directed to the perfect predictor were predominately directed to the novel feature. Remarkably, even when a feature that was 50% predictive of the correct response was present on those testing trials, the birds pecked the novel stimulus. This tendency suggests that not only was the increased allocation of attention to the relevant features learned, but so too was the decreased allocation of attention to the irrelevant features; the reduction in pecking at the irrelevant stimuli during training also supports this conclusion. In other words, it seems that the birds learned both to attend to the relevant features and to ignore the irrelevant features. We will delve into this issue next.

Negative associations and learning to ignore

Roembke, Wasserman, and McMurray (2016) recently trained pigeons to associate each of 16 object images with one particular report image or 'pexigram.' On each trial, two pexigrams were presented: one correct (a pexigram that was always presented with a particular object) and one incorrect (a randomly selected pexigram). For a given target object, only 8 pexigrams were available as potential report options (the correct pexigram and 7 possible foils); the other 8 foil pexigrams were withheld for testing. Pigeons could be learning a positive association between each particular object and the correct pexigram as well as learning a negative association between each particular object and the 7 incorrect pexigrams presented during training. However, there would be no opportunity for the pigeons to learn a negative association between a specific object and those incorrect pexigrams that had not been presented as foils during training.

In testing, when Roembke et al. (2016) presented each of the objects along with the correct pexigram and, as the alternative pexigram, a foil that had never been presented with that object (but had been presented with other objects) during training, pigeons' accuracy decreased, even when the correct pexigram was still available. These results suggest that pigeons were learning not only a positive association between a specific object and the correct pexigram, but also negative (or inhibitory) associations between a specific object and the incorrect pexigrams.

A process similar to that reported by Roembke et al. (2016) could underlie pigeons' observing responses in our study. The birds may have learned that the irrelevant features did not allow them to predict the correct response. They may also have learned that the imperfect predictor was an unreliable guide to accurate categorization; indeed, on 25% of

the trials, it directed the bird to make the incorrect response. Thus, when presented with new features in testing, the pigeons tended to peck those new features that had not developed a negative categorization association. It is still odd that the birds would peck the new features rather than the imperfect predictor—particularly given pigeons' well-known aversion to novel stimuli (neophobia; e.g., Clement & Zentall, 2003)—or even that, in some cases, they would peck the new features as often as they pecked the perfect predictor. But, Roembke et al. also noted that, even when the correct pexigram was available, pigeons might peck a pexigram that had never been presented with that specific object. Hence, both positive and negative associations appear to participate in category learning. The notion that irrelevant stimuli may be actively processed and negatively valenced stands in striking contrast to their being passively ignored.

Selective attention

Several decades ago, Sutherland and Mackintosh (1971) proposed that selective attention plays an important part in animal discrimination learning. In their words, “the evidence compels acceptance of the idea that discrimination learning in animals involves two processes: learning to which aspects of the stimulus to attend and learning what responses to attach to the relevant aspects of the stimulus situation (p. 21).” Our pigeons appear to have learned both to selectively attend to the relevant features of the stimuli and to choose the appropriate categorical response.

In the same vein, recent authors in the visual attention field deem selective attention to be a necessary requirement for adaptive learning and behavior. Given the inherent limitation of processing resources, only part of the available information can control an organism's behavior; so available stimuli compete with one another for behavioral control. Selective attention is understood to represent the resolution of this competition in favor of the stimuli that are most relevant for the task at hand (e.g., Chelazzi, Della Libera, & Santandrea, 2011; Desimone & Duncan, 1995). Our method—directly recording the location of birds' observing responses during learning of the categorization task—and our results—preferentially pecking at the most relevant features—allow us to conclude that pigeons do engage selective attention.

Some readers might find our conclusions to be controversial. In the categorization literature, a dominant notion is that multiple cognitive and neural systems allow an organism to learn about different aspects of the environment. Thus, many categorization theorists distinguish between analytic (explicit, focused, rule-based) processing, by which features or dimensions are found to diagnose category membership, and non-analytic (implicit, incidental, similarity-based) processing, by which the overall similarity among members of a category determines category membership (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Sloutsky, 2010).

