Learning & Behavior
https://doi.org/10.3758/s13420-024-00623-6

No evidence of real-world equivalence in chickens (*Gallus gallus domesticus*) categorizing visually diverse images of natural stimuli presented on LCD monitors

Jad Nasrini[1] · Robert R. Hampton[1]

Accepted: 3 January 2024

Abstract
Category learning is often tested with similar images that have no significance outside of the experiment for the subjects. By contrast, in nature animals often need to generalize a behavioral response like "eat" across visually distinct stimuli, such as spiders and seeds. Forming functional categories like "food" and "predator" may require conceptual rather than purely perceptual generalization. We trained free-range chickens to classify images assigned to one of four categories based on putative functional significance: inanimate objects, predators, food, and non-competing vertebrates. Images were visually diverse within each category, discouraging classification by perceptual similarity alone. In Experiment 1, chickens classified 80 images into four categories. Chickens then generalized to 80 new exemplars in each of three successive generalization tests. In Experiment 2, chickens saw new types of images to test whether their generalization was perceptual or functional. For example, chickens saw images of skunks for the predator category after training with images of hawks and snakes. Chickens used the "predator" response with these new images for both predators and non-threatening vertebrates, but not for objects or food, and did not successfully generalize any category other than predator. In Experiment 3, chickens categorized fractals as "food," and three of four chickens categorized a range of vertebrates they had not previously encountered as "predators," suggesting that chickens did not see the images as representing real world objects and animals. These results highlight constraints on the use of computer-generated images to assess categorization of natural stimuli in chickens.

Keywords Categorization · Discrimination learning · Avian cognition · Computerized animal testing

Introduction

In order to react appropriately to the visual environment, a chicken may need to discriminate between visually similar stimuli. A snake and a worm share many visual characteristics, but for a chicken one is a threat while the other is a prized food item. Conversely, a worm and a sunflower seed are visually distinct but both call for the same approach-and-consume response. The extent to which birds form abstract concepts like "predator" and "food" to categorize their natural environment has been subject to investigation and debate in the literature. Some reviews suggest that the conceptual classification abilities of non-human animals are frequently underestimated (e.g., Zentall et al., 2008), while others argue that we cannot determine whether animals rely on concepts to categorize (e.g., Chater & Heyes, 1994).

Birds categorize many types of arbitrary visual stimuli. Pigeons, for instance, learn to categorize images based on whether they contain humans (Herrnstein & Loveland, 1964), and based on whether the images depict benign or malignant human breast tissue (Levenson et al., 2015). The ability of pigeons to accurately categorize images of human tissue suggests that real-world equivalence and function are not necessary to categorize visually complex stimuli, because pigeons have no real-world experience with human breast cancer tissue. Pigeons can learn many category discriminations (Watanabe et al., 1993), and can even generalize learning within a category after a single trial (Bhatt et al., 1988). These impressive feats of visual categorization may not rely on what the images in each category represent functionally, but rather on perceptual similarity within categories

✉ Jad Nasrini
jad.nasrini@emory.edu

[1] Department of Psychology and Emory National Primate Research Center, Emory University, 36 Eagle Row, Atlanta, GA 30322, USA

(D'Amato & Van Sant, 1988). Some authors have therefore concluded that we cannot determine whether animals categorize in the same sense that humans do because without language, we cannot know what animals use to form categories (Chater & Heyes, 1994).

In contrast to the case for purely perceptual categorization, birds have been shown to group stimuli based on function in some contexts. In one study, pigeons were shown two stimuli (A and B) in succession followed by reinforcement. They were then trained to choose a particular response whenever stimulus A was presented. Finally, when tested with stimulus B, they chose the trained response that was previously associated with A more than would be predicted by chance (Zentall et al., 2003). This study and other evidence (reviewed by Zentall, 2006) suggest that pigeons can spontaneously learn to treat previously unassociated and distinct stimuli as functionally equivalent. A chicken may similarly treat two food items as functionally equivalent, even if they are visually distinct.

Birds have also been shown to learn differently about some ecologically relevant stimuli, like predators. Blackbirds, for instance, learn to mob an otherwise harmless bird or a plastic bottle, but their response to the plastic bottle is weaker and does not persist as long as their response to the harmless bird (Curio et al., 1978). Pigeons also tend to over-generalize learned categories if they are associated with a fear response, a bias that might be appropriate when failing to respond with avoidance could be fatal (McLaren, 1994). These findings suggest that learning about stimuli in a predator category may be different because of the functional significance of that category.

If birds can link stimuli based on their functional outcome, and they attend to characteristics that signify ecologically relevant categories, some semblance of spontaneous functional categorization may be possible in birds in cases where the category is visually diverse, but all exemplars elicit the same behavioral response in nature. A worm and a piece of fruit may therefore be categorized together not because they appear similar, but because they both cause a chicken to peck when encountered in the real world.

We designed the current study to test whether domestic chickens categorize visually diverse computer images with respect to real-world functional significance. Using a four-choice categorization task (as described by Wasserman & Astley, 1994), we introduced chickens to four categories of natural stimuli based on their putative functional significance: inanimate objects, predators, food, and non-competing vertebrates. We used visually similar images in different categories and visually distinct images within the same category to highlight the functional significance of images over perceptual features as the basis for categorization. For example, worms were in the "food" category and snakes were in the "predators" category despite sharing visual similarities, whereas fruit and ants are visually distinct but are both in the "food" category. If chickens categorize things in nature in part based on their functional properties, they may also learn to categorize visually diverse images of inanimate objects, predators, food items, and non-competing vertebrates, because each of those categories would be responded to differently in the real world.

Subjects and materials

Subjects and testing environment

Four adult free-range female chickens (*Gallus gallus domesticus*) with ad libitum access to food and water were used in this study. The testing apparatuses used for the study were introduced to the chickens' home environment, where they have been housed since they were purchased as chicks. Chickens were housed in an outdoor coop at night for protection from predators but foraged in a large yard after sunrise each day. They were supplied with chicken feed each morning and were free to forage in the open yard during the day with ad libitum access to the testing apparatuses.

