

Pigeons' Tracking of Relevant Attributes in Categorization Learning

Leyre Castro and Edward A. Wasserman
The University of Iowa

Most theories and experimental investigations of discrimination learning and categorization, in both humans and animals, hypothesize that attention must be allocated to the relevant attributes of the training stimuli for learning to occur. Attention has conventionally been inferred after learning has transpired rather than examined while learning is transpiring. We presented pigeons with a visual categorization task in which we monitored their choice accuracy through their responses to different report buttons; critically, we tracked the location of the pigeons' pecks to both the relevant and irrelevant attributes of the training stimuli using touchscreen technology, in order to find out where the birds may have been attending during the course of categorization learning. Pigeons readily mastered the categorization task; most importantly, as training progressed, they increasingly concentrated their pecks on the relevant features of the category exemplars, suggesting that the birds were tracking the relevant information to solve the task. When either new irrelevant features were introduced (Experiment 1) or when new relevant features were introduced and later the discriminative value of these new relevant features was reversed (Experiment 2), pigeons' choice accuracy and peck tracking were strongly affected. These results help elucidate the dynamics and interplay of attention and learning; they also suggest that peck tracking can be a suitable measure of the allocation of attention in pigeons, much as eyetracking is deemed to be a suitable measure of attention in humans.

Keywords: categorization, attention, peck tracking, animal cognition, pigeons

Animals such as rats and pigeons perform well in a variety of discrimination and categorization tasks. Rats learn to choose between a black and a white arm in a maze when the black and white arms are randomly located to the left and right (e.g., Lawrence, 1949). Furthermore, despite their less than stellar visual acuity, rats have recently proven to be capable of categorizing complex stimuli such as chairs versus flowers (Brooks et al., 2013). Pigeons, visually more adept than rats, can discriminate between color slides that contain people and color slides in which people are absent (Herrnstein & Loveland, 1964), between animals and non-animals (Roberts & Mazmanian, 1988), or between paintings by Monet and Picasso (Watanabe, Sakamoto, & Wakita, 1995). Moreover, pigeons not only learn such binary discriminations, but they also learn to categorize multiple complex stimuli within the same task. For example, Bhatt, Wasserman, Reynolds, and Knauss (1988) showed pigeons color slides depicting images from four different categories: cats, flowers, cars, and chairs. Only one image was presented at a time, and pigeons had to peck one out of four report buttons, depending on the category displayed. The birds rapidly reached high accuracy levels and responded far above chance to new exemplars from the four trained categories.

As Keller and Schoenfeld (1950) observed, categorization involves “generalization *within* classes and discrimination *between* classes (p. 155).” When an organism makes the same response

“car” to discriminably different cars, it is generalizing among all cars; when an organism makes the same response “flower” to discriminably different flowers, it is generalizing among all flowers; at the same time, it is discriminating flowers from cars. In order to master this categorization task, animals have to perceive and attend to those features that are common among cars and that distinguish cars from flowers (e.g., Soto & Wasserman, 2010). Thus, it seems intuitive, and almost obvious, to say that when animals learn to solve a discrimination or to categorize different images, they will have to focus on those features of the stimuli that are relevant to correctly perform the task and to disregard those features that are not relevant to task mastery.

Lawrence (1949) famously set forth the idea that those stimuli that are relevant to the solution of a discrimination problem will be paid more attention than those that are irrelevant. Several subsequent theorists expanded on that idea; for example, Mackintosh (1965, 1975) proposed that, during discrimination learning, relevant stimuli will become more accurate and reliable predictors of the outcome of a trial than those stimuli that are irrelevant. Consequently, attention to the relevant stimuli should increase, whereas attention to the irrelevant stimuli should decrease. According to this proposal, organisms gradually learn to attend to the characteristics along which the discriminative stimuli differ, with attention increasing because of the ability of a stimulus to predict the occurrence or nonoccurrence of an outcome.

Models of human categorization, many of them inspired by animal learning theories, have also included mechanisms by which attention to the relevant and irrelevant features of the stimuli changes during the course of learning. For example, according to Kruschke's (1992) ALCOVE (attention learning covering map) connectionist model, attention is equally allocated to all stimulus features at the beginning of learning; however, as training pro-

Leyre Castro and Edward A. Wasserman, Department of Psychology, The University of Iowa.

We thank Yuejia Teng for her assistance in conducting this project.

Correspondence concerning this article should be addressed to Leyre Castro, Department of Psychology, E11 Seashore Hall, The University of Iowa, IA City, IA 52242. E-mail: leyre-castroruiz@uiowa.edu

ceeds, attention gradually shifts to the relevant features as a result of error feedback.

William James (1890/1950) said that, "Everyone knows what attention is (. . .) It implies withdrawal from some things in order to deal effectively with others" (p. 403). Nowadays, attention is typically defined as selectively focusing on one aspect of the environment while ignoring others (Anderson, 2005). But, how can we empirically ascertain whether an organism is concentrating on one or another aspect of the environment? Attention cannot be measured directly, but must be inferred from an organism's behavior. Because there seems to be a close connection between the direction of eye gaze and changes in attention, and because there is empirical evidence suggesting that eye movements and attention operate simultaneously (e.g., Deubel & Schneider, 1996), eye movement has been used in human cognition studies as an overt measure of attention, although it is also known that attention and eye movements do not necessarily go together (e.g., Posner, 1980). Examining eye gaze in birds is an alluring option and the technology is currently being developed (see Yorzinski, Patricelli, Babcock, Pearson, & Platt, 2013, for usage of an eye-tracking device in peahens), but its way into the realm of experimental cognition research still has to be worked out. How then might we measure overt attention in animals?

Traditional studies in animal discrimination learning infer different levels of attention by examining the rate of learning of some events after other events have previously been presented. The rationale is that the future rate of learning should be influenced by the amount of attention previously deployed. If a cue has previously been highly attended, then learning of a new discrimination in which that cue is relevant should proceed more rapidly than if the cue had not been attended; but if the cue has not previously been attended, then learning of a new discrimination in which that cue is relevant should proceed more slowly than if the cue had been attended.

A classic design to study the role of attention in discrimination learning was advanced by Mackintosh and Little (1969). They initially trained two groups of pigeons on a visual discrimination containing compound stimuli of different colors and line orientations. One of the dimensions, say color, was relevant during the first learning phase, but the other dimension, line orientation, was not. Once the pigeons mastered that discrimination, a novel discrimination was presented in a second phase. For one group of pigeons, the dimension that was relevant in the first phase remained relevant in the second phase (intradimensional shift), whereas for the second group of pigeons, the dimension that was irrelevant in the first phase became relevant in the second phase (extradimensional shift). The intradimensional shift task was more rapidly learned than the extradimensional shift task; this disparity in the speed of learning is commonly interpreted as the result of the birds' increased attention to the relevant dimension during initial training. Because attention is more strongly allocated to the relevant dimension, future learning involving that dimension is facilitated (see, e.g., George & Pearce, 1999; Kruschke, 1992; Roberts, Robbins, & Everitt, 1988, for similar procedures and results).

In our current project, rather than inferring prior attention from later performance, we adopted the novel tactic of measuring the emergence and allocation of selective attention while learning was actually taking place, in order to shed fresh light on the dynamics and interplay of attention and learning. Recently, eyetracking

technology has been applied to study attention in human category learning. Rehder and Hoffman (2005) tracked participants' eye movements while they were solving categorization tasks in which some elements were relevant to correctly classifying the category exemplars, whereas other elements were not. They found that, as learning progressed, participants' allocation of attention gradually shifted toward the relevant features of the stimuli. Curiously, but possibly importantly, shifts in eye movements tended to follow rather than to lead changes in categorization accuracy.

A possible analog to eye gaze direction in humans may be peck location in birds. Prior to the exploration of peck location, several studies have looked at birds' peck rate under the assumption that, when a stimulus elicits a high rate of responding, more time and more attention is also allocated to it (e.g., Pearce, Esber, George, & Haselgrove, 2008). Most of these studies examining peck rate have not recorded the specific location where the birds were pecking, although there are some exceptions. For example, in Wasserman (1974), pigeons were presented 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 separately monitored during the acquisition of the discrimination. Of interest for our current discussion is that, when A and X were presented together, the rate of pecking to relevant stimulus A and to the common irrelevant stimulus X, similar at the beginning of training, quickly diverged. By the end of training, pigeons were directing all of their pecks to relevant stimulus A (for related results see Allan, 1993; Bermejo & Zeigler, 1998; Jenkins & Sainsbury, 1970).

More recently, touchscreen technology has been used to track the location of pigeons' pecks within complex images. Dittrich, Rose, Buschmann, Bourdonnais, and Güntürkün (2010) tracked pigeons' peck locations while the birds were learning a go-no go task which required discriminating between color pictures in which human figures were present (always followed by food reinforcement) and color pictures in which human figures were absent (never followed by food reinforcement). As pigeons' discriminative performance improved, they increasingly focused their pecks on areas occupied by human figures, the distinctively diagnostic features of the pictorial stimuli.

All of these prior studies monitoring peck rate and location strongly suggest that peck tracking might prove to be a powerful means of measuring pigeons' allocation of visual attention, in the same way that eyetracking is considered to be an effective, albeit imperfect, measure of human visual attention. Of course, we acknowledge that peck location should be cautiously interpreted as a measure of attention. But, note that the birds must somehow attend to the relevant dimensions of the stimuli to be able to successfully solve the categorization task. If peck tracking of the relevant dimensions gradually increases as categorization performance improves, then we will have converging evidence (peck location and performance accuracy) of the stimulus aspects being processed.

Furthermore, it should be noted that the allocation of attention may also depend on the complexity of the visual stimuli that are presented. Some research has found that category structure—that

is, the degree of variability and coherence within a category—can determine the ease of category learning in human children and adults: the less variability within a category, the easier the learning (e.g., Kloos & Sloutsky, 2008). Thus, categories containing a large number of irrelevant features may be more difficult to learn than categories containing a small number of irrelevant features. On the other hand, research in language and speech categorization has shown that variability in irrelevant information (e.g., in speaker voices) may actually help people better extract relevant information (e.g., particular phonemes) and, thereby promote generalization to new classification tasks (Rost & McMurray, 2009; see also, Apfelbaum, Hazeltine, & McMurray, 2013; Gómez, 2002).

