©American Psychological Association, 2026. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: 10.1037/com0000457

Supervision, Category Structure, and Selective Attention in Category Learning: A Comparative Approach

Hyungwook Yim
Department of Cognitive Sciences, Hanyang University, Seoul, Republic of Korea

Leyre Castro
Department of Psychology, University of Iowa, Iowa City, Iowa, USA

Jay I. Myung
Department of Psychology, Ohio State University, Columbus, Ohio, USA

Edward A. Wasserman
Department of Psychology, University of Iowa, Iowa City, Iowa, USA

Vladimir M. Sloutsky
Department of Psychology, Ohio State University, Columbus, Ohio, USA

5352 words

Author Note

The research was supported by the National Institutes of Health Grants R01HD078545 and P01HD080679 to VMS, and the National Research Foundation of Korea funded by the Ministry of Science and ICT (No. RS-2025-24534115) awarded to HY. Part of the study was presented at the 36th Annual Conference of the Cognitive Science Society. We have included an additional experiment that is not reported in the manuscript in the supplementary materials for full disclosure. The preprint of the manuscript can be found at https://osf.io/bt3wg. The data and simulation code in this study can be found on our OSF page (https://osf.io/9dg6t/).
Correspondence concerning this article should be addressed to Vladimir Sloutsky (Department of Psychology, The Ohio State University, Columbus, Ohio, USA. E-mail: sloutsky.1@osu.edu) or Hyungwook Yim (Department of Cognitive Sciences, Hanyang University, Seoul, Korea. E-mail: hwyim@hanyang.ac.kr).

Abstract

We investigated the interaction between supervision and category structure in visual categorization across two species that we assumed to have different abilities to selectively attend. Specifically, we compared humans and pigeons in a category learning task, in which the category structure (dense vs. sparse) and the level of supervision by corrective feedback (low vs. high) were manipulated. Results showed a benefit of supervision across species, which was particularly strong in learning the sparse categories. Moreover, both species benefitted from category structures that had multiple category-relevant dimensions (i.e., dense categories). In addition, humans, who have a more advanced ability to attend selectively, showed faster learning and better generalization overall. Finally, attention was optimized to the category-relevant dimension in the sparse category only in humans. Subsequent computational simulation of the data indicated that these patterns were well explained by a parameter that controls the organism's ability to flexibly switch attention to category-relevant dimensions.

151/250

Keywords: category learning; selective attention; supervision; category structure; comparative cognition

Supervision, Category Structure, and Selective Attention in Category Learning: A Comparative Approach

Categorization requires an organism to perceive and attend to those stimulus features that are common among exemplars of one category and that distinguish this category from others. For example, when learning how to distinguish alpacas from llamas, the shape of the ears is a relevant dimension—alpacas have pointier ears. Focusing on the shape of the ears, while ignoring other dimensions (e.g., four-legged, furry), helps one to distinguish the two camelid family animals that generally look alike. Indeed, when learning a category, both human adults (e.g., Ashby & Ell, 2001; Goldstone & Steyvers, 2001; Nosofsky, 1986; Rehder & Hoffman, 2005) and animals (e.g., Brooks et al., 2013; Castro & Wasserman, 2014, Freedman et al., 2002) tend to focus their attention on the features or dimensions that are relevant for successful categorization. The process of selectively attending to category-relevant dimensions facilitates learning and promotes generalization (Mackintosh, 1965).

The importance of selectively attending to a category-relevant dimension during learning has led major theories and models to incorporate selective attention as a key mechanism in explaining category learning (Kruschke, 1992; Love et al., 2004; Medin & Schaffer, 1978; Nosofsky, 1986; Smith & Minda, 1998). In these models, selective attention is assumed to stretch the scale of the attended dimension in a multi-dimensional psychological space. The stretched dimension becomes more sensitive to stimulus differences, which helps one learn the categories and generalize when new instances are given.