Chickens were fitted with leg-bands with attached radio frequency identification tags (Fig. 1; GiS mbH, Lenningen Germany) and had free access to the testing apparatuses during daylight hours on most days. Apparatuses consisted of a touchscreen (Elo Touch Solutions, Milpitas, CA, USA)

Fig. 1 Panel A: Sample leg bands with attached Radio Frequency Identification (RFID) tags. The arrow indicates an RFID tag. Panel B: A chicken fitted with an RFID leg band. The arrow indicates the leg band on the chicken's leg, which is scanned using an antenna on the ground to recognize the subject and administer the correct trial

connected to a laptop computer, an RFID antenna (GiS mbH, Lenningen Germany), and an automated food dispenser (Med Associates, Fairfax, VT, USA). Chickens initiated trials by stepping into a ground-level RFID antenna which allowed the computer to determine the appropriate next trial for each bird and for data to be associated with specific birds. On correct trials, chickens were provided a single fruit-flavored primate food pellet (TestDiet, Richmond, IN, USA). The subjects used in this study were naïve to cognitive testing prior to this study except that they had been trained to peck images on the computer screen for a food reward.

Statistical analyses

Analyses were conducted using SPSS version 27. Wherever we used accuracy scores, correct response proportions were calculated for each chicken separately then arcsine-transformed using the formula (2*arcsine(sqrt(proportion)) prior to analysis (Aron & Aron, 1999). We then compared the mean of arcsine-transformed proportions to accuracy predicted by chance. In cases where we tested the distribution of all four responses, or when a t-test was not appropriate, we used a separate goodness of fit chi-square test for each chicken based on the counts of responses.

Experiment 1: Chickens learned to classify visually diverse computer images into four categories

To determine whether chickens could learn to classify visually diverse stimuli into four categories, we first trained them on a set of 20 images in each category. We then introduced additional images depicting the same types of exemplars at three points during the experiment to test generalization and to increase the visual diversity within each category (as described by Wasserman & Astley, 1994). If chickens learn to group these visually diverse images, then they may be able to rely on non-visual characteristics, such as putative functional significance, to indicate category membership.

Procedure

Chickens were trained with images assigned to four putative functional categories: inanimate objects, predators, food, and non-competing vertebrates, with one category introduced at a time, in that order. Initially, each category contained a training set of 20 visually diverse images chosen to encourage functional categorization by including a variety of outlines and backdrops (Table 1, example images shown in Fig. 4). The number of exemplars within each member of each category was not held constant in experiment one, because we selected fewer predators that we thought the chickens were most likely to recognize, while optimizing for visual diversity in the other categories. Images for all experiments in this study were sourced using the category member names outlined in Table 1 as search terms in one of several image search engines. In some cases, we used more specific species names as search terms like “Cooper’s hawk” or “Red-tailed hawk,” but we do not identify species here because we cannot be positive that the pictures we used were in fact a particular species. The full stimulus set is available from the corresponding author upon request.

Each chicken was first presented with 20 trials using all 20 images in the object category, and only the object category response was available at test. This ensured they were rewarded on each trial. Next, 20 images from the predator category were added and object and predator trials were interleaved in a pseudo-random fashion such that each block of four trials included two object and two predator tests with both response options available. Trials with these two categories were presented to each chicken until it met a criterion of 80% correct in a session of trials including all 40 images.

Table 1 Types of category exemplars used for training and generalization in Experiments 1 and 2

Category	Experiment 1 and training in Experiment 2	Generalization in Experiment 2
Objects	Briefcase, chair, clock, crayons, cups, lamp, oil barrel, shed, steps, wood, table, tools, watering can	Garden rocks, car, plates, bricks
Predators	Coyote, hawk, raccoon, snake	Skunk, opossum, weasel, cat
Food	Ant eggs, clover, compost, cracked corn, earthworm, fruit, grasshopper, slug, sunflower seeds	Beetle, stink bug, ladybug, leafy greens
Non-competing Vertebrates	Beaver, bluebird, cedar waxwing, chickadee, eastern phoebe, rabbit, robin, tortoise, eastern towhee, woodpecker, Carolina wren	Heron, duck, sheep, armadillo

The same types of exemplars were used for training and generalization in Experiment 1, such that each category had the same number of snake images, for example. Those same types were used for training in Experiment 2, but new types of exemplars were introduced at generalization in Experiment 2, such as by introducing images of opossums, which were not previously presented in the predator category

At that point, 20 images from the food category were added counterbalanced in blocks of six trials, and three-category trials were administered until the same 80% correct criterion was met, at which point the final category of non-competing vertebrates was introduced, counterbalanced in blocks of eight trials. The initial training set therefore contained 80 images divided evenly among the four categories, and the categories were introduced one at a time until chickens met a criterion of 80% correct across all four categories in a single session of 80 trials before moving on the generalization phase.

In the generalization phase, three successive generalization tests were conducted, each involving 20 new images in each category. Generalization trials were intermixed randomly with trials using the previously trained images, and trials were counterbalanced in blocks of eight trials containing two trials from each category regardless of whether they were from the training or generalization set. Chickens therefore had to meet an accuracy criterion across 80 images combined before the first generalization test was presented, across 160 images before the second generalization test, and across 240 images before the third. Generalization images depicted the same types of exemplars within each category as in the training set (Fig. 4), but were novel. For example, after training on images of hawks as predators, generalization trials in Experiment 1 presented new images of hawks. Irrespective of whether the images were new or previously trained, chickens were only rewarded if they categorized correctly. Only the first administration of each new image was used to calculate generalization accuracy, preventing new learning from contaminating this measure of generalization. While all birds met the 80% criterion with the first set of images, some birds began to struggle after the first generalization test when the image set was expanded to 160 images. As a result, we lowered the criterion to 70% going forward.

Chickens initiated trials by placing their right leg into an antenna that read the RFID chip on their leg band. The apparatus would then present the chicken with a sample image that remained on the screen until the chicken pecked the image twice or left the apparatus. If the sample image was pecked twice, the sample would dim, and category symbols would appear at the corners of the screen. Pecking the correct category symbol twice automatically dispensed a food reward and ended the trial. Pecking an incorrect category symbol ended the trial without a food reward (Fig. 2). There was no programmed timeout following errors, and no explicit delay between trials, but chickens had to make the computer scan their RFID chip to initiate each trial. This required placing their right leg within an antenna on the ground in front of the apparatus. Because the time subjects took to reposition their legs correctly and rescan their RFID chip between trials varied, and chickens could leave one testing apparatus for another between trials to continue testing immediately, we did not use a timeout following incorrect trials. After several training sessions, but before the first generalization test, a correction procedure was implemented whereby an incorrect trial would be repeated once with all response options available, and a second time with only the correct category response shown if the second attempt was also incorrect. The correction procedure was removed right before the first generalization test for the remainder of the study.