In addition, studies of visual categorization with human adults (Homa, Cross, Cornell, Goldman, & Schwartz, 1973), human infants (Hayne, 1996), and pigeons (Wasserman & Bhatt, 1992) have shown that the number of exemplars in a category during the learning phase strongly affects later transfer performance; the more exemplars presented during the initial learning phase, the better the classification of novel exemplars during the later transfer phase. Presumably, a large number of individual instances enhances categorical knowledge by increasing the salience of the relevant features of the category and/or by reducing the salience of features that are irrelevant for the categorization task.

Thus, it could be that, on the one hand, variability in the irrelevant features of a category slows learning (e.g., Wasserman & Bhatt, 1992); when more exemplars are presented, the organism is exposed to a larger number of features, and the process of learning which features are relevant and which features are irrelevant can be retarded. In relation to attention, it may take the organism longer to identify the relevant features when there are more rather than fewer irrelevant features.

On the other hand, variability in the irrelevant features may improve later generalization to novel exemplars of the learned categories by increasing the salience of the relevant features of the categories. Even if learning to focus attention on the relevant features is slower, once this learning has taken place, it might be more resistant to disruption generated by the introduction of other irrelevant features. Thus, when novel exemplars are presented, generalization may be improved after training with a large number of irrelevant features.

In our present project, two groups of pigeons had to learn to categorize two different artificial categories; each of the categories was associated with one particular response button. The categories contained two relevant features and two irrelevant features (Group Fixed Irrelevant) or two relevant features and eight irrelevant features (Group Variable Irrelevant); so, the groups differed in the amount of irrelevant information contained within the categories. Category exemplars always consisted of four features: two relevant and two irrelevant (see Figure 1). Thus, for the Fixed Irrelevant group the irrelevant features were always presented on every trial, whereas for the Variable Irrelevant group they varied from trial to trial. Based on prior research, learning might be expected to be slower in Group Variable Irrelevant than in Group Fixed Irrelevant, but generalization to novel category exemplars might be expected to be superior in Group Variable Irrelevant than in Group Fixed Irrelevant.

Critically, when a category exemplar was presented on the screen, the pigeons had to peck it several times. Only the areas occupied by the relevant and irrelevant features were “active” for

monitoring pecks; pecks at the black background did not count. The pigeons did not have to peck at any specific feature of the category exemplar; pecks to the relevant and irrelevant features equally satisfied the observing response requirement. The pigeons did, however, have to find and process the relevant features (which were randomly located, from trial to trial, in two of four possible locations) in order to solve the category discrimination task. Because the pigeons were free to distribute their pecks to the two relevant and two irrelevant features of every category exemplar, the birds could adopt three possible strategies: (1) look at the category exemplar and then randomly peck at the features before making their final choice response, (2) randomly peck at the features and then look at the category exemplar before making their final choice response, or (3) look at the category exemplar and peck at one or both of the relevant features before making their final choice response. Only the third strategy would support peck tracking of the features that are relevant for the categorization task.

Experiment 1

Method

Subjects. The subjects were eight feral pigeons (*Columba livia*) maintained at 85% of their free-feeding weights by controlled daily feedings. The pigeons were divided into two groups: Fixed Irrelevant (two irrelevant features, present in all exemplars) and Variable Irrelevant (eight irrelevant features, varying from exemplar to exemplar). The pigeons had served in unrelated studies prior to the present project.

Apparatus. The experiment used four 36 × 36 × 41 cm operant conditioning chambers 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). The portion of the screen that was viewable by the pigeons was 28.5 cm × 17.0 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 located in the center of the rear wall opposite the touchscreen. Illumination during the experimental sessions was provided by a houselight mounted 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 14 multicolored 3 × 3-cm squares (features) were used to create the different category training exemplars. Each of the 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 6-cm apart (both vertically and horizontally) from the two adjacent features, and all were connected by a white line. Every category exemplar contained two relevant features and two irrelevant features. There were two relevant features for Category A and two different relevant features for Category B. Irrelevant features were common to Categories A and B. For the Fixed Irrelevant group, there were two irrelevant features, so both of them were present in all exemplars.

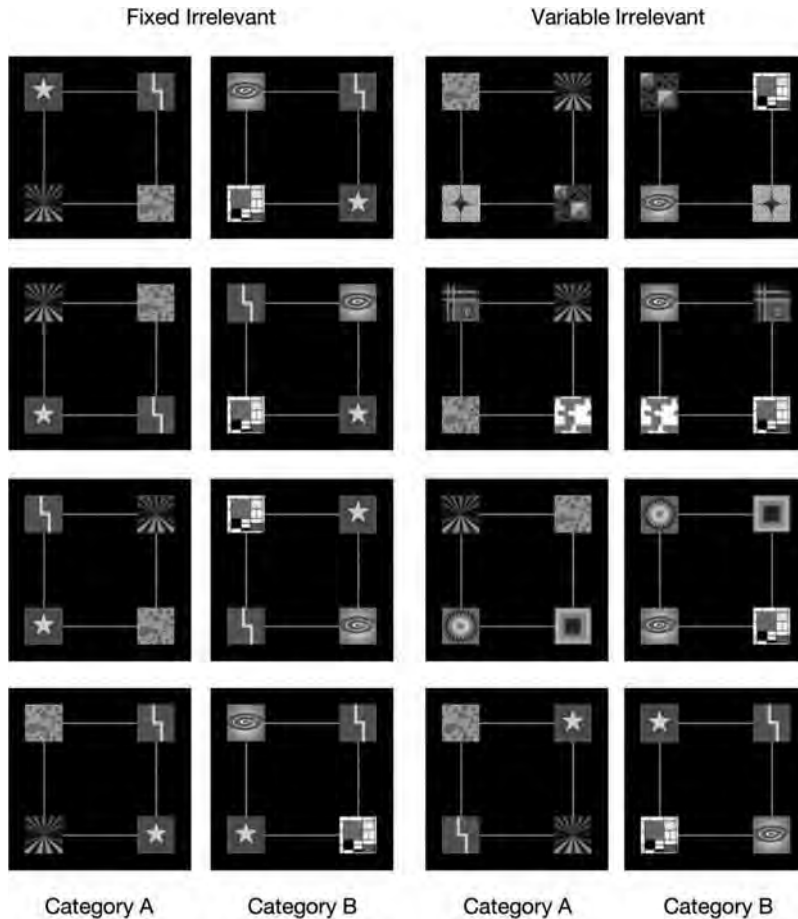


Figure 1. Examples of Category A and Category B training exemplars in Experiment 1. The two relevant features for Category A and the two relevant features for Category B are the same for group Fixed Irrelevant and group Variable Irrelevant. The irrelevant features are constant across exemplars for group Fixed Irrelevant (two irrelevant features), but vary from exemplar to exemplar for group Variable Irrelevant (eight irrelevant features).

For the Variable Irrelevant group, there was a total of eight irrelevant features, so they varied from exemplar to exemplar (in order to avoid an extremely large number of exemplars, the eight irrelevant features were divided into four pairs, so that the features in each pair always appeared together and never with any of the six other irrelevant features).

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 determining where the relevant features would be presented. All possible combinations of features and locations were used to create the exemplars in each category (see Figure 1 for a sampling of exemplars). Thus, the total number of unique training exemplars was 24 for Category A and 24 for Category B in Group Fixed Irrelevant. In Group Variable Irrelevant, because of the larger number of irrelevant features, the total number of unique training exemplars was 96 for Category A and 96 for Category B.

To create the testing exemplars, we replaced the irrelevant training features with new irrelevant features. Eight new irrelevant features were used to create a total of 64 testing exemplars, 32 per

category (see a sampling of these exemplars in Figure 2). Both Groups Fixed Irrelevant and Variable Irrelevant were presented with the same 64 testing exemplars. 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 with the training exemplars, all of the relevant and new irrelevant features appeared equally often in each of the four corner locations.

Procedure.

Training. Daily training sessions comprised 144 trials; half presented Category A exemplars and half presented Category B exemplars, in a random fashion. At the start 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 response requirement (from 5 to 15 pecks) to any of the features—relevant or irrelevant—in the display. This requirement was adjusted to the performance of each pigeon. If the bird was consistently pecking, but not meeting the discrimination criterion (see ahead) in a timely fashion, then the number of pecks was raised to increase the cost of making an incorrect response (at the end of the

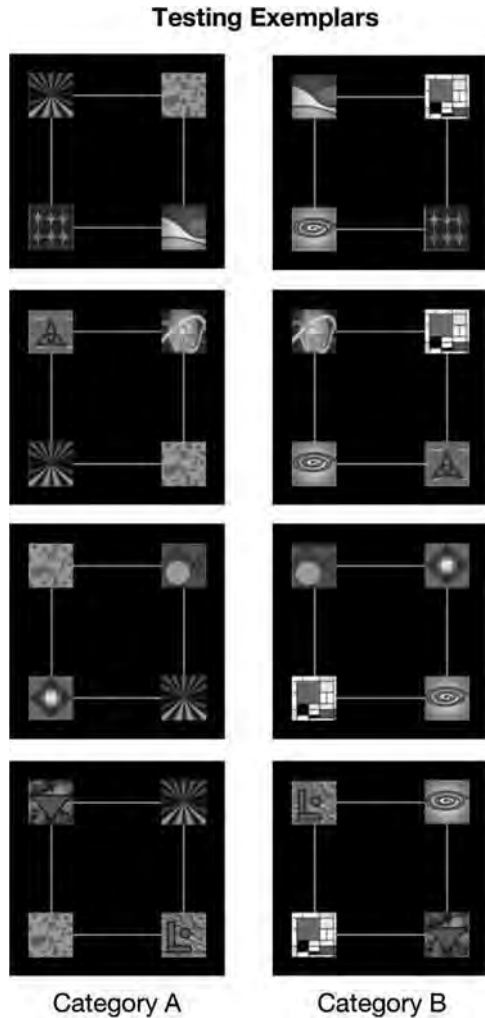


Figure 2. Examples of Category A and Category B testing exemplars in Experiment 1. The relevant features for both Category A and Category B are maintained, but new irrelevant features are introduced (eight irrelevant features in total). Testing exemplars were the same for groups Fixed Irrelevant and Variable Irrelevant.

initial training phase, the observing requirements for the four pigeons in the Fixed Irrelevant group were 5, 5, 8, and 10 pecks; the observing requirements for the four pigeons in the Variable Irrelevant group were 5, 5, 8, and 15 pecks). 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 prevent a 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 deliv-

ered and the intertrial interval (ITI) ensued; the ITI randomly ranged from 8 to 12 s. 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.

We trained the birds until they reached an accuracy level of 85% for each of the categories on 2 consecutive days, to ensure that categorization performance had reached a high level. Then, we started the testing phase.