The structure of the category has also been found to affect learning. One way to define category structure is by category density (Gentner, 1981; Kloos & Sloutsky, 2008; Sloutsky,

2010). Category density can be defined as a combination of the similarity among the members in the same category and their distinctiveness from members in other categories. For example, dense categories have multiple inter-correlated features that are relevant for the category, while only a few features are irrelevant. Basic level categories such as ‘dog’ and ‘cat’ can be considered to be dense. For example, cats generally have upright/pointed ears, retractable claws, and move with a soft and quiet gait. However, dogs generally have floppy/cropped ears, non-retractable claws, and move with a heavier/more pronounced gait. Since category-relevant information is redundant, even infants are able to learn dense categories (e.g., Quinn et al., 1993). On the other hand, sparse categories have many features that are shared across different categories, with only few features being relevant to distinguish them. Categories such as ‘llama’ and ‘alpaca’ can be considered to be sparse categories because they have many overlapping features, and only a few features are relevant for differentiating them. Due to their having little category-relevant information, sparse categories can be hard to learn and may require extensive training even for adults (Palmeri et al., 2004; Vong et al., 2019).

Density is a relative concept, which depends on the contrast between the features of the specifically chosen categories. For example, if we now contrast llamas with red pandas, both categories become 'dense' due to their vastly different, non-overlapping physical features. Category density naturally interacts with the ability to attend selectively. When learning sparse categories, the ability to attend to relevant dimensions may be more critical, as little information determines category membership. In contrast, dense categories have multiple intercorrelated features relevant to category membership, supporting learning even

without selective attention (Goldstone, 1994). Studies show that 6-to-8-month-olds, who have less developed selective attention than adults due to their immature prefrontal cortex, struggle to learn sparse categories (Yim et al., 2012). Similarly, comparative studies with pigeons suggest that they may require relatively dense categories for successful learning (Castro et al., 2020; Castro et al., 2026; Wasserman & Castro, 2021).

Supervision is another factor that affects category learning. In supervised learning, the learner is given corrective feedback (e.g., verbal responses like “correct” or “incorrect”, or differential auditory signals indicating a right or wrong answer, or the occurrence/nonoccurrence of food reward in the case of animals) after their choices, whereas in unsupervised learning, the learner develops categories without being provided any differential feedback about the correctness of their choices. The different outcomes resulting from supervised and unsupervised category learning have been well documented in humans (e.g., Colreavy & Lewandowsky, 2008; Love, 2002; Love et al., 2004; Pothos et al., 2011). However, few studies have systematically examined the effect of supervision along with category density and selective attention. Kloos and Sloutsky (2004) found that, when supervision was provided, learning was enhanced for sparse categories in both 4-to-5 year-olds and adults, but the enhancement was even greater for children. As supervision acts as a top-down signal and aids category learning, children with immature selective attention benefited more from supervision than did adults. On the other hand, when learning dense categories, supervision hindered learning in both children and adults. Considering the role of supervision, these results are counterintuitive. However, one thing to note is that Kloos and Sloutsky provided supervision through *verbal instructions* that gave information about the

features that were relevant to category membership (e.g., "*Ziblets have a lightbody*"). This required the participants to maintain the instructions active in working memory, which may increase cognitive load and hinder learning (e.g., Wan & Sloutsky, 2026; Zeithamova & Maddox, 2006). Therefore, it is not clear whether supervision, when provided without increasing cognitive load, will hinder the learning of dense categories.

In the present study, we systematically examined how supervision, category density, and selective attention affect category learning, as well as how these factors interact. We did so in both human adults and pigeons, evolutionarily distant species that may very well differ in their ability to attend selectively. Instead of using explicit verbal instructions, an obvious impossibility with pigeons, we manipulated supervision in terms of the amount of corrective feedback provided. Finally, to more assiduously compare the results of humans and pigeons, we utilized a computational model to simulate the obtained data patterns and to infer the latent parameters that may drive performance differences between the target species.

Experiment 1

Method

In Experiment 1, human adults learned to discriminate two flower-like artificial categories, in which we manipulated category density and level of supervision. Participants were presented with category members one at a time and had to choose whether it belonged to one category or the other. We were especially interested in the number of trials it would take to learn the categories. After participants learned the categories by reaching a set criterion, we introduced a generalization phase. In the generalization phase, participants were presented with new exemplars of the studied

categories, where irrelevant features were substituted with novel features. We expected that the ability to generalize would provide valuable evidence for how well participants had selectively attended to the category relevant dimensions during learning.

Participants

One hundred undergraduate students at The Ohio State University participated for course credit (66 females, *M* = 19.31 years, *SD* = 2.04). Each participant was randomly assigned to one of four conditions created by crossing two independent variables: Density (Dense vs. Sparse) × Supervision (Low vs. High). There were 25 participants in each condition. We did not conduct a power analysis prior to the experiment, and the sample size was based on previous, related experiments. The experiment was approved by the Institutional Review Board at The Ohio State University, Columbus, Ohio, USA.