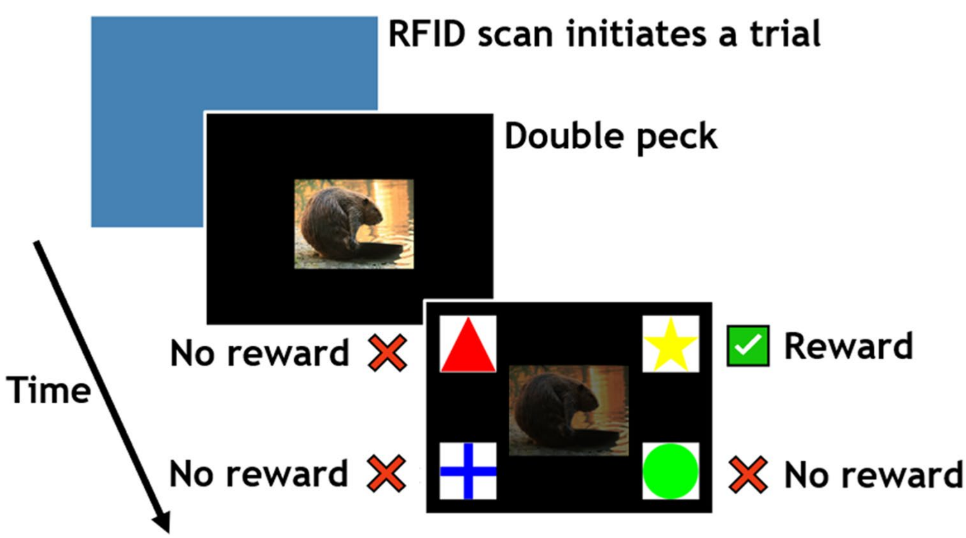

Fig. 2 Sequence on a four-category trial. The subject stepped up to the testing apparatus displaying a blue screen, and the subject's RFID tag was scanned. A sample image was then displayed until the subject pecked on it twice, at which point the image dimmed and four category responses were presented. Only the correct category response was rewarded with a food pellet

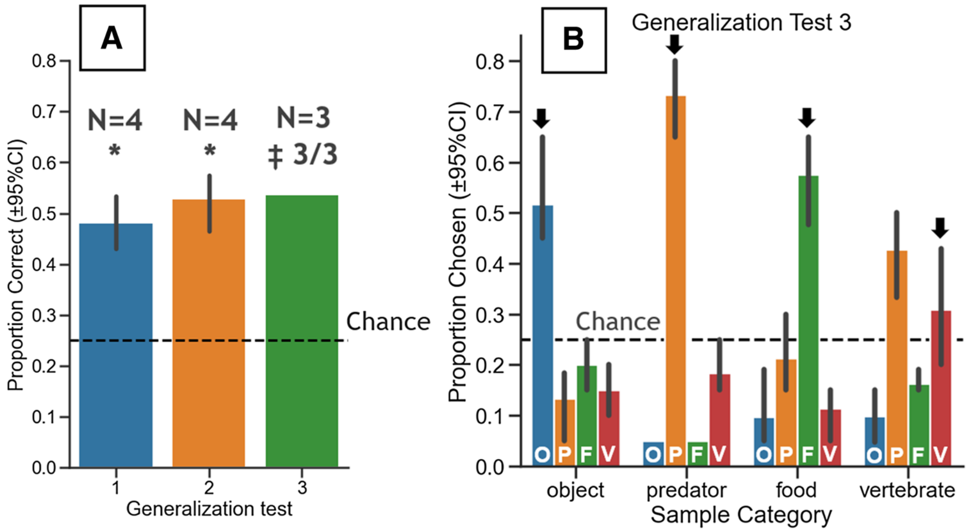

Fig. 3 Panel **A**: Average proportion correct on the three successive generalization tests in Experiment 1. Birds successfully generalized across all three tests in Experiment 1. Panel **B**: Proportion of responses given by the three chickens on the third generalization test. Chickens chose the correct response most frequently on all but the vertebrate category, where predator was the most frequent response. One of the four chickens never reached criterion with the training set of 240 images, so that chicken was never tested on generalization test 3. A breakdown of response categories on generalization tests 1 and 2 is similar to panel B is provided in Online Supplemental Material Fig. 2. In Panel **A**, * Refers to a significant one-sample t-test with α=.05. ‡ Refers to the number of Chi-squared tests that indicated response proportions significantly different from chance performance out of three (one test was conducted for each bird) because a t-test was not appropriate. In Panel **B**, arrows point to the correct response on each sample category (e.g., responding with the "object" category label to new exemplars in the object category). Initials on the bars refer to the response categories: O: Object; P: Predator; F: Food; V: Non-competing vertebrate