Testing. Each testing session comprised 144 training trials and 16 testing trials, randomly presented, for a total of 160 trials. A total of four testing sessions were given in order to present all 64 testing stimuli; thus, each particular testing exemplar was presented only once. Testing exemplars were the same for Groups Fixed Irrelevant and Variable Irrelevant. The relevant features for Categories A and B were maintained, but eight novel irrelevant features were presented in the testing exemplars, 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 our teaching the birds the correct responses to the testing exemplars. The designations of correct or incorrect for choice responses (based on the testing exemplars belonging to Category A or B) on testing trials were for scoring purposes only. No correction trials were thus necessary on testing trials. Testing sessions and training sessions were given on alternate days, so that there was always one day of training between testing sessions.

Data analysis. We calculated the birds' percentage of pecks at the relevant features (*relevant pecks*) over the total number of daily pecks. In order to analyze relevant pecks as well as choice accuracy, we used analyses of variance (ANOVAs) in which an alpha level of .05 was adopted. Because of the factorial nature of all our ANOVAs, we report partial eta squared (η_p^2) as the measure of standardized effect size and we include 95% confidence intervals (CIs) for η_p^2 . In order to compute upper and lower boundaries of the CIs for η_p^2 , we obtained the noncentrality parameter of the noncentral *F* distribution, as well as CIs for the noncentrality parameter using methods implemented in the Methods for the Behavioral, Educational, and Social Sciences (MBESS; Kelley, 2007a,2007b) R statistical package (R Development Core Team, 2007) and followed procedures described by Smithson (2003).

Results

Training. Group Fixed Irrelevant took a mean of 7.00 ($SD = 0.81$), 11.75 ($SD = 3.77$), 17.75 ($SD = 7.93$), and 25.75 ($SD = 14.86$) days to reach categorization accuracy levels of 65%, 75%, 85%, and 85% on 2 consecutive days, respectively. Group Variable Irrelevant took a mean of 8.75 ($SD = 4.92$), 13.50 ($SD = 5.74$), 24.00 ($SD = 15.18$), and 34.50 ($SD = 22.78$) days to reach the same categorization accuracy levels of 65%, 75%, 85%, and 85% on 2 consecutive days, respectively. There was a numerical speed advantage for the Fixed Irrelevant group, but an ANOVA with group and criterion (65% vs. 75% vs. 85% vs. 85%_2) as factors yielded no significant differences between the groups; at all criterion levels, all $F_s < 1$.

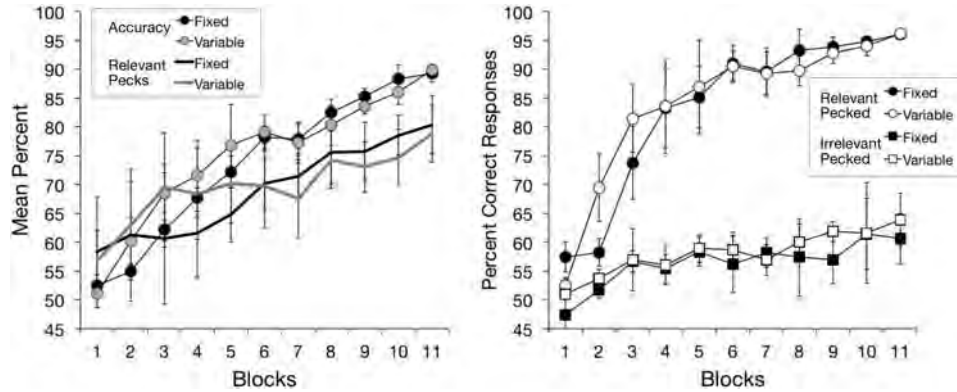


Figure 3. All daily training sessions until reaching 85% accuracy level for 2 consecutive days were grouped into 11 training blocks. Left: mean percent accuracy and relevant pecks across training for groups Fixed Irrelevant and Variable Irrelevant in Experiment 1. Right: accuracy scores across training depending on whether the relevant or the irrelevant features had been pecked just before making the choice response. Error bars indicate the standard error of the means.

In order to examine the relation between categorization accuracy and feature peck-tracking, we constructed Vincentized learning curves by grouping all of the training sessions until reaching the 85% criterion on 2 consecutive days into 11 blocks (11 was the smallest number of sessions that it took a bird to reach 85% accuracy on 2 consecutive days). Blocks were composed of one to six sessions, depending on the pigeon; any remaining sessions were placed into the middle blocks (Kling & Riggs, 1971). Mean accuracy on the last training block was 89% for the Fixed Irrelevant group and 90% for the Variable Irrelevant group. A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 11 (blocks) ANOVA yielded a main effect of training block, $F(10, 30) = 28.09$, $p < .001$, $\eta_p^2 = .90$, 95% CI [.77, .92], confirming that accuracy in both groups increased as training progressed. There was no main effect of group nor an interaction.

Next, we looked at the birds' percentage of pecks at the relevant features. A similar number of relevant and irrelevant pecks, approximately 50% each, would mean that the birds were indiscriminately pecking at the features in the display, regardless of their predictive value. But, this was not the case. As can be seen in Figure 3 (left), as categorization accuracy increased, relevant pecks progressively increased as well. By Block 5 in Group Variable Irrelevant and by Block 6 in Group Fixed Irrelevant, 70% of the pecks were directed at the relevant features of the exemplars; by the end of training, relevant pecks reached 79% and 80% in Groups Variable Irrelevant and Fixed Irrelevant, respectively. A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 11 (blocks) ANOVA yielded a main effect of training block, $F(10, 30) = 9.53$, $p < .001$, $\eta_p^2 = .76$, 95% CI [.46, .79], confirming that relevant pecks increased as training progressed in both groups. There was no main effect of group nor an interaction. In addition, we compared the percentage of relevant pecks in the last training block in each of the groups against a hypothesized mean of 50% (the case in which the birds would randomly peck at relevant and irrelevant cues). This difference was statistically significant in both groups: 95% CI [64.56, 96.19],¹ $t(3) = 6.11$, $p < .005$, in Group Fixed Irrelevant, and 95% CI [63.19, 94.31], $t(3) = 5.88$, $p < .005$, in Group Variable Irrelevant confirming that, indeed, relevant pecking was higher than expected by chance at the end of training.

Clearly, although not specifically required to do so, the birds learned to predominately peck at the relevant features of the category exemplars. So, both accuracy and relevant pecks gradually increased over training. But, why should the birds more frequently come to peck at the relevant than the irrelevant features of the images?

To shed light on this issue, we looked at birds' categorization accuracy on trials in which their pecks were directed either at the relevant or the irrelevant features of the category exemplars. As Figure 3 (right) shows, after the birds pecked at the relevant features of the category exemplars, their accuracy was very high; however, after the birds pecked at the irrelevant features of the category exemplars, their accuracy was much lower. A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 11 (blocks) \times 2 (type of feature pecked: relevant vs. irrelevant) ANOVA on choice accuracy revealed a significant main effect of the type of feature pecked, $F(1, 6) = 337.50$, $p < .001$, $\eta_p^2 = .98$, 95% CI [.86, .99], confirming that accuracy was lower when the birds pecked at the irrelevant features of the category exemplars (overall, 58%) than when they pecked at the relevant features of the category exemplars (overall, 81%). There was also a significant main effect of block, $F(10, 60) = 19.34$, $p < .001$, $\eta_p^2 = .76$, 95% CI [.60, .80], and a Block \times Type of feature interaction, $F(10, 60) = 5.28$, $p < .001$, $\eta_p^2 = .47$, 95% CI [.19, .54], disclosing that the disparity in accuracy due to the type of feature pecked widened as training progressed.

Testing. Accuracy and relevant pecks in testing are displayed in Figure 4 (top). Both Groups Fixed Irrelevant and Variable Irrelevant showed excellent transfer performance to the novel testing stimuli. In both groups, accuracy for novel testing stimuli was very high (84% and 86% for the Fixed Irrelevant group and the Variable Irrelevant group, respectively), albeit slightly lower than accuracy for the familiar training stimuli (91% for both the Fixed Irrelevant group and Variable Irrelevant group). The same

¹ Here and throughout, for the t -test comparisons against a hypothesized mean of 50%, the 95% CI refers to the 95% confidence interval of the mean of interest.

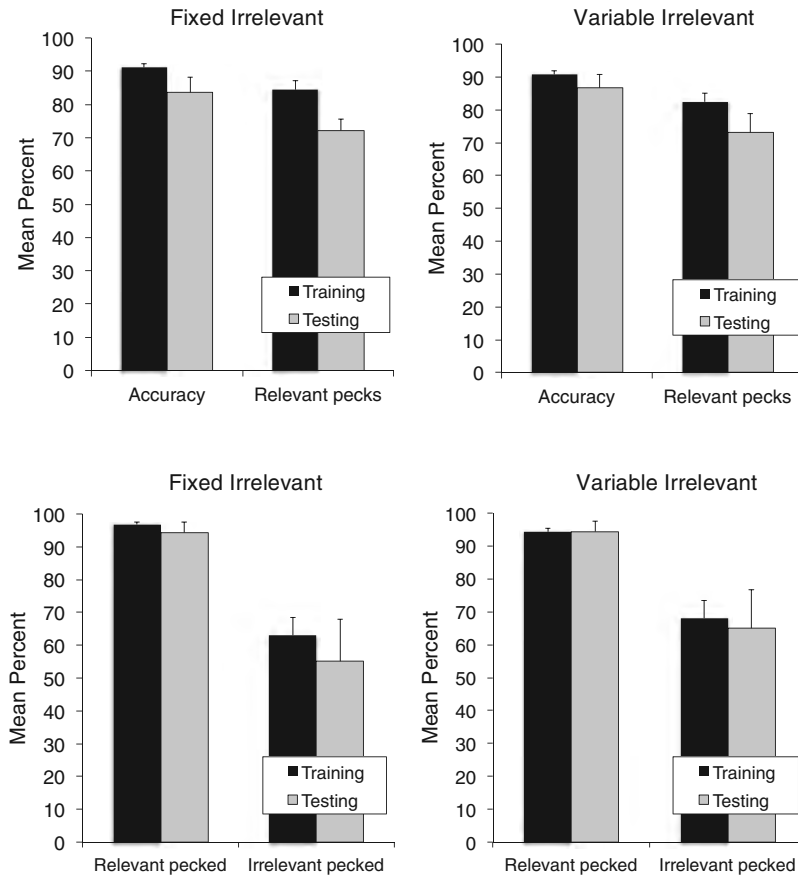


Figure 4. Top: mean percent accuracy and relevant pecks in testing in Experiment 1 for groups Fixed Irrelevant (left) and Variable Irrelevant (right). Bottom: testing accuracy scores depending on whether the relevant or the irrelevant features had been pecked just before making the choice response for groups Fixed Irrelevant (left) and Variable Irrelevant (right). Error bars indicate the standard error of the means.

was true for relevant feature pecking; in both groups, the percentage of pecks directed at the relevant features of the novel testing stimuli was high (72% and 73% for the Fixed Irrelevant group and the Variable Irrelevant group, respectively), albeit slightly lower than the percentage of pecks directed at the relevant features of the familiar training stimuli (84% and 82% for the Fixed Irrelevant group and the Variable Irrelevant group, respectively).