Recently, Smith et al. (2012) compared four species—humans, macaques, capuchin monkeys, and pigeons—in learning two categorization tasks involving sine-wave disks varying in spatial frequency and orientation. In the rule-based task, variation in only one dimension, frequency or orientation, carried the information required to solve the task; here, optimal performance requires the learner to discover this unidimensional rule. In the

information-integration task, attention to both frequency and orientation is required to solve the task; here, optimal performance requires the learner to integrate information from both dimensions in a holistic, non-analytic way. Humans and macaques showed better performance in the rule-based task than in the information-integration task; capuchin monkeys performed poorly, but they too showed a preference for the rule-based task. In contrast, pigeons learned both tasks equally fast. Smith et al. concluded that “pigeon minds appeared not to preferentially appreciate one-dimensional task solutions” and “showed no tendency to learn rules or to apply dimensional analysis (p. 2365).” According to these authors, pigeons learn to gradually associate specific responses to a global stimulus pattern, but not to specific relevant properties of the stimuli; therefore, they lack the capacity for rule formation and selective attention.

Our results do not support Smith et al.’s (2012) analysis of pigeons’ cognitive capacities (see also Lea & Wills, 2008, Lea et al., 2009, and Wills et al., 2009, for further results and discussion). The mere fact that pigeons learned rule-based and information-integration tasks at similar rates does not prove that pigeons are incapable of selective attention. Rather, one possible inference from Smith et al.’s study is that, at least for pigeons, there is no advantage, in the particular procedure with the particular stimuli given, in using a single dimension rule compared to using overall exemplar similarity. It is also possible, as Edmunds, Milton, and Wills (2015) have argued, that spatial frequency and orientation were not separable dimensions for the pigeons in Smith et al.’s experiment. In order to use a single dimension rule, separability of the dimensions is a prerequisite; yet, Smith et al. provided no evidence that spatial frequency and orientation were in fact separable for the pigeons. The features in our stimuli were placed rather far apart to properly record peck tracking; that spatial separation may well have enhanced feature separability and attention to the most relevant feature. Different stimuli and procedures must be explored before strong conclusions are made about the cognitive system of an organism.

It could be argued that, in our experiment, a better case for selection attention would have been made had all of the pigeons’ pecks been directed to the perfect predictor. But, selective attention mechanisms are never perfect nor do they allow an organism to exhibit flawless performance. In humans, interference from irrelevant information is frequently observed (e.g., Monsell, 2003; Rein & Markman, 2010). For example, in the classic Stroop task, people have difficulty verbalizing the color of a word (the relevant attribute) when the printed word (the irrelevant attribute) names a different color (MacLeod, 1991). Pigeons are also influenced by irrelevant information (for example the spatial location of the features), but they nevertheless exhibited clear tracking of the most relevant information.

From the multiple systems view of category learning (e.g., Ashby et al., 1998), selective attention in category learning likely requires the involvement of brain structures mediating what is called ‘executive function’: namely, the prefrontal cortex (PFC). On the other hand, similarity-based categorization can be accomplished with more primitive brain regions, such as the inferotemporal cortex and basal ganglia (e.g., Ashby et al., 1998). Pigeons, in particular, and birds, in general, do not have a brain structure homologous to the PFC; nevertheless, the avian nidopallium caudolaterale (NCL) and the mammalian PFC share several anatomical and physiological features (Güntürkün, 2012). Although there is little

research in executive control functions in avian species, recent evidence suggests that the NCL is involved in higher-order cognition and cognitive flexibility (e.g., Veit & Nieder, 2013)—the functional domains of the PFC. Thus, despite lacking a laminated forebrain, both crows (Veit & Nieder, 2013) and pigeons (Castro & Wasserman, 2016b) can exhibit clear behavioral hallmarks of executive function. Given our present results (see also Lea & Wills, 2008) and the emerging neuroanatomical evidence about brain structures supporting high-order cognitive function in avian species, suggesting that selective attention is well within the pigeons' abilities should no longer be deemed to be far-fetched.