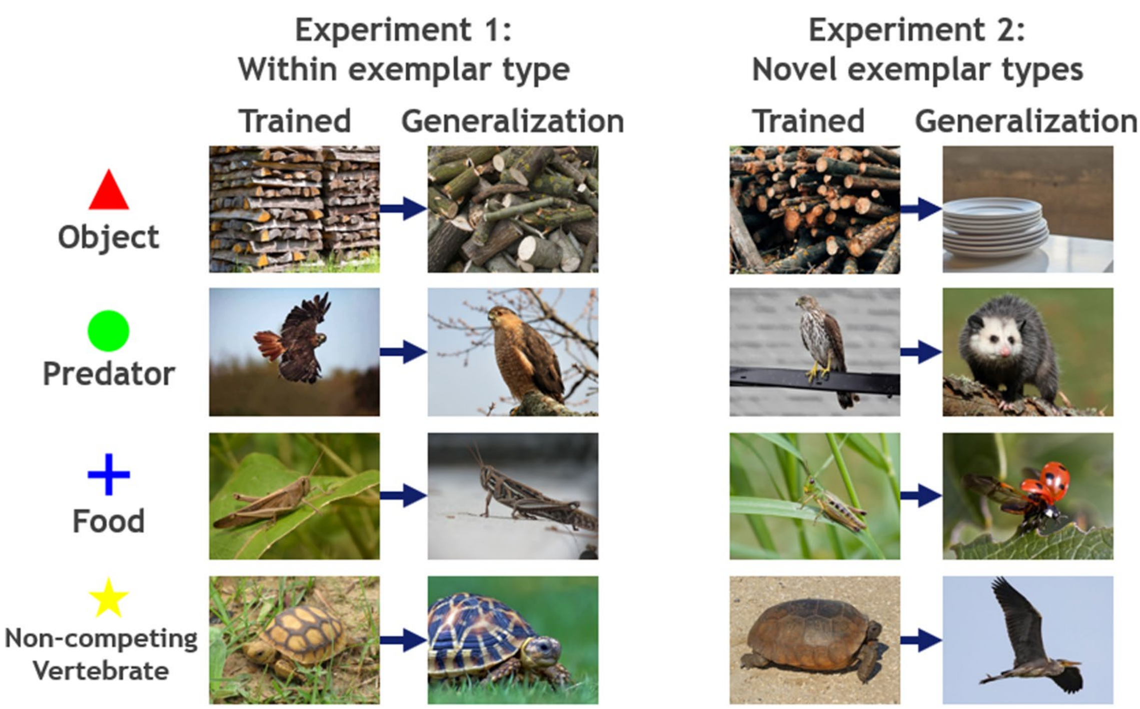

Fig. 4 Sample training and generalization images used in Experiments 1 and 2. In Experiment 1, the images added for each generalization test depicted the same types of exemplars trained previously. In Experiment 2, probe images were added that depicted new types of exemplars

Results and discussion

Chickens learned to classify visually diverse images into the categories we created. All four chickens met the training criterion on the original 80-image set across the four categories (mean trials to criterion before generalization test 1 = 29,323.75 including correction trials, before generalization test 2 = 23,795.50; before generalization test 3 = 25,443.33. Training data from the initial training set are shown in Online Supplementary Material (OSM) Fig. 1). Over the course of Experiment 1, the four chickens completed an average of 338 trials per day each (SD = 105.70 trials) excluding days with no trials completed, with a maximum of 1,970 trials completed in one day by a single chicken.

On the first and second generalization tests, the four chickens collectively categorized the new exemplars correctly on the first exposure more often than would be predicted by chance (chance is 25% correct; Panel A of Fig. 3, first generalization test $t(3) = 8.083$, p = 0.004; second generalization test $t(3) = 8.493$, p = 0.003). One of the four chickens did not meet the accuracy criterion for the set of 240 images after the second generalization test before the end of the experiment, so the third generalization test was performed by only three chickens. Because two of those three chickens achieved identical accuracy on the third generalization test, there was insufficient variance for a t-test to be appropriate to test whether their accuracy was different from chance. Instead, we performed a Chi-squared test for each chicken independently, based on the observed frequency of correct responses across the categories compared to that expected by chance (25%, or 20/80 correct responses). All three chickens were more accurate than would be predicted by chance ($X^2(1, N = 80) = 38.400$, 35.267; 35.267; all $p < .001$). A similar Chi-squared test for the first two generalization tests with all four chickens is reported in OSM Table 1, and those individual analyses produce the same result as the group analysis. Notably, despite generalizing better than would be expected by chance overall, chickens often misclassified non-competing vertebrates as "predators," and by the third generalization test, "predator" was the most common response on vertebrate generalization trials (Panel B of Fig. 3, OSM Fig. 2).

Experiment 2: Chickens generalized to new types of exemplars only on the "predator" category

The generalization results from Experiment 1 show that chickens learned to classify the visually diverse types of exemplars into four categories. Accurate performance in Experiment 1 cannot be explained by memorization of individual images because the birds generalized on first exposure to novel exemplars. However, this generalization performance does not demonstrate that the birds categorized on the basis of the functional properties of the depicted objects. Generalization could have been achieved on perceptual similarity alone, because the new exemplars were all pictures of objects or animals on which the birds had been trained. That is, they were novel pictures of worms or red-tailed hawks, but they were still pictures of worms and hawks. Similar accuracy could be achieved if chickens learned multiple sub-categories based on each *type* of exemplar and used the same category response button for multiple sub-categories. For example, a chicken could have learned a perceptual category of the visual features of worms, and another of fruits, and learned to provide the same "food" response for either of these perceptual categories without relying on the commonality of a behavioral response to a food item. To address this possibility, we designed Experiment 2 with generalization tests that maintained the same types of exemplars within each category at training but introduced new types of exemplars in the generalization phase. If chickens grouped images into categories based on functional significance of what the images depicted, then they should generalize the category responses to new exemplars with the same function, even if they look quite different from the images used in training.

Procedure

The training procedure for Experiment 2 was similar to Experiment 1, except that the chickens started the experiment with all four categories present from the beginning. A new set of 80 images that depicted the same types of exemplars from the previous experiment was used. The first trials with each of these new images therefore constituted an additional generalization test, albeit without the intermixed training trials. After the chickens met an accuracy criterion of 70% correct on a session containing all 80 images across the four categories, a probe set of 80 new images was added intermixed with regular training trials. Unlike generalization tests in Experiment 1, these new probe images depicted new types of exemplars within each category. For example, subjects were trained and tested with images of coyotes and hawks in Experiment 1, whereas in Experiment 2 they were trained with images of coyotes and hawks, then tested for generalization with images of opossums and cats (Fig. 4). Trials with the new exemplars were always reinforced regardless of how the chickens categorized, so that feedback on novel items does not contaminate categorization of the other items in the same category. Table 1 lists the new types of exemplars that were introduced to each category in this experiment.

Results and discussion

On the first trial with each new image depicting the same types of exemplars trained in Experiment 1, chickens generalized more than would be expected by chance on all categories except non-threatening vertebrates, where "predator" was the most common response instead, similar to the third generalization test in Experiment 1 (Fig. 5, objects: $t(3) = 3.77$, $p = .03$; predator $t(3) = 6.17$, $p = .01$; food: $t(3) = 4.68$, $p = .02$; vertebrates: $t(3) = 1.23$, $p = .30$). Including those first trials, mean trials to criterion in Experiment 2 was 6,420.25, substantially quicker than training before generalization tests 2 and 3 in Experiment 1, likely due to the smaller overall stimulus set size in this experiment compared to the end of Experiment 1. The effect of set size on learning rate may suggest that chickens memorized responses to some individual images despite being able to generalize to new exemplars in some cases.