Materials

Flower-like artificial categories were used in the experiment (see Figure 1A). For the learning phase, there were two sets of categories (i.e., Category A and Category B) for each density condition (i.e., dense and sparse). Every exemplar had a gray hexagon in the middle with six colored shapes on each side. Each of the six locations was considered a dimension. For the dense categories, three of the six dimensions had variable colors and shapes, whereas the other three dimensions had the same shape and color within a category but different from those in the contrast category; the latter three were category relevant dimensions. For example, Category A in the dense condition had three pink triangles in the upper-left, bottom-left, and bottom-right locations, that were the same for all members of Category A, whereas

Category B had three blue semicircles in the same location (marked as 'R' in Figure 1A). Other locations, which involved irrelevant dimensions, changed their features independently in the following fashion: upper-right location (square in yellow, green, or orange), right location (trapezoid in light green, purple, or orange), left location (house-shape in yellow, or green).

For the sparse categories, five of the six dimensions varied in their color and shape, so that they were irrelevant, whereas one was constant, the relevant dimension. For example, one category in the sparse condition (i.e., Category A) had a pink triangle, whereas the other category (i.e., Category B) had a blue semi-circle in the bottom-right location (marked as 'R' in Figure 1A). Other locations, which involved irrelevant dimensions, changed their features independently in the following fashion: upper-right location (square in yellow or orange), right location (trapezoid in light green or purple), bottom-left (trapezoid in light green or purple), left location (house-shape in orange or yellow), upper-left (triangle in pink or blue). Generalization stimuli were similar to those in the learning phase, except that the irrelevant dimensions had new values. Each category set had 18 unique exemplars, and there were two sets each for the learning phase and the generalization phase. As there were two density conditions, the total number of exemplars used in the experiment was 144 (18 exemplars × 2 category sets (Category A/B) × 2 phases (learning/generalization) × 2 density conditions (dense/sparse)).