References

- Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychological Review*. 1998; 105:442–481. [PubMed: 9697427]
- Brainard DH. The psychophysics toolbox. *Spatial Vision*. 1997; 10:433–436. [PubMed: 9176952]
- Castro L, Wasserman EA. Pigeons' tracking of relevant attributes in categorization learning. *Journal of Experimental Psychology: Animal Learning and Cognition*. 2014; 40:195–211. [PubMed: 24893218]
- Castro L, Wasserman EA. Attentional shifts in categorization learning: Perseveration but not learned irrelevance. *Behavioural Processes*. 2016a; 123:63–73. [PubMed: 26548717]
- Castro L, Wasserman EA. Executive control and task switching in pigeons. *Cognition*. 2016b; 146:121–135. [PubMed: 26407340]
- Chelazzi L, Della Libera C, Sani I, Santandrea E. Neural basis of visual selective attention. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2011; 2:392–407. [PubMed: 26302199]
- Clement TS, Zentall TR. Choice based on exclusion in pigeons. *Psychonomic Bulletin & Review*. 2003; 10:959–964. DOI: 10.3758/bf03196558 [PubMed: 15000545]
- Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*. 1995; 18:193–222. DOI: 10.1146/annurev.neuro.18.1.193
- Deubel H, Schneider WX. Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*. 1996; 36:1827–1837. DOI: 10.1016/0042-6989(95)00294-4 [PubMed: 8759451]
- Dittrich L, Rose J, Buschmann JU, Bourdonnais M, Güntürkün O. Peck tracking: a method for localizing critical features within complex pictures for pigeons. *Animal Cognition*. 2010; 13:133–143. [PubMed: 19557444]
- Edmunds CER, Milton F, Wills AJ. Feedback can be superior to observational training for both rule-based and information-integration category structures. *The Quarterly Journal of Experimental Psychology*. 2015; 68:1203–1222. DOI: 10.1080/17470218.2014.978875 [PubMed: 25397975]
- George DN, Pearce JM. A configural theory of attention and associative learning. *Learning & Behavior*. 2012; 40:241–254. [PubMed: 22926999]
- Gibson BM, Wasserman EA, Frei L, Miller K. Recent advances in operant conditioning technology: A versatile and affordable computerized touch screen system. *Behavior Research Methods, Instruments and Computers*. 2004; 36:355–362.
- Güntürkün O. The convergent evolution of neural substrates for cognition. *Psychological Research*. 2012; 76:212–219. DOI: 10.1007/s00426-011-0377-9 [PubMed: 21881941]
- Kemler Nelson DG. The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior*. 1984; 23:734–759.
- Kruschke JK. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*. 1992; 99:22–44. [PubMed: 1546117]
- Lea SEG, Wills AJ. Use of multiple dimensions in learned discriminations. *Comparative Cognition and Behavior Reviews*. 2008; 3:115–133.
- Lea SEG, Wills AJ, Leaver LA, Ryan CME, Bryant CML, Millar L. A Comparative Analysis of the Categorization of Multidimensional Stimuli: II. Strategic Information Search in Humans (*Homo*