A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 2 (stimuli: training vs. testing) \times 4 (sessions) ANOVA on choice accuracy yielded a main effect of stimuli, $F(1, 6) = 15.15, p < .01, \eta_p^2 = .72, 95\% \text{ CI } [.08, .85]$; overall accuracy was slightly higher for the familiar training stimuli (91%) than for the novel testing stimuli (85%). No other main effects or interactions were statistically significant.

A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 2 (stimuli: training vs. testing) \times 4 (sessions) ANOVA on the percentage of relevant pecks again yielded a main effect of stimuli, $F(1, 6) = 12.15, p < .01, \eta_p^2 = .67, 95\% \text{ CI } [.04, .83]$; the percentage of relevant pecks to the familiar training stimuli was slightly higher (83%) than the percentage of relevant pecks to the novel testing stimuli (72%). No other main effects or interactions were significant. In addition, we compared the percentage of relevant pecks in

testing in each of the groups against a hypothesized mean of 50%. This difference was statistically significant in both groups for training trials: 95% CI [72.07, 96.76], $t(3) = 8.87, p < .005$, in Group Fixed Irrelevant, and 95% CI [65.99, 98.93], $t(3) = 6.27, p < .005$, in Group Variable Irrelevant, and for testing trials: 95% CI [54.06, 90.46], $t(3) = 3.81, p < .01$, in Group Fixed Irrelevant, and 95% CI [60.63, 85.45], $t(3) = 5.90, p < .01$, in Group Variable Irrelevant, confirming that relevant pecking was higher than expected by chance for both training and testing exemplars.

Although both accuracy and relevant pecking to the testing exemplars were robust, there was nonetheless a small decrement in accuracy compared to the training exemplars. In order to find possible reasons for this decrement, we looked at birds' categorization accuracy on trials when their observing pecks were directed either at the relevant or at the irrelevant exemplar features. As Figure 4 (bottom) shows, when the birds pecked at the relevant features of the exemplars, accuracy was very high, regardless of the irrelevant stimuli being novel or familiar. However, when the birds pecked at the irrelevant features of the exemplars, their accuracy was much lower. A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 2 (stimuli: training vs. testing) \times 4 (sessions) \times 2 (type of feature pecked: relevant vs. irrelevant) ANOVA on

choice accuracy revealed a significant effect of the type of feature pecked, $F(1, 6) = 115.20$, $p < .001$, $\eta_p^2 = .95$, 95% CI [.66, .97], confirming that accuracy was lower when the birds pecked at the irrelevant features of the displays (overall, 63%) than when they pecked at the relevant features of the displays (overall, 95%). Importantly, this disparity in accuracy was very similar regardless of the irrelevant features being novel, 61% versus 94%, or familiar, 66% versus 96%. It seems, therefore, that the drop in accuracy observed with the testing stimuli was due to the decrease in relevant pecks to the testing stimuli compared to the training stimuli.

Discussion

Birds in both the Fixed Irrelevant and Variable Irrelevant groups mastered their respective categorization tasks and reached very high accuracy levels. We did not, however, find a statistically significant difference in training days to criterion between Group Fixed Irrelevant and Group Variable Irrelevant. In testing, the two groups were virtually identical, with similarly high accuracy to both training and testing stimuli. So, it seems that the difference in the variability of irrelevant features did not affect the birds' learning or generalization performance. Of course, it could be that our present procedure was not sensitive enough to detect the effect of this variable; perhaps a larger number of irrelevant features or a more extreme irrelevant-relevant ratio would have yielded a different outcome.

Our pigeons not only learned to categorize the complex visual stimuli, but in doing so they also learned which stimuli were relevant and which stimuli were irrelevant to solving the task. At the beginning of training, the birds started pecking similarly at both the relevant and irrelevant features of the category exemplars. However, as training progressed, the pigeons increasingly concentrated their pecks on the relevant features of the category exemplars, thereby suggesting that these pecks were not randomly distributed; rather, the birds were tracking the relevant information to solve the categorization task. These results thus suggest that peck tracking can be a suitable measure of the allocation of overt attention in pigeons, just as eyetracking can be a suitable measure of the allocation of overt attention in humans.

In addition, our birds' peck tracking helps illuminate the mechanisms involved in the generalization of categorization to novel exemplars. The fact that we did not observe any generalization decrement when the birds pecked at the relevant features of the new testing exemplars suggests that the generalization decrement in accuracy that is frequently found in categorization and concept learning tasks (e.g., Bhatt et al., 1988; Roberts & Mazmanian, 1988; Young & Wasserman, 1997; Wright, 1997) might be due to an attentional failure rather than to the absence of a fully developed category or concept.

This observation raises another issue. Only the irrelevant features of the testing displays were novel; nonetheless, the birds showed an increased tendency to peck at these novel irrelevant features rather than at the familiar relevant features of the testing stimuli. This tendency is surprising because pigeons tend to prefer familiar over unfamiliar stimuli. It could be that the pigeons had not only learned to approach the relevant features of the stimuli, but to avoid the irrelevant features as well. When new irrelevant features were substituted for old irrelevant features in the testing

exemplars, these new features (that the birds had not learned to attend or to ignore) may have commanded more attention than the old irrelevant features (that the birds had learned to ignore); consequently, the new irrelevant features may have diverted attention away from the familiar relevant features. We will return to this issue in the General Discussion.

Experiment 2

Generalization testing in Experiment 1 involved the presentation of category exemplars that contained the same relevant features as in training, but new irrelevant features. In Experiment 2, we trained the same birds with category exemplars that contained the same irrelevant features as in original training, but new relevant features for both Category A and Category B. So, the categorization task in Experiment 2 was effectively a new categorization task, in which we expected the pigeons' accuracy to start at 50% chance level. However, learning might proceed rapidly or slowly depending on what the birds had learned in Experiment 1. If the birds had learned to allocate their attention to the relevant features without any corresponding learning to ignore the irrelevant features, then learning should proceed at the same speed as in Experiment 1. However, if the birds had learned to allocate their attention to the relevant features and, concurrently, to ignore the irrelevant features, then learning might proceed even faster than in Experiment 1, because the pigeons' attention would be directed away from the old irrelevant features and toward the new relevant features, the very features to which they must now attend in order to successfully solve the task.

Moreover, it could be that, under these circumstances, we might observe a difference in performance between the Fixed Irrelevant and Variable Irrelevant groups. The Fixed Irrelevant group had been presented with category exemplars that each contained the same two irrelevant features, whereas the Variable Irrelevant group had been presented with category exemplars that contained irrelevant features that varied from trial to trial. Furthermore, because there were eight irrelevant features in Variable Irrelevant training, but only two irrelevant features in Fixed Irrelevant training, the individual irrelevant features occurred at a much lower rate in the former condition than the latter. If the pigeons had learned to ignore the irrelevant features in the category exemplars, then this learning might well have been stronger in Group Fixed Irrelevant than in Group Variable Irrelevant, thereby giving the Fixed Irrelevant group a category learning advantage over the Variable Irrelevant group. Thus, the Fixed Irrelevant group might show high accuracy and relevant pecking earlier because strong avoidance of the irrelevant features should facilitate approaching the new relevant features from the beginning; the Variable Irrelevant group, however, might take longer to reach high accuracy and relevant pecking because avoidance of the irrelevant features was not so strongly established.

After the birds mastered the new relevant categorization task, we added another phase in which we maintained the informational value of the relevant and irrelevant features, but we switched the contingencies between each category and its associated response button. In this reversal phase, the response button that was previously correct for Category A became the correct response button for Category B, and vice versa. From this switching point onward, the pigeons should have continued attending to the relevant fea-

tures of the category exemplars, thereby promoting perhaps faster learning than in the original learning phase in Experiment 1.

Method

Subjects and apparatus. The same eight pigeons that were studied in Experiment 1 served as subjects in Experiment 2. All of the pigeons remained in their original group: Fixed Irrelevant or Variable Irrelevant. The apparatus was the same as in Experiment 1.

Procedure. The initial procedure in Experiment 2 was identical to that in Experiment 1, except for the replacement of the relevant features of the stimuli. Birds were shown two categories with new relevant features and the same irrelevant features as in Experiment 1. Now, two new relevant features defined Category A and two new relevant features defined Category B. The total number of unique training exemplars was 24 for Category A and 24 for Category B in Group Fixed Irrelevant; that number was 96 for Category A and 96 for Category B in Group Variable Irrelevant, just as in Experiment 1.

Once the pigeons reached 85% correct on 2 consecutive days for each of the two categories, we maintained the birds on this training regimen for 2 weeks more before starting the reversal phase in order to ensure stable categorization performance. In the subsequent reversal phase, all of the category exemplars were the same, but the correct response for each of the categories was switched; now, the response button that had been associated with Category A was associated with Category B and the response button that had been associated with Category B was associated with Category A. All other procedural details were the same as in prior training phases.

Results

New relevant training. Overall, the new category discrimination with new relevant cues and old irrelevant cues in Experiment 2 was learned faster than the original category discrimination in Experiment 1 (see Table 1 for a comparison of days to criterion in the different experimental phases). Group Fixed Irrelevant took a mean of 5.25 ($SD = 1.71$), 6.75 ($SD = 1.89$), 9.75 ($SD = 4.19$), and 15.50 ($SD = 8.18$) days to reach categorization accuracy levels of 65%, 75%, 85%, and 85% on 2 consecutive days, respectively. Group Variable Irrelevant took a mean of 6.50 ($SD = 2.38$), 11.75 ($SD = 1.70$), 16.75 ($SD = 5.37$), and 18.50 ($SD = 4.30$) days to reach the same categorization accuracy levels of 65%, 75%, 85%, and 85% on 2 consecutive days, respectively. Statisti-

cal analysis suggested that the Fixed Irrelevant group took fewer sessions to reach each of the criterion levels. A Group \times Criterion (65% vs. 75% vs. 85% vs. 85%_2) ANOVA showed a statistically significant main effect of group, $F(1, 6) = 7.22$, $p = .01$, $\eta_p^2 = .55$, 95% CI [.00, .76]. Thus, overall, pigeons in the Fixed Irrelevant group learned faster than pigeons in the Variable Irrelevant group.