On generalization trials where the images depicted new types of exemplars for each category, chickens correctly categorized predator images more often than would be expected by chance (Fig. 6, $t(3) = 7.09$, $p = .01$), but did not do so for any of the other three categories (Fig. 6, objects: $t(3) = 0.80$, $p = .48$; food: $t(3) = .84$, $p = .46$; vertebrates: $t(3) = -0.73$, $p = .54$).

Notably, "predator" was the most common response to new non-competing vertebrate images but not for new exemplar types of food or objects. Thus, better than chance accuracy with predator images and failure with object and food images cannot be explained solely as a bias to use the "predator" response. Chickens used the predator response similarly in the last generalization test in Experiment 1 and in the first training session of this experiment, which depicted the same types of exemplars as those used in training and testing in Experiment 1. One possibility is that chickens treat all vertebrate animals as potential predators. Indeed, this might be an adaptive conservatism, particularly given we are not certain they have encountered all the species used. Pigeons are reported to over-generalize categories associated with a fear response (McLaren, 1994). Alternatively, the chickens may have learned a perceptual category based on visual characteristics that are common to the exemplars we used in both the predator and non-competing vertebrate categories. If so, it is notable that they did not learn a similar perceptual category for either objects or food, because they generalized to new images depicting the same category members, but not to new images depicting newly introduced members to each category. We attempt to adjudicate these possibilities in Experiment 3.

Experiment 3: Chickens categorized fractals as "food," and most categorized all novel vertebrates as "predators"

The images used in the "predator" and "non-competing vertebrate" categories in both previous experiments contained a mixture of animals that we would expect our chickens have had firsthand experience with, and some they might not. The chickens' use of the predator category for all vertebrates more than would be predicted by chance could therefore be driven either by shared visual characteristics, or by an adaptively

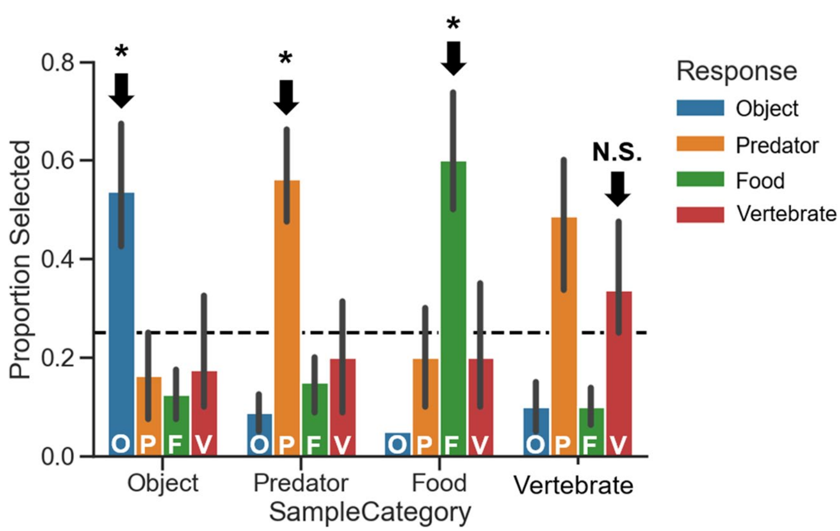

Fig. 5 Proportion of responses given by the chickens on the first trial with each image in the training phase of Experiment 2, which is equivalent to the generalization tests of Experiment 1, but without the intermixed training trials. Arrows point to the correct response on each sample category (e.g., responding with the "object" category label to novel exemplars in the object category). Significance indicators (* for significant, N.S. for non-significant) refer to a one-sample t-test with $\alpha = .05$. Error bars represent the 95% confidence interval around each mean. The dashed line indicates chance. Initials on the bars refer to the response categories: O: Object; P: Predator; F: Food; V: Non-competing vertebrate

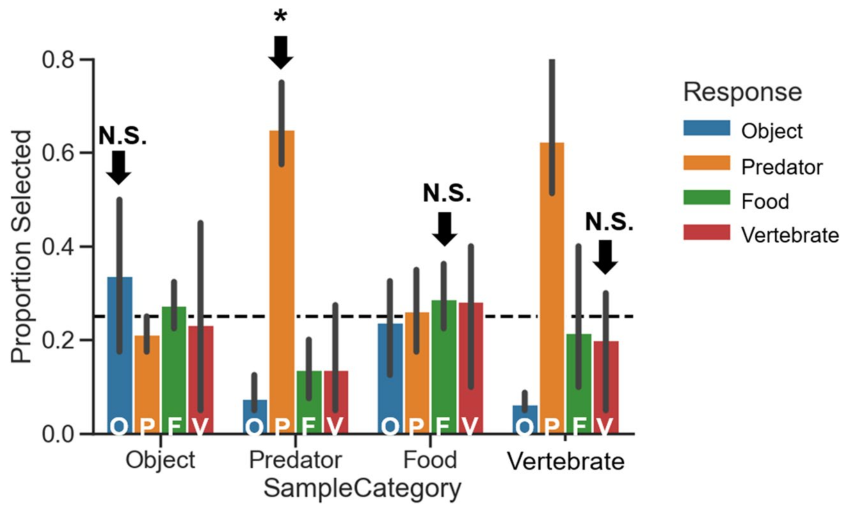

Fig. 6 Proportion of responses given by the chickens to each category of probe images in the generalization phase of Experiment 2. Arrows point to the correct response on each sample category (e.g., responding with the "object" category label to novel exemplars in the object category). Significance indicators (* for significant, N.S. for non-significant) refer to a one-sample t-test with $\alpha = .05$. Error bars represent the 95% confidence interval around each mean. The dashed line indicates chance. Initials on the bars refer to the response categories: O: Object; P: Predator; F: Food; V: Non-competing vertebrate

conservative functional classification that associates fear with all vertebrates with which the chickens have had firsthand experience. To disambiguate these possibilities, we designed Experiment 3 to introduce images of vertebrate animals that the chickens would never have seen in their environment. Additionally, to test whether the chickens had developed any biases for any of the responses, we also introduced fractal images as probes that should be impossible to categorize.

If the chickens relied on their firsthand experience with animals in their environment to group together all vertebrates, then they should not generalize that response to these new vertebrates with which they have never had firsthand experience. Additionally, if their categorization relies on what each image depicts in the real world, they should not know how to categorize the fractals and respond randomly.