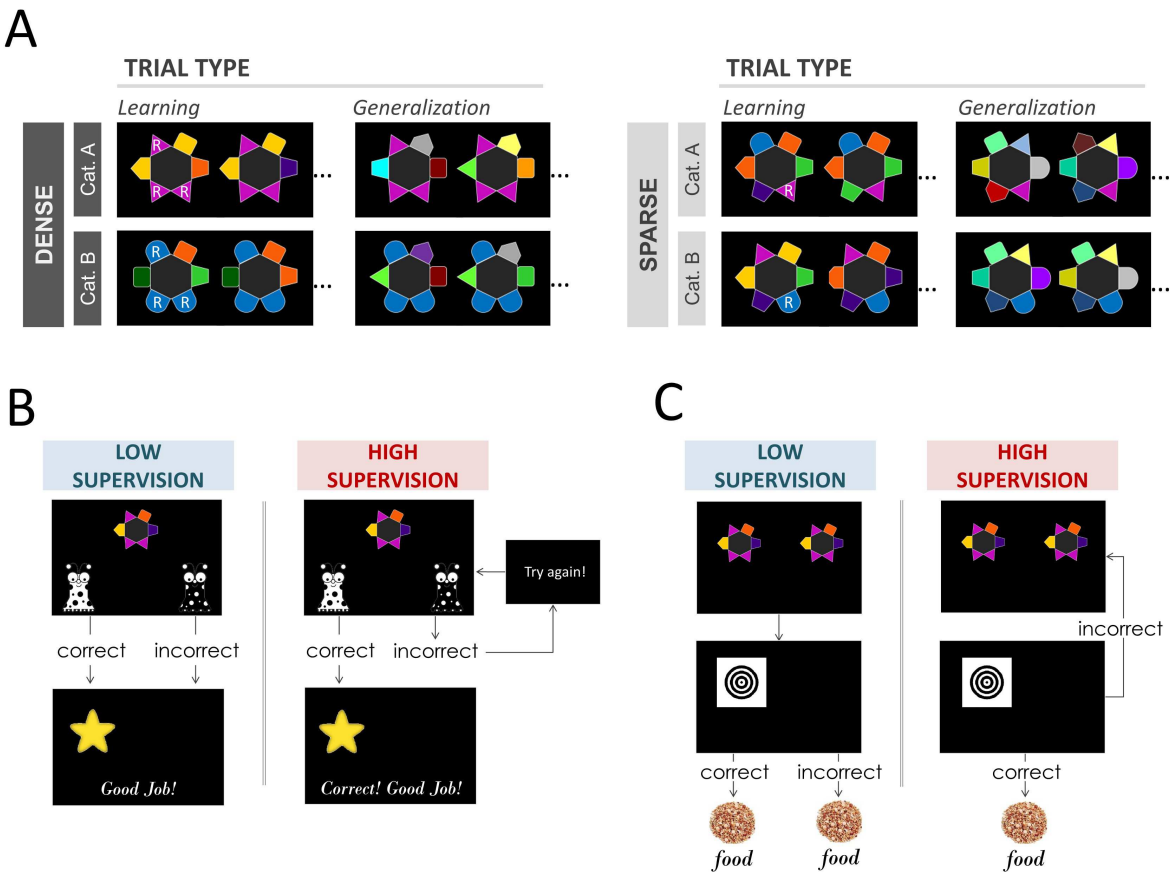

Figure 1. Stimulus structure and experimental design. (A) The dense and sparse stimulus structures used in the learning phase and the generalization phase. The letter 'R' represents the relevant dimension of the category; the letter was not shown to the participants. (B) Experimental procedure in Experiment 1. (C) Experimental procedure in Experiment 2.

Design and procedure

Participants were told they were going to play a game where a cartoon character (i.e., Dora the Explorer) was investigating magic flowers from a different planet, which would turn into black or white aliens when she touched them. Participants were told that their job was to help Dora discover which flowers would turn into black aliens and which flowers would turn into white aliens, so that they should carefully look at the flowers.

There were 12 trial blocks, with 12 learning trials per block. Half of the learning trials

had Category A exemplars, whereas the other half had Category B exemplars, with exemplars presented in a random order. On each trial, an exemplar was presented in the top center of the computer screen, and an image of a white alien and a black alien were each presented in the bottom left and right sides of the screen. Left and right locations were randomized across participants (see Figure 2B). On each trial, participants used a computer mouse to click on the image of a white or black alien to indicate their decision about how the presented flower would change.

In the low-supervision condition, when the participant made a decision, regardless of its accuracy, a star appeared on the correct side for 500 msec[1]. A "Good Job!" sign was also simultaneously presented at the bottom of the screen. In the high-supervision condition, the star was only presented when the participant made a correct response; in addition, if the participant made an incorrect response, then a "Try again!" sign was presented in the center of the screen for 500 msec. Then the participant was presented with the same exemplar and was asked to make another choice.

When a participant obtained more than 10 correct trials in a 12-trial block, a generalization phase was introduced before proceeding to the next block. Trials in the generalization phase were identical to trials in the learning phase except: (1) stimuli from the generalization set were used, (2) there were six trials, and (3) a blank screen was presented for 500 msec after the participant's response, without any feedback. The experiment was

[1] The participants may have used the star to infer the correct answer. However, the results show that the method provided a weaker supervisory signal as participants required more trials to reach the learning criterion and showed lower generalization accuracy. These findings indicate that the manipulation was effective as intended.

terminated either when the participant responded to all 12 blocks or when the participant responded incorrectly for only one trial across two consecutive blocks (i.e., 23 correct responses out of 24 trials). The experiment was controlled by MatLab® with Psychtoolbox-3 extensions (Brainard, 1997; Pelli, 1997; http://psychtoolbox.org/).

Transparency and openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow JARS (Kazak, 2018). All data, analysis code, and research materials are available at https://osf.io/9dg6t, and none of the experiments was preregistered. Data were analyzed using JASP, version 0.19 (JASP Team, 2024) and plotted using matplotlib, version 3.2.1 (Hunter, 2007). This study's design and its analysis were not pre-registered.

Results

We first determined the number of participants who successfully learned the two categories. Successful learners were defined as those who reached an accuracy higher than .85 in each block for two consecutive blocks, and who displayed asymptotic performance after reaching criterion. There were more learners in the dense conditions and in the high-supervision conditions (see Table 1). The number of learners was statistically different for density (Z = 2.49, p = .013, 95% CI [.56, 4.76]), but not for supervision (Z = 1.27, p = .20, 95% CI [-.47, 2.20]; binary logistic regression with Supervision and Density as factors). The non-learners were excluded from all subsequent analyses unless mentioned otherwise.

Table 1.

Proportion of learners in Experiment 1. The numbers in parentheses represent the number of

learners and the total number of participants.