Just as in Experiment 1, we constructed Vincentized learning curves by grouping all of the training sessions until reaching the criterion of 85% on 2 consecutive days into eight blocks (eight was the smallest number of sessions that it took a bird to reach criterion). Mean accuracy on the last training block was 90% for the Fixed Irrelevant group and 90% for the Variable Irrelevant group (Figure 5, left). A 2 (group) \times 8 (blocks) ANOVA on choice accuracy scores yielded a main effect of block, $F(7, 42) = 56.52$, $p < .001$, $\eta_p^2 = .90$, 95% CI [.82, .92], confirming that accuracy increased as training progressed in both groups. There was also a main effect of group, $F(1, 6) = 8.62$, $p < .01$, $\eta_p^2 = .59$, 95% CI [.00, .79], due to slightly higher overall accuracy in Group Fixed Irrelevant (75%) than in Group Variable Irrelevant (70%). The Group \times Block interaction was not significant, $F < 1$.

Next, we examined pigeons' feature tracking behavior. Pecks at the relevant features are displayed, along with accuracy, in Figure 5 (left). During original training in Experiment 1, we had seen that as accuracy increased, relevant pecks gradually increased as well. But, the pattern of relevant pecks was different now and, curiously, a large difference between the groups emerged. Group Fixed Irrelevant predominately pecked at the relevant features of the stimuli from the very beginning (80% in Block 1), and its relevant pecking remained very high throughout the entire training period (last block, 88%). Group Variable Irrelevant modestly pecked at the relevant features of the stimuli at the beginning of training (60% in Block 1), and progressively increased to 72% in the last training block; however, feature tracking in the Variable Irrelevant group never attained the high level of the Fixed Irrelevant group. A 2 (group) \times 8 (blocks) ANOVA on the percentage of relevant pecks yielded a main effect of group, $F(1, 6) = 34.16$, $p < .001$, $\eta_p^2 = .85$, 95% CI [.28, .92], confirming that relevant pecking was higher in Group Fixed Irrelevant than in Group Variable Irrelevant. There was no main effect of block nor a Group \times Block interaction, $F < 1$. Thus, overall, no significant changes in relevant pecking occurred during training. In addition, we compared the percentage of relevant pecks in the last training block in each of the groups against a hypothesized mean of 50%. This difference was statistically significant in both groups: 95% CI [73.80, 100], $t(3) = 8.35$, $p < .005$, in Group Fixed Irrelevant, and 95% CI [57.52, 86.68], $t(3) = 4.82$, $p < .01$, in Group Variable Irrelevant, confirming that relevant pecking was higher than expected by chance in the last training block.

We also examined birds' accuracy when the observing pecks were directed at either the relevant or the irrelevant features of the category exemplars. As in Experiment 1, when the birds pecked at the relevant features, accuracy was very high; after pecking at the irrelevant features, their accuracy was much lower (Figure 6, left). Nonetheless, although accuracy in the Fixed Irrelevant group remained around 50% when pigeons had pecked at the irrelevant features, accuracy in the Variable Irrelevant group increased throughout training. A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 8 (blocks) \times 2 (type of feature pecked: relevant vs. irrelevant) ANOVA on choice accuracy revealed a significant

Table 1
Number of Days to Criterion in Experiment 1 (Original Training) and Experiment 2 (New Relevant Training and Reversal Training)

Group	Original training	New relevant	Reversal
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Fixed irrelevant	25.75 (14.86)	15.50 (8.19)	21.75 (6.02)
Variable irrelevant	34.50 (22.78)	18.50 (4.20)	37.75 (22.07)

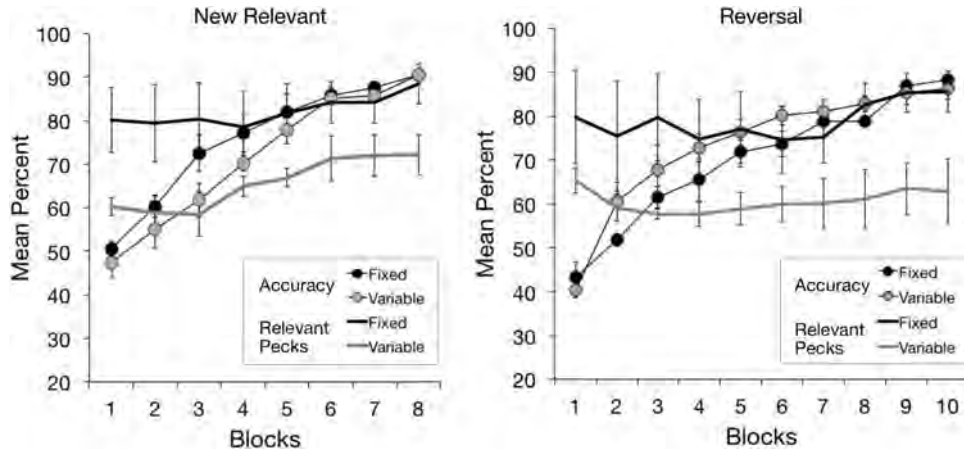


Figure 5. Mean percent accuracy and relevant pecks across training blocks for groups Fixed Irrelevant and Variable Irrelevant in the new relevant phase of Experiment 2 (left) and in the reversal phase of Experiment 2 (right). Error bars indicate the standard error of the means.

main effect of the type of feature pecked, $F(1, 6) = 361.80, p < .001, \eta_p^2 = .98, 95\% \text{ CI } [.87, .99]$, confirming that accuracy was lower when the birds pecked at the irrelevant features (overall, 58%) than when the birds pecked at the relevant features of the displays (overall, 79%). There was also a significant main effect of block, $F(7, 42) = 42.72, p < .001, \eta_p^2 = .88, 95\% \text{ CI } [.77, .90]$, and a Block \times Type of feature interaction, $F(7, 42) = 14.60, p < .001, \eta_p^2 = .71, 95\% \text{ CI } [.48, .76]$, showing that the disparity in accuracy due to the type of feature pecked increased as training progressed. Finally, there was a Group \times Block interaction, $F(7, 42) = 3.39, p < .01, \eta_p^2 = .36, 95\% \text{ CI } [.04, .46]$, and a Group \times Type of feature interaction, $F(1, 6) = 43.97, p < .001, \eta_p^2 = .88, 95\% \text{ CI } [.36, .94]$, indicating that the dependency of accuracy on the feature pecked differed between groups; although accuracy after pecking at the relevant features increased similarly in both groups, accuracy after pecking at the irrelevant features remained at

chance level for the Fixed Irrelevant group, but increased for the Variable Irrelevant group.

Reversal training. Switching the associative significance of the relevant cues dramatically disrupted the birds' categorization performance. Learning to discriminate two new categories with new relevant cues and old irrelevant cues had proceeded much faster than the original learning in both groups, documenting a positive carryover effect. However, switching the response buttons associated with each of the categories now resulted in the birds' initial accuracy falling below chance (as expected), followed by slow learning of the new contingencies. As can be seen in Table 1, the number of days that it took the birds to reach the criterion of 85% correct in 2 consecutive days was very similar for original training (Experiment 1) and for this reversal phase. Group Fixed Irrelevant took a mean of 8.25 ($SD = 3.86$), 13.25 ($SD = 5.05$), 20.25 ($SD = 5.50$), and 21.75 ($SD = 6.02$) days to reach catego-

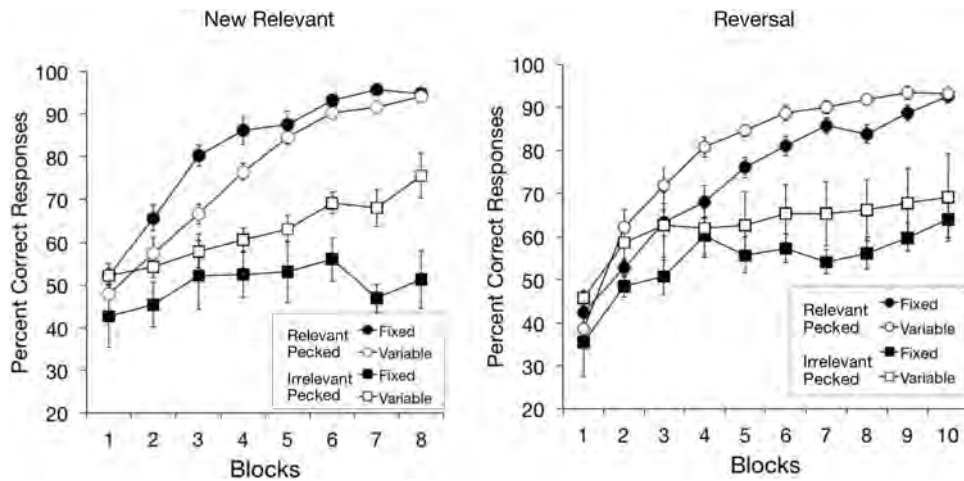


Figure 6. Mean percent accuracy across training depending on whether the relevant or the irrelevant features had been pecked just before making the choice response for groups Fixed Irrelevant and Variable Irrelevant in the new relevant phase of Experiment 2 (left) and in the reversal phase of Experiment 2 (right). Error bars indicate the standard error of the means.

ization accuracy levels of 65%, 75%, 85%, and 85% on 2 consecutive days, respectively. Group Variable Irrelevant took a mean of 10.50 ($SD = 7.93$), 15.50 ($SD = 13.52$), 25.75 ($SD = 19.55$), and 37.75 ($SD = 22.06$) days to reach categorization accuracy levels of 65%, 75%, 85%, and 85% on 2 consecutive days, respectively. A Group \times Criterion (65% vs. 75% vs. 85% vs. 85%_2) ANOVA yielded no significant differences between the groups; at all criterion levels, all $F_s < 1$.

As in prior stages, we constructed Vincentized learning curves by grouping all of the training sessions until reaching 85% on 2 consecutive days into 10 blocks (10 was the smallest number of sessions that it took a bird to reach 85% accuracy on 2 consecutive days). Mean accuracy on the last training block was 88% for the Fixed Irrelevant group and 86% for the Variable Irrelevant group. A 2 (group) \times 10 (blocks) ANOVA yielded a main effect of training block, $F(9, 54) = 47.3, p < .001, \eta_p^2 = .89, 95\% \text{ CI } [.80, .91]$, confirming that accuracy increased as training progressed. There was also a main effect of group, $F(1, 6) = 7.84, p < .01, \eta_p^2 = .57, 95\% \text{ CI } [.00, .77]$; overall accuracy was slightly higher in Group Variable Irrelevant (73%) than in Group Fixed Irrelevant (69%). The Group \times Block interaction was not significant, $F < 1$.