Procedure

The training procedure for Experiment 3 was identical to that of Experiment 2, and we used the same initial training images that were used in Experiment 2. These images were presented again as an initial training set in this Experiment to ensure that the chickens were still proficient on the originally learned categorization task. After chickens reached an accuracy criterion of 70% on the training set, we introduced 80 new probe images (Fig. 7). Forty images were of bats, elephants, red macaws, ostriches, chameleons, Komodo dragons, penguins, and ibexes, all vertebrates the subjects have never seen before. Forty additional images were computer-generated fractals that varied in color and shape and were intended to be uncategorizable in the sense that they did not depict any of the types of exemplars used throughout this study. Probe trials were always rewarded regardless of response, firstly because there is no "correct" response for the fractal stimuli by definition, and secondly so that feedback on some items in the "all vertebrates" category does not contaminate categorization of the other items in the same category.

Results and discussion

All four chickens categorized fractals as "food" more than would be predicted by chance (Fig. 8, $X^2(3, N{=}40) = 26.4$; 21.6; 83.6; 31.0; all $p<.001$), and three of four chickens categorized the novel all-vertebrates probes as "predators" more than would be predicted by chance (Fig. 8, $X^2(3, N = 40) = 2.5$, $p = .457$; 9.8, $p = .020$, 14.4, $p = .002$, 11.8, $p = .008$). Given that the fractals were so reliably categorized as food despite the fact that they do not represent anything with a real-world equivalent for the chickens, it's likely that the chickens did not rely on what the images represented in their environment, and instead learned perceptual categories that reflect a visual similarity between the food images we presented and the fractals. Although the new probe images introduced in the new all-vertebrates category further increased the visual diversity of vertebrates that the chickens categorized as predators, these images represented vertebrates with which the chickens would not have had any firsthand experience. Combined with the finding from the fractal images, these results suggest that chickens grouped images into categories based on perceptual similarity from the outset. If chickens had relied on the real-world equivalents of images of things like worms, grasshoppers, and fruit to label those images as "food," they would not use the same label

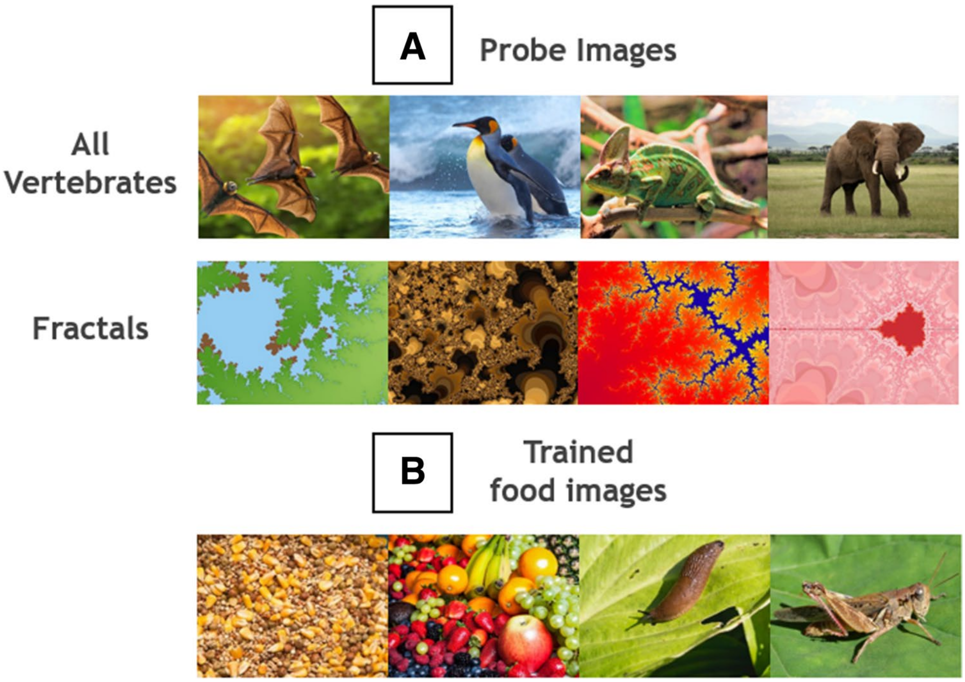

Fig. 7 Panel **A**: example images from the new probe categories introduced in Experiment 3, depicting a wide range of animals never previously encountered by the chickens, and computer-generated fractals. Panel **B**: a few examples of trained images from the food category are shown to demonstrate that the fractals may have mimicked some of the fine-detailed features typical of some but not all types of exemplars in the food category

to categorize the fractals they had never previously seen. It is likewise unlikely that the chickens categorized the new images of vertebrates as predators because of a real-world equivalence given that they would not have had firsthand experience with most of the vertebrates depicted in this set. If real-world experience was necessary for categorizing predator images, chickens might have categorized the novel vertebrates randomly. It's still possible that these previously unencountered vertebrates shared enough visual similarities with animals that the chickens had previously encountered, which led the chickens to recognize them as potential predators. While we cannot address that possibility with the current study, their response to the fractal images would not suggest that they were relying on firsthand experience, and their response to the fractals images was even more consistent than that to the new vertebrates.

General discussion

In our study, chickens learned to categorize visually diverse images and generalized that learning to new exemplars successfully, but only when the new exemplars depicted the same types of category members as those presented during training. In Experiment 1 and the training phase of Experiment 2, chickens correctly generalized their category response to objects, predators, and food, but tended to categorize all vertebrates as "predators." When new images were introduced that depicted new types of exemplars in each category, thereby making the images harder to categorize without reference to the functional significance of each category, the chickens only appeared to generalize the "predator" response, and used the predator response for stimuli most closely resembling those that the chickens were previously trained to categorize in a separate non-competing vertebrate category. If chickens have a real tendency to label all vertebrates as predators, this may be adaptive because ignoring a potential predator could be lethal. In Experiment 3, when always-rewarded probe trials were added that depicted presumably uncategorizable fractals, chickens reliably categorized them as food. Three of four chickens also used the predator response on probe trials that depicted a wide range of vertebrates that they had no prior firsthand experience with.