		Sparse	*Dense*
Adults	*Low-Supervision*	**0.72** (18/25)	**0.96** (24/25)
	High-Supervision	**0.84** (21/25)	**1.0** (25/25)

Additionally, we determined how many trials it took participants to learn the two categories (i.e., trials-to-criterion; TTC) across the different conditions (see Figure 2A). TTC was defined identically to how learners were determined. The dense-high-supervision condition showed the fastest learning (M = 38.4, SD = 17.32) followed by the dense-low-supervision condition (M = 44.5, SD = 29.07), the sparse-high-supervision condition (M = 64.57, SD = 30.47), and the sparse-low-supervision condition (M = 78.0, SD = 27.69). A density × supervision between-subjects analysis of variance (ANOVA) revealed a statistically significant main effect for density ($F(1, 84) = 27.79$, $p < .001$, $\eta^2 = .24$, 95% CI of η^2 [.10, .39]), a trend for supervision ($F(1, 84) = 2.98$, $p = .09$, $\eta^2 = .03$, 95% CI of η^2 [.00, .13]), but no sign of an interaction ($F(1, 84) = .42$, $p = .52$, $\eta^2 = .004$).

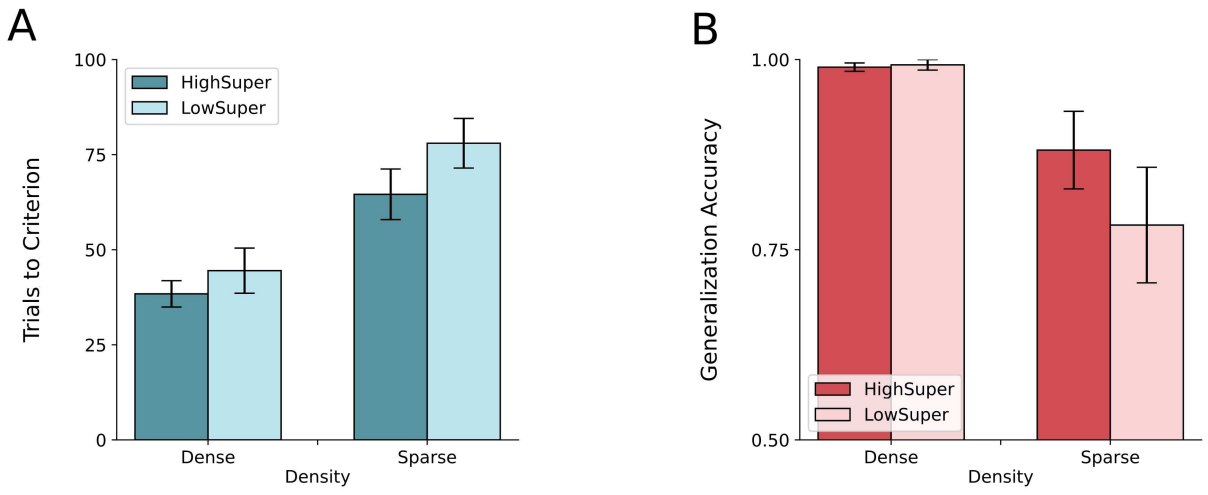

Figure 2. Results from Experiment 1 with adults. (A) Trials-to-criterion, and (B) Generalization accuracy. (*Note that error bars represent +/- 1 SEM*)

For the generalization trials, the highest accuracy was shown in the two dense conditions (high-supervision: M = .99, SD = .028; low-supervision: M = .99, SD = .034), followed by the sparse-high-supervision condition (M = .88, SD = .23), and the sparse-low-supervision condition (M = .78, SD = .32). A density × supervision between-subjects ANOVA disclosed a statistically significant main effect for density ($F(1, 84) = 16.0$, $p < .001$, $\eta^2 = .16$, 95% CI of η^2 [.04, .30]). There was no effect of supervision ($F(1, 84) = 1.43$, $p = .24$, $\eta^2 = .01$, 95% CI of η^2 [.00, .10]), and no interaction ($F(1, 84) = 1.62$, $p = .21$, $\eta^2 = .02$). Generalization performance was above chance (.50) for all conditions (high-super/dense: $t(24) = 88.64$, $p < .001$, Cohen's d = 17.73, 95% CI of d [12.51, 22.58]; high-super/sparse: $t(20) = 7.47$, $p < .001$, Cohen's d = 1.63, 95% CI of d [.96, 2.28]; low-super/dense: $t(23) = 71.00$, $p < .001$, Cohen's d = 14.49, 95% CI of d [10.13, 18.54]; low-super/sparse: $t(17) = 3.72$, $p = .002$, Cohen's d = .88, 95% CI of d [.32, 1.41]).