Pecks at the relevant features, along with categorization accuracy, are displayed in Figure 5 (right). The differences in relevant pecking between groups that emerged in the new relevant phase continued during this reversal phase. Group Fixed Irrelevant pecked at the relevant features of the exemplars from the very beginning (80% in Block 1) and its relevant pecking remained very high throughout the entirety of reversal training (last block, 85%). Relevant feature pecking in Group Variable Irrelevant was rather low at the beginning of reversal training (65% in Block 1) and it remained at a similar level until the end of this phase (last block, 63%). A 2 (group) \times 10 (blocks) ANOVA on the percentage of relevant pecks yielded a main effect of group, $F(1, 6) = 37.30, p < .001, \eta_p^2 = .86, 95\% \text{ CI } [.31, .93]$, confirming that relevant pecking was higher in Group Fixed Irrelevant than in Group Variable Irrelevant. There was no main effect of block on relevant pecking nor a Group \times Block interaction, $F_s < 1$. Thus, overall, no significant changes in relevant pecking occurred during training. In addition, we compared the percentage of relevant pecks in the last training block in each of the groups against a hypothesized mean of 50%. This difference was statistically significant in Group Fixed Irrelevant: 95% CI [70.38, 100], $t(3) = 7.47, p < .005$, but not in Group Variable Irrelevant: 95% CI [39.21, 86.35], $t(3) = 1.72, p = .09$, confirming that relevant pecking was higher than expected by chance at the end of training in Group Fixed Irrelevant but not in Group Variable Irrelevant.

Next, we examined pigeons' accuracy when observing pecks were directed at either the relevant or the irrelevant features of the category exemplars. As in prior phases, when the birds pecked at the relevant features, accuracy was very high; after pecking at the irrelevant features, their accuracy was much lower (Figure 6, right). A 2 (group: Fixed Irrelevant vs. Variable Irrelevant) \times 10 (blocks) \times 2 (type of feature pecked: relevant vs. irrelevant) ANOVA on choice accuracy revealed a significant main effect of the type of feature pecked, $F(1, 6) = 116.94, p < .001, \eta_p^2 = .95, 95\% \text{ CI } [.66, .97]$, confirming that accuracy was lower when the birds pecked at the irrelevant features (overall, 59%) than when the birds pecked at the relevant features (overall, 76%). There was also a significant main effect of block, $F(9, 54) = 24.15, p < .001,$

$\eta_p^2 = .80, 95\% \text{ CI } [.65, .83]$, and a Block \times Type of feature interaction, $F(9, 54) = 4.10, p < .001, \eta_p^2 = .41, 95\% \text{ CI } [.11, .48]$, showing that the disparity in accuracy due to the type of feature pecked increased as training progressed. Finally, there was a main effect of group, $F(1, 6) = 17.34, p < .001, \eta_p^2 = .74, 95\% \text{ CI } [.10, .86]$, due to overall accuracy being higher in Group Variable Irrelevant (71%) than in Group Fixed Irrelevant (64%).

Discussion

Learning with new relevant cues proceeded faster (overall, an average of 17 days) than original training (Experiment 1; overall, an average of 30 days). Accuracy started at chance level in both groups and gradually increased to 90%. But, Group Fixed Irrelevant was significantly faster (an average of 15 days) to master the new discrimination than Group Variable Irrelevant (an average of 18.5 days). The reason for this difference could be that birds in Group Fixed Irrelevant directed most of their pecks at the relevant features from the beginning of training (80% in Block 1), whereas birds in Group Variable Irrelevant, although pecking at the relevant features more often than at the irrelevant features from the beginning of training as well (60% in Block 1), did so to a lesser degree.

Thus, the number of irrelevant features in the category exemplars now supported a between-groups disparity. It could be that, during initial training in Experiment 1, learning to suppress responding to the irrelevant features was more effective in the Fixed Irrelevant group (where the same irrelevant features appeared on 100% of the trials) than in the Variable Irrelevant group (where the same irrelevant features appeared on only 25% of the trials). So, when new relevant features were introduced, the Fixed Irrelevant birds' observing responding was more strongly directed to these new features than to the old irrelevant features that they had learned to ignore, whereas the Variable Irrelevant birds' observing responding was more diffusely distributed (although still directed to the new relevant features more than to the old irrelevant features).

Reversal learning was slower (overall, an average of 30 days) than training with new relevant cues (overall, an average of 17 days); indeed, reversal learning took the birds the same number of sessions as original learning (Experiment 1; overall, an average of 30 days). This slow reversal learning is difficult to explain from an attentional perspective because the birds were already pecking at and therefore allocating their attention to the cues that were relevant to the discrimination. According to traditional attentional theories (e.g., Mackintosh, 1975) as well as more recent attentional learning models (e.g., George & Pearce, 2012), the allocation of attention to relevant cues, higher than chance in both new relevant and reversal training, should have facilitated learning in both phases. Although new relevant training indeed proceeded faster than original training, reversal training took much longer. Although reversed choice responses were required in the new relevant and reversal phases, the relevant features of the task remained the same throughout the two phases; therefore, the attention to the relevant features that had been established during the new relevant phase was appropriate for the reversal phase; nevertheless, more rapid learning was not observed. It seems that the stimulus-response learning from the new relevant phase greatly interfered with learning the opposite response to the same category stimuli,

and this interference was much stronger than any facilitation that established attention to the relevant cues might have generated.

In addition, relevant feature pecking in Group Variable Irrelevant did not increase during reversal learning, remaining around 60%. When looking at these birds' categorization accuracy depending on the type of feature pecked, we did see that, although the birds were clearly more accurate after pecking at the relevant features, they also improved their accuracy after pecking at the irrelevant features. Indeed, overall accuracy of the Variable Irrelevant group was slightly higher than overall accuracy of the Fixed Irrelevant group. The decrease in overall relevant pecks in Group Variable Irrelevant, compared to original and new relevant training, could then be a consequence of this improvement.

It is not obvious why birds in the Variable Irrelevant group maintained a low level of relevant pecking in Experiment 2. It seems that the introduction of new relevant cues and, especially, the reversal in the contingencies between each category and its associated response button disrupted their prior strategy. As mentioned above, it could be that, in Experiment 1, birds in the Fixed Variable group had learned to avoid the irrelevant cues, whereas birds in the Variable Irrelevant group had learned to approach the relevant cues. The absence of the relevant cues that birds in the Variable Irrelevant group had learned to attend to in Experiment 1 may be the cause of these birds' decrease in relevant pecking.

Still, the birds had to look at the relevant cues to be able to choose the correct response; so, it could be that the Variable Irrelevant birds were now finding the relevant cues and looking at them either before or after completing the required number of pecks at the display, and completing this requirement in the most convenient location (see General Discussion about birds' preferences for specific locations).

General Discussion

Several theories of categorization and discrimination learning hypothesize an intimate interrelation between attention to the stimulus attributes relevant to learn a task and mastery of that task (e.g., Mackintosh, 1975; George & Pearce, 2012; Kruschke, 1992; LePelley, 2004). In animals, most empirical support for these theories comes from transfer studies, in which the amount of attention allocated to a cue in original learning is assumed to determine how later learning about that cue proceeds; thus, attention is inferred after learning has already taken place. Our results show that pigeons' attention to relevant and irrelevant attributes of the training exemplars (assessed by directly recording the location of birds' observing responses) can effectively be monitored during learning of a categorization task, and that such peck tracking can shed fresh light on the category learning process itself.

We found that as accuracy progressively increased, pecks at the relevant features of the category exemplars increased as well (Experiment 1); this relationship suggests that as pigeons were learning to categorize the stimuli, they were also paying increasing attention to the relevant attributes of the training stimuli. This result implies a direct relationship between pigeons' categorization behavior and their allocation of attention to the relevant stimulus information. In addition, this finding joins the results of prior work suggesting that animals do indeed attend more to those stimulus attributes that are relevant than irrelevant to solving a discrimination problem (e.g., Dittrich et al., 2010; Dopson, Esber, & Pearce,

2011; George & Pearce, 1999, 2012). Furthermore, the progressive increase in relevant feature tracking (Experiment 1) supports theories which propose that attention is not merely engaged by the inherent salience of a stimulus (e.g., Rescorla & Wagner, 1972), but is subject to dynamic changes depending on the ability of that stimulus attribute to predict reinforcing consequences (e.g., George & Pearce, 2012; Kruschke, 1992; LePelley, 2004; Mackintosh, 1975).

Interestingly, birds' categorization accuracy was much higher after pecking the relevant features than after pecking the irrelevant features of the category exemplars; precisely because the likelihood of the bird choosing correctly is so much higher after pecking the relevant features, such relevant pecking is prone to increase. So, it seems that the rise in tracking the relevant features follows the increase in choice accuracy; that is, when, after paying attention to a specific cue, the choice response is correct and the bird consequently receives reinforcement, more attention will be paid to that cue on future trials. A similar trend was reported by Dittrich et al. (2010), in which pigeons had to learn a person present versus person absent discrimination task: as accuracy rose, the birds were more likely to peck at the distinctive features of the images.

It could be argued that pigeons tend to peck at the relevant cues because they are strong predictors of the outcome and, consequently, they acquire high associative strength. Indeed, relevant cues are relevant because they are reliable and strong predictors of the correct outcome; so, it is extremely difficult to disentangle attention from associative strength and complex designs are required in order to do so. 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 experiments, it is conceivable that the processing mechanisms in George and Pearce's study are similar to our present study. This is an issue that we hope to address in future research.

Critical for the generalizability of our findings is the observation that not only animals', but humans' attentional behavior follows a common pattern. In a categorization task similar to our own, in which the elements of the compound stimuli were separately located on the computer screen, Rehder and Hoffman (2005) found that people's eye movements toward the relevant elements of the category stimuli tended to follow rather than to precede improvements in task accuracy; correct responses were already very high before people fully deployed their attention to the relevant features of the category stimuli.

Note that we are not proposing that changes in peck tracking in pigeons (or eyetracking in humans) will not happen until a high level of accuracy has been attained. Rather, we are suggesting that the delivery of reinforcement after presentation and pecking of a specific stimulus attribute will incrementally influence the amount of attention that will later be allocated to that attribute; the amount of attention allocated to that attribute will, in turn, influence the likelihood of obtaining reinforcement after its presentation. So, both changes in attention and increases in choice accuracy are, most likely, progressive and synergetic processes.