Under the conditions in our study, it's likely that chickens learned the categories based on perceptual features, not function. We gave the subjects a task that would allow

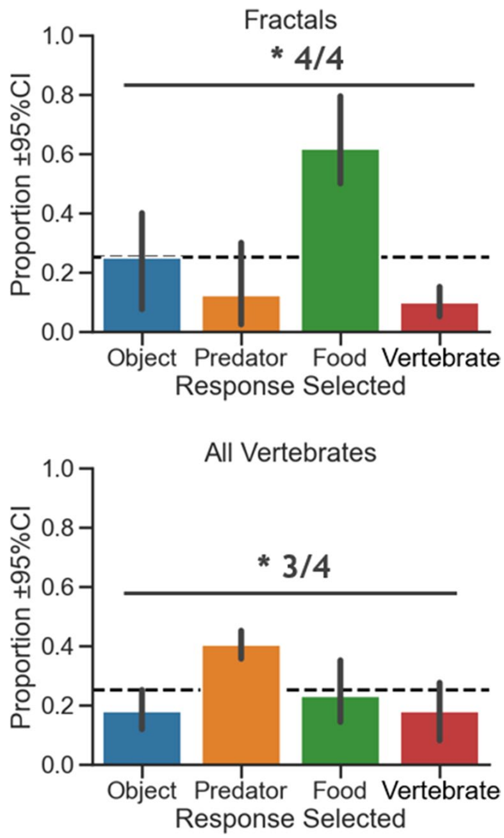

Fig. 8 Average use of the four category responses on the two probe types in Experiment 3. Chickens reliably categorized fractals as food (top panel), and mostly categorized the novel "all vertebrates" probes as predators (bottom panel). Asterisks refer to the number of Chi-squared tests that indicated response proportions significantly different from chance out of four (one test was conducted for each bird)

them to categorize according to either perceptual similarity or functional relevance, and designed the stimulus set to encourage functional relevance above perceptual similarity, but categorization appears to have been controlled by perceptual rather than functional properties. The finding that chickens reliably categorized fractals as food suggests that correspondence between the images and the real world was not relevant for categorization.

We chose the categories in this study based on the behavioral response that would be predicted from a chicken when encountering a member of each category and did not find evidence that chickens relied on the real-world function of the stimuli to categorize them. Nevertheless, it may be that chickens do categorize their environments based on functional significance, but that the specific categories we selected for their putative function are not the ones that chickens spontaneously use. We selected the stimuli across the four categories to introduce visual diversity in the background and foreground of images within each category, as well as some visual similarity between members of different categories, based on our own perception. For example, we purposely included both flying and land predators, as well as flying and land non-competing vertebrates. Despite our attempt, there may have been particular features of the stimulus set that encouraged perceptual rather than functional categorization. Future studies may benefit from using machine vision approaches to investigate the visual characteristics of complex images that allowed the chickens in our study to categorize them without reference to their functional significance. Despite some visual similarity between categories, and considerable visual diversity within each category, the chickens in our study did not appear to rely on the real-world equivalents of the images to categorize them more accurately.

It is notable that the fractal probe images we used in Experiment 3 may have been similar to the repeating finely detailed features in some but not all of the images used in the food category, such as a pile of cracked corn or seed (Fig. 7). If that is the feature that the chickens relied on to categorize the fractals, that would support the idea that they had learned to categorize visually diverse stimuli by simultaneously learning multiple sub-categories for each response option we provided. Nevertheless, we cannot determine based on our data alone whether the fine texture feature was critical for categorizing the fractals as the chickens did, and fine texture is certainly not common to all food items that chickens consume in the real-world.

All four chickens in our experiment tended to over-generalize the initially trained predator category to include all vertebrates, and none of them showed the reverse pattern of responding with the vertebrate label to predators. There is some evidence that birds engage in different forms of learning for predators compared to other stimuli in their environment. In the tradeoff between generalizing and discriminating, responses associated with fear are more likely to be generalized because ignoring them is more costly (McLaren, 1994). It is therefore plausible that our chickens learned to categorize predators differently from the other categories. Unfortunately, we cannot directly address this possibility in our study, because all chickens initially learned the predator category before the non-competing vertebrate category. The longer reinforcement history on the predator category could equally be responsible for the special status of this category. Notably, however, birds learned the inanimate object category first, and this category was not overgeneralized as the predator category was. Even though the initial training criterion was evaluated based on average accuracy across the four categories, the initial training data show that chickens improved on each of the four categories over time as they were introduced (OSM Fig. 1). Despite the vertebrate category being introduced last, chickens gradually achieved better than chance performance on that category before finishing training, so their propensity to label non-competing vertebrates as predators is not because they never learned to discriminate the vertebrates. Especially given the propensity of chickens to categorize fractals as food, further study is necessary to determine whether vertebrate animals or predators are special categories for chickens when learning to discriminate and generalize images.

We did not control the real-life experience of our chickens with any of the natural equivalents of our stimuli. It is possible that the chickens have not had firsthand experience with the real-world function of each of the category members our images depicted. For example, raccoons are primarily nocturnal, and while we have found them at the chicken coop on multiple occasions, we did not track whether these specific birds encountered a raccoon. Nevertheless, these chickens produce alarm calls in response to hawks flying overhead, and approach a wide range of food items to consume them. They also encounter a variety of objects in the yard. Compared to most laboratory animals, these chickens have more experience interacting with diverse foods, objects, and animals in their environment.

Finally, it's important to consider whether there is any correspondence for a chicken between two-dimensional computer images and their real-world equivalents, regardless of categorization. It's possible that chickens do use shared functional significance to classify animals and objects in their environment, but that the images we presented our chickens did not correspond to their experience of anything in the real world and were instead categorized as sets of shapes and colors that have some similarities and dissimilarities. The avian visual system is considerably different from the human visual system and may not respond the same way to computer screens (as reviewed by Weisman & Spetch, 2010). In particular, birds can see color in the ultraviolet range where humans cannot, and birds have a tetrachromatic color space compared to the human trichromatic color space for which computer displays are designed (Cuthill et al., 2000). Viewing images from a short distance can also drive birds to focus on local details instead of global features (Cavoto & Cook, 2006), and we did not control viewing distance in our experiment. Two-dimensional images on a flat computer display may also not convey important depth and size features that may be relevant for recognizing the stimuli depicted in computer images (Dawkins et al., 1996). It is notable however, that previous studies with male chickens have used moving abstract stimuli on standard video displays to elicit the appropriate alarm calls from male chickens based on the location of the moving stimulus (Evans & Marler, 1992), so detailed colored features in particular may have been a key limitation in the recognizability of our stimuli.