- sapiens) but Not in Pigeons (*Columba livia*). *Journal of Comparative Psychology*. 2009; 123:406–420. DOI: 10.1037/a0016851 [PubMed: 19929109]
- Le Pelley ME, Mitchell CJ, Beesley T, George DN, Wills AJ. Attention and associative learning in humans: An integrative review. *Psychological Bulletin*. 2016; 142:1111–1140. <http://dx.doi.org/10.1037/bul0000064>. [PubMed: 27504933]
- Mackintosh NJ. A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*. 1975; 82:276–298. DOI: 10.1037/h0076778
- MacLeod CM. Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*. 1991; 109:163–203. DOI: 10.1037/0033-2909.109.2.163 [PubMed: 2034749]
- Monsell S. Task switching. *Trends in Cognitive Sciences*. 2003; 7:134–140. DOI: 10.1016/s1364-6613(03)00028-7 [PubMed: 12639695]
- Nicholls E, Ryan CME, Bryant CML, Lea SEG. Labeling and family resemblance in the discrimination of polymorphous categories by pigeons. *Animal Cognition*. 2011; 14:21–34. DOI: 10.1007/s10071-010-0339-4 [PubMed: 20652343]
- Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
- Rehder B, Hoffman AB. Eyetracking and selective attention in category learning. *Cognitive Psychology*. 2005; 51:1–41. DOI: 10.1016/j.cogpsych.2004.11.001 [PubMed: 16039934]
- Rein JR, Markman AB. Assessing the concreteness of relational representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010; 36:1452–1465. DOI: 10.1037/a0021040
- Roembke TC, Wasserman EA, McMurray B. Learning in rich networks involves both positive and negative associations. *Journal of Experimental Psychology: General*. 2016; 145:1062–1074. DOI: 10.1037/xge0000187 [PubMed: 27336324]
- Sloutsky VM. From perceptual categories to concepts: What develops? *Cognitive Science*. 2010; 34:1244–1286. [PubMed: 21116483]
- Smith JD, Berg ME, Cook RG, Murphy MS, Crossley MJ, Boomer J, Spiering B, Beran MJ, Church BA, Ashby FG, Grace RC. Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews*. 2012; 36:2355–2369. DOI: 10.1016/j.neubiorev.2012.09.003 [PubMed: 22981878]
- Smithson, M. Confidence intervals. Thousand Oaks, CA: SAGE Publications, Inc; 2003.
- Styles, E. The psychology of attention. 2nd. New York, NY: Psychology Press; 2006.
- Sutherland, NS., Mackintosh, NJ. Mechanisms of animal discrimination learning. New York: Academic Press; 1971.
- Veit L, Nieder A. Abstract rule neurons in the endbrain support intelligent behavior in corvid songbirds. *Nature Communications*. 2013; 4doi: 10.1038/ncomms3878
- Wasserman EA. Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*. 1974; 103:284–297.
- Wills AJ, Inkster AB, Milton F. Combination or Differentiation? Two theories of processing order in classification. *Cognitive Psychology*. 2015; 80:1–33. DOI: 10.1016/j.cogpsych.2015.04.002 [PubMed: 26057479]
- Wills AJ, Lea SEG, Leaver LA, Osthaus B, Ryan CME, Suret MB, Bryant CML, Chapman SJA, Millar L. 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*. 2009; 123:391–405. DOI: 10.1037/a0016216 [PubMed: 19929108]

Figure 1.

Representative Category A and Category B training (left) and testing (right) exemplars. Each training exemplar contained one feature, the perfect predictor, present in all exemplars (the rainbow for Category A and the green spiral for Category B), and one imperfect predictor that appeared along with the perfect predictor of a given category on 75% of the trials (the red bubbles for Category A and the Mondrian colored squares for Category B); on the other 25% of the trials, this imperfect predictor appeared along with the perfect predictor of the opposite category. In addition, there were two irrelevant features, common to Categories A and B: one fixed irrelevant feature present in all exemplars and two variable irrelevant features, each of them present in half of the exemplars. In testing exemplars, some training features were replaced. In the top row, exemplars in which the perfect predictor was replaced with a novel feature. In the second row, exemplars in which the imperfect predictor was

replaced with a novel feature. In the third row, exemplars in which the variable irrelevant feature was replaced with a novel feature. In the fourth row, exemplars in which the fixed irrelevant feature was replaced with a novel feature.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

Mean percentage of correct responses (thick continuous lines) and mean percentage of pecks to each of the four types of features on trials containing the imperfect predictor of the same category (left) and trials containing the imperfect predictor of the opposite category (right) across training blocks often daily sessions each. The dotted line, at 25%, represents the chance level for percentage of pecks, whereas the dashed line, at 50%, represents the chance level for percentage of correct responses. Error bars indicate the standard error of the means.

Figure 3.

Mean percentage of correct responses depending on the feature that the birds had pecked before making the choice response on trials containing the imperfect predictor of the same category (left) and on trials containing the imperfect predictor of the opposite category (right) across training blocks containing ten daily sessions each. The dashed line, at 50%, represents the chance level for percentage of correct responses. Error bars indicate the standard error of the means.

Figure 4.

Mean percentage of correct responses on the different type of testing trials (in different shades of grey; dark grey for trials in which the imperfect predictor of the same category was present, light grey for trials in which the imperfect predictor of the opposite category was present, and diagonal stripes for trials in which the imperfect predictor was replaced by a new feature), as well as on training trials during the testing phase (in black). The dashed line, at 50%, represents the chance level for percentage of correct responses. Error bars indicate the standard error of the means.

Figure 5.

Mean percentage of pecks to each of the different features on the different type of testing trials. Pecks on training trials during the testing phase (in black) are presented for comparison purposes in both the top and the bottom graphs. Pecks on testing trials are presented in grey, except for pecks at the new features that appear in a black-and-white diagonal striped pattern. The dashed line, at 25%, represents the chance level for percentage of pecks. Error bars indicate the standard error of the means.