We excluded the non-learners from the analysis because their TTC could not be measured (as they never reached the criterion), and many did not complete the generalization phase. However, it is possible that the exclusion may introduce bias in interpreting the data, especially for the sparse conditions, which had a higher proportion of non-learners. Therefore, we also report analyses of TTC and generalization performance that include non-learners, in the interest of transparency. Since the non-learners did not have a true TTC value, we used 144, which was the maximum possible trials, and we used .5 for their generalization accuracy. Similar to the previous analysis, the dense-high-supervision condition showed the fastest learning (M = 38.4, SD = 17.32), followed by the dense-low-supervision condition (M = 48.5, SD

= 34.72), the sparse-high-supervision condition (M = 77.28, SD = 40.71), and the sparse-low-supervision condition (M = 96.48, SD = 38.18). A density × supervision between-subjects analysis of variance (ANOVA) revealed a statistically significant main effect for density (F(1, 96) = 40.84, p < .001, η^2 = .29, 95% CI of η^2 [.15, .42]), and supervision (F(1, 96) = 4.64, p = .03, η^2 = .03, 95% CI of η^2 [.00, .13]), but no interaction (F(1, 96) = .45, p = .50, η^2 = .003).

For the generalization trials, the highest accuracy was shown in the two dense conditions (high-supervision: M = .99, SD = .028; low-supervision: M = .97, SD = .10), followed by the sparse-high-supervision condition (M = .82, SD = .26), and the sparse-low-supervision condition (M = .70, SD = .30). A density × supervision between-subjects ANOVA disclosed a statistically significant main effect for density (F(1, 96) = 28.86, p < .001, η^2 = .22, 95% CI of η^2 [.09, .36]), but not for supervision (F(1, 96) = 2.65, p = .11, η^2 = .02, 95% CI of η^2 [.00, .11]) or interaction (F(1, 96) = 1.49, p = .23, η^2 = .012). Overall, we did not find a significant difference in the results when analyzing the data including the non-learners.

Experiment 2

In Experiment 1, humans showed an effect of density and supervision during learning, supported by comparing the number of learners and trials to criterion in each condition. Generalization was affected by category density, but not by supervision. In Experiment 2, we studied the behavior of pigeons, which generally exhibit less robust selective attention (e.g., Castro et al., 2020), in a highly similar category learning task. Pigeons were given the same four tasks as in Experiment 1, except for necessary modifications in the training procedures.

Method

Participants

The subjects were 16 homing pigeons (*Columba livia*) maintained at 85% of their free-feeding weights by controlled daily feedings. The pigeons had served in unrelated studies prior to the present project. The pigeons were randomly assigned to each of the conditions, resulting in four pigeons for each condition. We did not conduct a power analysis prior to the experiment, and the sample size[2] was based on previous, related experiments. All procedures were approved by the Institutional Animal Care and Use Committee at the University of Iowa.

Materials

The experiment used four 36 × 36 × 41 cm operant conditioning chambers detailed by Gibson et al. (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 cm. Pecks to the touch screen were processed by a serial controller board outside the box. A rotary dispenser delivered 45-mg pigeon pellets through a vinyl tube into a food cup located in the center of the rear wall opposite the touchscreen. Illumination during the experimental sessions was provided by a house light mounted on the upper rear wall of the chamber. The pellet dispenser and houselight were controlled by a digital I/O interface board. Each chamber was controlled by its own Windows® computer. The program to run the experiment was developed in MatLab® with Psychtoolbox-3 extensions (Brainard, 1997; Pelli, 1997; http://psychtoolbox.org/).

[2] Although the sample size is typical for comparative research, we acknowledge that the limited statistical power warrants cautious interpretation of the results. We observed robust main effects of Density and Supervision on pigeons' performance measures. There was also a trend toward an interaction between Density and Supervision; however, this interaction did not reach statistical significance for either trials-to-criterion or generalization accuracy. It is possible that these interactions may have reached statistical significance with a larger sample size.