Our experimental design not only permitted us to monitor changes in selective attention during categorization learning, it also allowed us to examine how peck tracking and/or attention

proceeds once initial learning has been established and changes in the category exemplars and categorization rules are introduced. When the relevant stimulus features were maintained and new irrelevant features were introduced (Experiment 1, testing phase), birds' choice accuracy was still very high, but a bit lower than with the original training exemplars. This small decrement in accuracy was actually due to an increase in pigeons' pecking these new irrelevant features. So, it seems that even after the birds had learned to allocate their attention to the relevant features, and this strategy was still highly efficient, new irrelevant features nonetheless commanded some attention; the birds actually pecked at the new irrelevant features more than they had pecked at the original irrelevant features. This tendency to peck at the new irrelevant features 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 original irrelevant features; that is, the birds probably learned both to attend to the relevant features and to ignore the irrelevant features.

When we created new categories with the same irrelevant features as the original training stimuli but with new relevant features, we observed the first disparity between the Fixed Irrelevant and Variable Irrelevant groups. Group Fixed Irrelevant learned the task faster and its overall accuracy was higher than Group Variable Irrelevant. Although in both groups accuracy began at chance level, relevant peck tracking in Group Fixed Irrelevant was high from the outset, whereas there was a drop in relevant peck tracking in Group Variable Irrelevant compared to the end of initial training (although it was still above chance level). This disparity in relevant peck tracking might be the reason for the difference in learning rate and overall accuracy between the groups. Relevant peck tracking progressively increased throughout training in the Variable Irrelevant group, but it remained much lower than in the Fixed Irrelevant group.

During the final categorization reversal phase, accuracy progressively rose in both groups, but overall accuracy was now higher in the Variable Irrelevant group than in the Fixed Irrelevant group. Relevant pecking was very high from the beginning in the Fixed Irrelevant group, but comparatively low (not different from chance level at the end of the reversal phase) in the Variable Irrelevant group; still, pigeons in the Variable Irrelevant group managed to successfully solve the task. So, even when accuracy in the Variable Irrelevant group gradually increased until it reached very high levels, its relevant pecking stayed low throughout the entire reversal phase. This dissociation between accuracy and relevant pecking also suggests that pecking at the relevant cues is not a necessary consequence of the relevant cues being the best predictors of the occurrence of the outcome.

Theoretical explanations: George and Pearce's (2012) Model Predictions

Overall, our findings support theories that propose that the allocation of attention to a stimulus changes during the course of learning depending on the role the stimulus plays in the discrimination to be learned (e.g., Mackintosh, 1975; George & Pearce, 2012; Kruschke, 1992; LePelley, 2004). In order to better understand our pattern of results, we focused on George and Pearce's (2012) recent formalization of an attentional theory of associative learning. The advantage of George and Pearce's model over prior

attentional models is that it considers attention in two different ways, a distinction for which there is extensive evidence (see Pearce & Mackintosh, 2010, for a review), but which other models like Mackintosh (1975) or Kruschke (1992) do not contemplate.

George and Pearce proposed (see also, LePelley, 2004, for a similar suggestion) one type of attention that allows the organism to focus on stimuli whose significance remains uncertain (equivalent to the associability concept in Pearce & Hall, 1980); this type of attention, implemented in the σ parameter, allows the organism to focus on stimuli whose significance remains uncertain. The σ parameter, or conditionability, increases on trials in which the outcome is surprising, whereas it decreases on trials in which the outcome is predicted. The second type of attention allows the organism to focus on significant events in the environment, strengthening the salience of relevant stimuli and weakening the salience of irrelevant stimuli (equivalent to the attentional parameter in Mackintosh, 1975). This second type of attention is implemented in the α parameter. In George and Pearce's model, α increases as the stimulus becomes a reliable predictor of the outcome of a trial, whereas α decreases when the stimulus does not contribute to predicting the outcome of a trial.

Because of our interest in dynamic changes in the allocation of attention to the relevant and irrelevant stimuli in a discrimination, we analyzed the predictions of George and Pearce's (2012) model, concentrating on response strength predictions and on variations in the α and σ parameters. We used the MATLAB simulator provided by George and Pearce, setting the stimulus learning rate θ parameter at 0.1 and both the initial α and σ at 0.4 (parameters were chosen following George & Pearce's own selections for their simulations; in any case, different parameters do not change the pattern of the simulations). Also, because we had two different outcomes (one correct response button for Category A and a second correct response button for Category B) rather than an outcome that could either be present or absent, we set the correct outcome at 1 for Category A and at -1 for Category B (as per the suggestion of David George, personal communication, June 20, 2013).

Simulations of our experimental tasks are depicted in Figure 7. During initial categorization learning (Experiment 1), response strength and attention to the relevant cues (the α relevant parameter) is predicted to progressively increase in both groups (Figure 7, left panel), just as accuracy and peck tracking increased for our birds (Figure 3, left); also, conditionability (the σ parameter) is predicted to progressively decrease as the outcome is correctly predicted. The model predicts faster learning in the Variable Irrelevant group because the greater the number of irrelevant cues, the lower their individual α and, consequently, the easier it is for the α of the relevant cues to increase and contribute to the correct prediction of a trial outcome. Although we did not find a trend toward faster learning in the Variable Irrelevant group, this prediction does agree with studies of language and speech categorization (Apfelbaum et al., 2013; Gómez, 2002; Rost & McMurray, 2009), which have shown that variability in irrelevant information (e.g., in speaker voices) significantly improves identification of relevant category information (e.g., particular phonemes). Perhaps this result can also be found in a visual categorization task if the relevant and irrelevant characteristics of the stimuli belong to different dimensions or if the relevant-irrelevant information ratio were more extensively manipulated.

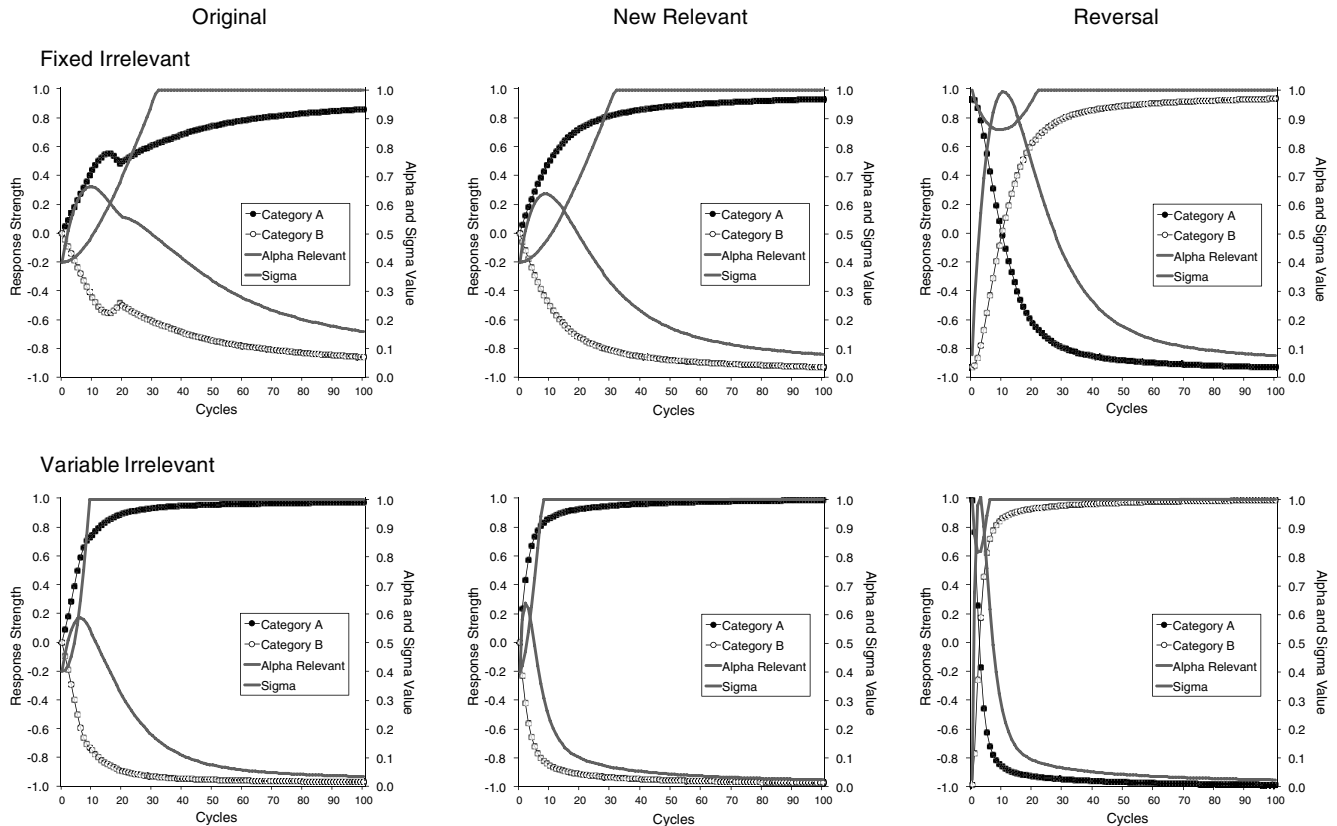


Figure 7. Simulated predictions of George and Pearce's (2012) model for each of the two groups (Fixed Irrelevant and Variable Irrelevant) in Experiment 1 (Original) and 2 (New Relevant and Reversal). Correct Category A responses move upward, whereas correct Category B response move downward. Alpha represents the strength of attention to relevant stimulus attributes and Sigma represents the conditionability of the exemplars in each category.

When new relevant features were introduced (Experiment 2), George and Pearce's (2012) model predicts faster learning compared to the original phase (Figure 7, middle panel), because attention is no longer being allocated to the irrelevant features (the α value of the irrelevant features had decreased from 0.4 to 0.05 during the initial learning phase). Indeed, it took our birds fewer days to learn the categorization task with the new relevant features than it took them to learn the original discrimination (see Table 1). Also, pigeons in the Fixed Irrelevant group concentrated their pecks on the new relevant features from the very beginning of new relevant training, but there was an immediate drop in relevant pecks for the Variable Irrelevant group, which only slowly recovered to prior levels. George and Pearce's model predicts that because for both groups the α value for the irrelevant features should be equally low, all of the pigeons ought to concentrate their pecks on the relevant features. Hence, the obtained difference in relevant pecks between groups cannot be explained by George and Pearce's model, unless it is argued that the α value for the irrelevant features in the Variable Irrelevant group did not reach the same low level as in the Fixed Irrelevant group by the end of original training, perhaps because the eight irrelevant features in the former case were presented less frequently than the

two irrelevant features in the latter case. Consequently, these irrelevant cues in the Variable Irrelevant group could still have recruited some attention when the new relevant cues were introduced.