Some authors have suggested that pigeons learn to categorize images of natural stimuli more readily than artificial ones (Soto & Wasserman, 2010), and there is some evidence that pigeons transfer learning between real-world objects and high-resolution images (Aust & Huber, 2006; Spetch & Friedman, 2006). Nevertheless, Weisman and Spetch (2010) argue that categorization of ecologically "natural" stimuli is no different from categorization of paintings or tumors unless image to real-world correspondence is demonstrated on a case-by-case basis, because there are many conditions under which that correspondence is absent. The fact that chickens in our study categorized novel abstract fractals as food even more reliably than natural images of food suggests that even though the chickens may have initially *appeared* to rely on functional significance, they likely never relied on recognizing the real-world equivalents of the images.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13420-024-00623-6.

Acknowledgements We acknowledge support from the National Science Foundation (BCS-1632477; BCS-1946767) and from the National Institutes of Health's Office of the Director, Office of Research Infrastructure Programs (P51OD011132). We thank Tara Dove-VanWormer, Tristan Correa, and Jessica Winer for their assistance with data collection.

Open practices statement/Availability of data and materials The datasets, materials, and full stimulus set generated and/or analyzed during the current study are available from the corresponding author on reasonable request, and the experiments in this study were not preregistered.

Funding We acknowledge support from the National Science Foundation (BCS-1632477; BCS-1946767) and from the National Institutes of Health's Office of the Director, Office of Research Infrastructure Programs (P51OD011132).

Declarations

Conflicts of interest The authors have no competing interests to declare.

Ethics approval All procedures involving animals were approved by the Emory University Institutional Animal Care and Use Committee (IACUC).

Consent to participate and consent for publication Not applicable.

References

Aron, A., & Aron, E. (1999). *Statistics for psychology* (2nd ed.). Prentice Hall, Inc. https://psycnet.apa.org/record/1998-06362-000

Aust, U., & Huber, L. (2006). Picture-object recognition in pigeons: Evidence of representational insight in a visual categorization task using a complementary information procedure. *Journal of Experimental Psychology: Animal Behavior Processes, 32*(2), 190–195. https://doi.org/10.1037/0097-7403.32.2.190

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & 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. https://doi.org/10.1037/0097-7403.14.3.219

Cavoto, B. R., & Cook, R. G. (2006). The contribution of monocular depth cues to scene perception by pigeons. *Psychological Science, 17*(7), 628–634. https://doi.org/10.1111/j.1467-9280.2006.01755.x

Chater, N., & Heyes, C. (1994). Animal concepts: Content and discontent. *Mind & Language, 9*(3), 209–246. https://doi.org/10.1111/j.1468-0017.1994.tb00224.x

Curio, E., Ernst, U., & Vieth, W. (1978). Cultural transmission of enemy recognition: One function of mobbing. *Science (New York, N.Y.)*, *202*(4370), 899–901. https://doi.org/10.1126/science.202.4370.899

Cuthill, I. C., Partridge, J. C., Bennett, A. T. D., Church, S. C., Hart, N. S., & Hunt, S. (2000). Ultraviolet vision in birds. In P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, & T. J. Roper (Eds.), *Advances in the study of behavior* (Vol. 29, pp. 159–214). Academic Press. https://doi.org/10.1016/S0065-3454(08)60105-9

D'Amato, M. R., & Van Sant, P. (1988). The person concept in monkeys (Cebus apella). *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 43–55. https://doi.org/10.1037/0097-7403.14.1.43

Dawkins, M. S., Guilford, T., Braithwaite, V. A., & Krebs, J. R. (1996). Discrimination and recognition of photographs of places by homing pigeons. *Behavioural Processes, 36*(1), 27–38. https://doi.org/10.1016/0376-6357(95)00013-5

Evans, C. S., & Marler, P. (1992). Female appearance as a factor in the responsiveness of male chickens during anti-predator behaviour and courtship. *Animal Behaviour, 43*(1), 137–145. https://doi.org/10.1016/S0003-3472(05)80079-X

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science (New York, N.Y.)*, *146*(3643), 549–551. https://doi.org/10.1126/science.146.3643.549

Levenson, R. M., Krupinski, E. A., Navarro, V. M., & Wasserman, E. A. (2015). Pigeons (Columba livia) as trainable observers of pathology and radiology breast cancer images. *PLoS One, 10*(11), e0141357. https://doi.org/10.1371/journal.pone.0141357

McLaren, I. P. L. (1994). Representation development in associative systems. In J. A. Hogan & J. J. Bolhuis (Eds.), *Causal mechanisms of behavioural development* ((1st ed.), pp. 377–402). Cambridge University Press. https://doi.org/10.1017/CBO9780511565120.018.

Soto, F. A., & Wasserman, E. A. (2010). Error-driven learning in visual categorization and object recognition: A common elements model. *Psychological Review, 117*(2), 349–381. https://doi.org/10.1037/a0018695

Spetch, M. L., & Friedman, A. (2006). Pigeons see correspondence between objects and their pictures. *Psychological Science, 17*(11), 966–972. https://doi.org/10.1111/j.1467-9280.2006.01814.x

Wasserman, E. A., & Astley, S. L. (1994). A behavioral analysis of concepts: Its application to pigeons and children. In *The psychology of learning and motivation: Advances in research and theory* (Vol. *31*, pp. 73–132). Academic Press. https://doi.org/10.1016/S0079-7421(08)60409-0

Watanabe, S., Lea, S. E. G., & Dittrich, W. H. (1993). What can we learn from experiments on pigeon concept discrimination? In *Vision, brain, and behavior in birds* (pp. 351–376). The MIT Press.

Weisman, R., & Spetch, M. (2010). Determining when birds perceive correspondence between pictures and objects: A critique. *Comparative Cognition & Behavior Reviews, 5*, 117–131. https://doi.org/10.3819/ccbr.2010.50006

Zentall, T. R. (2006). Functional equivalence in pigeons. *The Behavior Analyst Today, 7*(2), 262–274. https://doi.org/10.1037/h0100080

Zentall, T. R., Clement, T. S., & Weaver, J. E. (2003). Symmetry training in pigeons can produce functional equivalences. *Psychonomic Bulletin & Review, 10*(2), 387–391. https://doi.org/10.3758/BF03196496

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K. R., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews, 3*. https://doi.org/10.3819/ccbr.2008.30002

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