316 ***Design and Procedure***

317 The category exemplars used in the experiment were identical to those used in
318 Experiment 1. During the learning phase, pigeons were given daily sessions containing 144
319 trials. Among the 144 trials, half contained exemplars from Category A and the other half
320 contained exemplars from Category B. On each trial, two identical exemplars were presented
321 on the right and left side of the computer screen (see Figure 1C). In the low-supervision
322 condition, after five pecks to either of these two images, a target stimulus (i.e., black concentric
323 rings on a white background) was presented on the correct side of the screen; the correct side
324 for Category A and Category B was counterbalanced across pigeons. After one peck to this
325 target stimulus, food reward was delivered. All trials ended with food reinforcement, regardless
326 of the specific side that the pigeons chose when the exemplars were presented. Intertrial
327 intervals were 5000 msec in duration. The high-supervision condition was very similar to the
328 low-supervision procedure; however, the last peck to the category exemplar had to be on the
329 correct side for the target stimulus to be presented. If the last peck was on the wrong side, then
330 the screen turned black for 1000 msec and the same category exemplar was again presented.
331 The pigeons had to complete the response requirement again until the correct response was
332 made. Responses from the correction trials were not analyzed. When a bird reached a learning
333 criterion of .85, the generalization phase was introduced the next day and continued for 4 days.
334 The generalization phase consisted of 162 trials a day, which included 144 training trials and 18
335 randomly intermixed testing trials. The testing trials presented the generalization stimuli; on
336 these trials, no correction procedure was given (to avoid teaching the birds the correct
337 response) so that pigeons' choices were always reinforced, regardless of the pigeons being in

the low- or high-supervision conditions.

Results

All birds were trained until reaching the learning criterion, except for one bird in the low-Supervision/sparse condition that did not reach the criterion after 76 days (10,944 trials; see Table 2). This bird was excluded from all following analyses. We did not conduct a statistical analysis for the number of learners as the experiment was intended to provide learning trials until the birds learned the categories.

Table 2.

Proportion of learners in Experiment 2. The numbers in the parentheses represent the number of pigeons that reached the learning criterion and the total number of pigeons.

		Sparse	*Dense*
Pigeons	*Low-Supervision*	**0.75** (3/4)	**1.0** (4/4)
	High-Supervision	**1.0** (4/4)	**1.0** (4/4)

Trials-to-criterion were calculated as in Experiment 1 (see Figure 3A). The two high-supervision conditions exhibited faster learning (dense: M = 288, SD = 0; sparse: M = 612, SD = 72), than the two low-supervision conditions (dense: M = 1836, SD = 1685; sparse: M = 5664, SD = 3169). A density × supervision between-subjects ANOVA disclosed a statistically significant main effect for density ($F(1, 11) = 6.12$, $p = .031$, $\eta^2 = .17$, 95% CI of η^2 [.00, .53]) and supervision ($F(1, 11) = 15.46$, $p = .002$, $\eta^2 = .42$, 95% CI of η^2 [.02, .70]) with a suggestive interaction ($F(1, 11) = 4.36$, $p = .061$, $\eta^2 = .12$, 95% CI of η^2 [.00, .48]). Generalization performance was also higher in the high-supervision conditions (dense: M = .94, SD = .46; sparse: M = .85, SD = .09; see Figure 3B) than in the low-supervision conditions (dense: M = .88, SD = .07; sparse: M = .67, SD = .11)

across all density conditions. A density × supervision between-subjects ANOVA showed a statistically significant main effect for density ($F(1, 11) = 14.03$, $p = .003$, $\eta^2 = .39$, 95% CI of η^2 [.003, .68]), and supervision ($F(1, 11) = 9.14$, $p = .012$, $\eta^2 = .25$, 95% CI of η^2 [.00, .59]) but not an interaction ($F(1, 11) = 2.19$, $p = .17$, $\eta^2 = .06$). Additionally, generalization performance was above chance (.50) for all conditions, except for the low-supervision/sparse condition (high-super/dense: $t(3) = 25.40$, $p < .001$, Cohen's $d = 12.70$, 95% CI of d [3.34, 21.85]; high-super/sparse: $t(3) = 8.15$, $p = .004$, Cohen's $d = 4.07$, , 95% CI of d [.89, 7.32]; low-super/dense: $t(3) = 10.64$, $p = .002$, Cohen's $d = 5.32$, , 95% CI of d [1.27, 9.49]; low-super/sparse: $t(2) = 2.54$, $p = .13$, Cohen's $d = 1.47$, 95% CI of d [-.32, 3.16]). The results should be interpreted with caution due to the relatively small sample size.