The predictions of George and Pearce's (2012) model for our reversal phase are shown in Figure 7 (right, panel). According to the model, when the trained features are now paired with the opposite outcomes, there is a transient decrease in attention to the relevant features (because they initially predict the incorrect outcome) that is accompanied by a transient increase in attention to the irrelevant features. Indeed, relevant pecks dropped in both groups when the category-response contingencies were reversed. But, in Group Variable Irrelevant, contrary to expectations, relevant pecks did not increase throughout reversal training. It seems that, because pecking at the relevant features was no longer reliably followed by food reinforcement, the birds' peck tracking was drastically disrupted, and they developed other strategies to solve the task. Nonetheless, Group Fixed Irrelevant responded in closer accord with the predictions of George and Pearce's model; relevant feature tracking fell slightly at the beginning of the reversal phase (from 88% in the last block of new relevant training to 80% in the first block of reversal training), and increased again (up to 85% in the last block of reversal training) throughout the reversal

phase (although these numerical variations were not statistically significant).

Also, it can be seen that σ , the conditionability of the stimulus exemplars, which was very low at the outset of the reversal phase, steeply increased when the stimulus exemplars began predicting the opposite outcome. Here, the difference in the two types of attention (that other models such as Mackintosh (1975) or Kruschke (1992) have no way of representing) is clearly revealed: α of the relevant cues is high throughout the entire reversal phase, because their status as relevant cues does not change, but they now predict the opposite outcome, so their significance becomes uncertain and, consequently, σ greatly increases at the beginning of the reversal phase and then gradually decreases as the new contingencies are learned.

Peck Tracking as a Measure of Birds' Attention

We believe that our results support the further development of peck tracking as an effective measure of birds' overt attention during visual categorization learning, much as eyetracking has come to serve as an effective measure of overt attention in human category learning (Rehder & Hoffman, 2005). Nonetheless, we acknowledge that peck location is not a perfect measure of attention. For example, peck location might not be solely determined by attention, but might be influenced by other factors as well.

Indeed, we observed that some of our birds preferred to peck particular spatial locations. In Group Fixed Irrelevant, two out of the four birds concentrated their pecks on the bottom left and bottom right areas of the exemplars, whereas the other two birds distributed their pecks similarly over all four locations. In Group Variable Irrelevant, two out of the four birds also concentrated their pecks on the bottom left and bottom right areas of the exemplars, whereas the other two birds distributed their pecks on the bottom left, bottom right, and top right areas of the exemplars. Note that each category exemplar had two relevant and two irrelevant features; so a bird that pecks only at the bottom two features will find at least one relevant feature in one of those two locations on 75% of the trials. If the features appearing in those two locations are both irrelevant (which happens on 25% of the trials), then the bird still has a 50% chance of making the correct response. Such a bird can therefore reach a very high level of categorization accuracy (87.5% correct) without ever having to peck at the top two nonpreferred locations. Hence, the spatial location of the features can have an important influence on peck location and can negatively affect the tracking of relevant features. Modest adjustments in the placement of the attributes of the training stimuli might minimize the influence of such position preferences.

Nevertheless, when one relevant and one irrelevant feature was presented in those preferred locations, even those four birds (two in each group) that had exhibited a strong preference for the bottom locations chose the relevant rather than the irrelevant feature between 80% and 98% of the time (in the last block of original, new relevant, and reversal training, regardless of group). Therefore, even when the birds' peck distributions were strongly influenced by the spatial location of the features, when spatial location was "neutralized," a strong tendency to peck at the rele-

vant rather than the irrelevant feature could clearly be observed, again attesting to the promise of our peck tracking technique.

Conclusions

Rehder and Hoffman (2005) summarized the results of their eyetracking category learning experiments as follows:

The first [finding] is that participants learned to allocate attention to stimulus dimensions in a way that optimized their ability to discriminate categories (. . .) The second finding is that learners tend to fixate all stimulus dimensions early in learning (. . .) The third finding is that changes in eye fixations to only relevant dimensions tend to occur after errors have been greatly reduced (pp. 38–39)

Point by point, these results closely accord with the findings of our peck tracking pigeon project. Our birds learned to allocate their attention to the relevant features of the category exemplars, the very features that allowed them to successfully solve the discrimination; the birds pecked equally often to the relevant and irrelevant features of the exemplars at the start of training; and peck tracking of the relevant features increased as a consequence of accuracy rising when the relevant features were pecked.

These striking empirical parallels between humans and pigeons given challenging visual categorization problems strengthens the case for common mechanisms participating in higher-order cognitive behaviors. Our experimental strategy opens the door to a world of possibilities for research in comparative cognition. It shows that, just as do people, pigeons have complex problem-solving abilities that extend beyond simple discrimination learning. Plus, our investigative strategy can very effectively and inexpensively be deployed to concurrently monitor organisms' attentional and categorization behavior in real time, thereby allowing us to study the dynamic interplay of these factors in higher-order cognition.

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.
- Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). New York, NY: Worth.
- Apfelbaum, K. S., Hazeltine, E., & McMurray, B. (2013). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology, 49*, 1348–1365. doi:10.1037/a0029839
- 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. doi:10.1016/S0166-4328(97)00121-6
- Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., Jr., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 219–234. doi:10.1037/0097-7403.14.3.219
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436. doi:10.1163/156856897X00357
- Brooks, D. I., Ng, K. H., Buss, E. W., Marshall, A. T., Freeman, J. H., & Wasserman, E. A. (2013). Categorization of photographic images by rats using shape-based image dimensions. *Journal of Experimental Psychology: Animal Behavior Processes, 39*, 85–92. doi:10.1037/a0030404

- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827–1837. doi:10.1016/0042-6989(95)00294-4
- Dittrich, L., Rose, J., Buschmann, J.-U., Bourdonnais, M., & Güntürkün, O. (2010). Peck tracking: A method for localizing critical features within complex pictures for pigeons. *Animal Cognition*, *13*, 133–143. doi:10.1007/s10071-009-0252-x
- Dopson, J. C., Esber, G. R., & Pearce, J. M. (2011). Changes in attention to an irrelevant cue that accompanies a negative patterning discrimination. *Learning & Behavior*, *39*, 336–349. doi:10.3758/s13420-011-0029-3
- 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: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: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 and Computers*, *36*, 355–362. doi:10.3758/BF03195582
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436. doi:10.1111/1467-9280.00476
- Hayne, H. (1996). Categorization in infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 10, pp. 79–120). Norwood, NJ: Ablex.
- Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, *146*, 549–551. doi:10.1126/science.146.3643.549
- Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining a prototype. *Journal of Experimental Psychology*, *101*, 116–122. doi:10.1037/h0035772
- James, W. (1890/1950). *The principles of psychology*. New York, NY: 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, NY: Appleton-Century-Crofts.
- Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology*. New York, NY: Appleton-Century-Crofts.
- Kelley, K. (2007a). Methods for the behavioral, educational, and educational sciences: An R package. *Behavior Research Methods*, *39*, 979–984. doi:10.3758/BF03192993
- Kelley, K. (2007b). *MBESS: Methods for the behavioral, educational, and social sciences*. R package (Version 0.0.8) [software] Retrieved from <http://CRAN.R-project.org/>
- Kling, J. W., & Riggs, L. A. (1971). *Woodworth and Schlossberg's experimental psychology* (3rd ed.). New York: Holt, Rinehart, and Winston.
- 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:10.1037/0096-3445.137.1.52
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44. doi:10.1037/0033-295X.99.1.22
- 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:10.1037/h0058097
- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, *57*, 193–243. doi:10.1080/02724990344000141
- Mackintosh, N. J. (1965). Selective attention in animal discrimination learning. *Psychological Bulletin*, *64*, 124–150. doi: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:10.1037/h0076778
- Mackintosh, N. J., & Little, L. (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science*, *14*, 5–6. doi:10.3758/BF03336395
- 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:10.3758/LB.36.3.188
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532–552. doi:10.1037/0033-295X.87.6.532
- Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In C. Mitchell, and M. Le Pelley (Eds.) *Attention and associative learning: From brain to behaviour* (pp. 11–39). Oxford, UK: Oxford University Press.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25. doi:10.1080/0033558008248231
- R Development Core Team. (2007). *R: A language and environment for statistical computing [Software]*. R Foundation for Statistical Computing, Vienna, Austria Retrieved from <http://www.R-project.org/>
- Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41. doi:10.1016/j.cogpsych.2004.11.001
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.
- Roberts, A. C., Robbins, T. W., & Everitt, B. J. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *The Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, *40*, 321–341.
- Roberts, W. A., & Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 247–260. doi:10.1037/0097-7403.14.3.247
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 339–349. doi:10.1111/j.1467-7687.2008.00786.x
- Smithson, M. (2003). *Confidence intervals*. Thousand Oaks, CA: Sage Publications, Inc.
- Soto, F. A., & Wasserman, E. A. (2010). Error-driven learning in visual categorization and object recognition: A common elements model. *Psychological Review*, *117*, 349–381. doi:10.1037/a0018695
- Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, *103*, 284–297. doi:10.1037/h0036872
- Wasserman, E. A., & Bhatt, R. S. (1992). Conceptualization of natural and artificial stimuli by pigeons. In W. K. Honig & J. G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 203–223). Hillsdale, NJ: Erlbaum.
- Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeon's discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, *63*, 165–174. doi:10.1901/jeab.1995.63-165

- Wright, A. A. (1997). Concept learning and learning strategies. *Psychological Science*, 8, 119–123. doi:10.1111/j.1467-9280.1997.tb00693.x
- Yorzinski, J. L., Patricelli, G. L., Babcock, J. S., Pearson, J. M., & Platt, M. L. (2013). Through their eyes: Selective attention in peahens during courtship. *Journal of Experimental Biology*, 216, 3035–3046. doi:10.1242/jeb.087338
- Young, M. E., & Wasserman, E. A. (1997). Entropy detection by pigeons: Response to mixed visual displays after same-different discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 157–170. doi:10.1037/0097-7403.23.2.157

Received September 6, 2013

Revision received February 9, 2014

Accepted February 10, 2014 ■