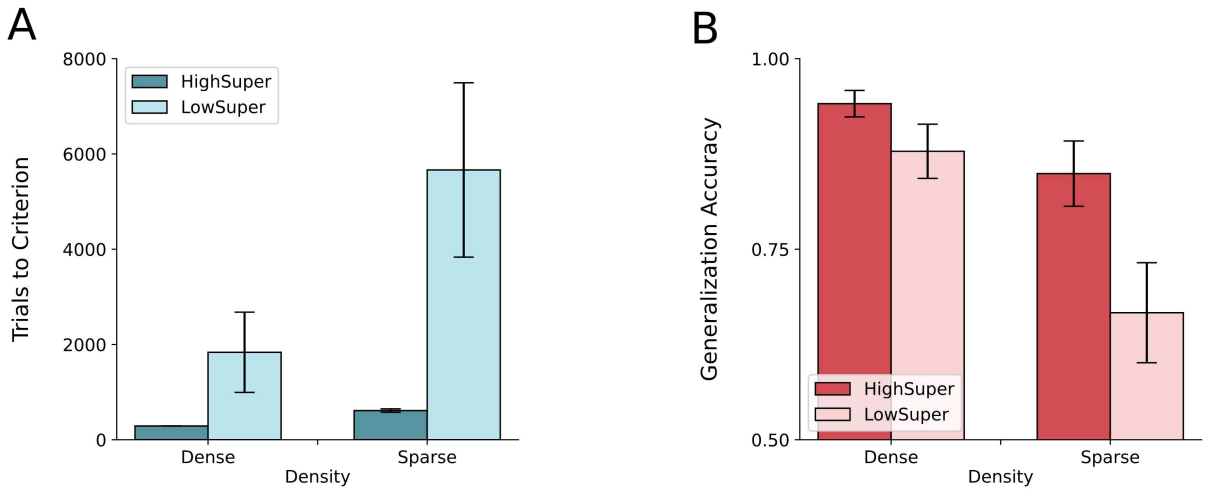

Figure 3. Results from Experiment 2 with pigeons. (A) Trials-to-criterion, and (B) Generalization accuracy. *Note that error bars represent +/- 1 SEM*.

The results evidenced a tendency for an interaction in the trials-to-criterion data, which implies that pigeons benefited more from a stronger supervisory signal in the sparse condition than in the dense condition. During generalization, pigeons evidenced both an effect of density

and supervision, unlike the adults.

Computational Model Simulation

Method

The overall pattern of results across the two species may be explained by differences in the ability to selectively attend to the category relevant information. If selective attention is highly functional, then it should be easy to pinpoint the category relevant information and to ignore the category irrelevant information. As a consequence, category learning will be fast and generalization accuracy will be high, resembling the results observed in humans. On the other hand, if selective attention is less functional, pinpointing the relevant information will be difficult and category learning may have to rely on remembering each exemplar (i.e., rote memory). Also, generalization will not be robust because the overall similarity of the exemplars will have to be used when categorizing the exemplars rather than relying only on the category relevant information, resembling the results found in pigeons.

To further examine this account, we used a computational model of category learning. To do so, we chose ALCOVE (Kruschke, 1992; Lee & Navarro, 2002) to identify which component of the model contributes to the overall pattern that changes across groups. ALCOVE is an exemplar-based network model that represents the exemplars in a multidimensional space and incorporates attentional learning. Importantly, ALCOVE is able to simulate not just an outcome but also the learning process of category learning (c.f., Nosofsky, 1986).

We took a simulation approach where we changed each parameter one at a time and examined the model predictions. Therefore, the simulation results should be interpreted with caution, and should be taken as a proof-of-concept rather than a full mechanistic explanation.

The four main parameters of ALCOVE are: the attention learning rate (λ_α), the output learning rate (λ_w), the decision scaling parameter (ϕ), and the specificity constant (c). Most importantly, the λ_α parameter controls how flexibly the model is able to allocate attention throughout learning; it has been a useful parameter in explaining the results of a wide variety of category learning studies (e.g., Nosofsky & Kruschke, 2002). The λ_w parameter controls how well the model associates each exemplar with the correct responses (i.e., category membership). The ϕ parameter reflects the consistency of choice responses (i.e., temperature parameter in a softmax function). Finally, the *c* parameter controls the activation of the psychological exemplar space and tunes (or scales) the psychological exemplar space. A smaller *c* parameter will activate more exemplars near the encountered exemplar, while a larger *c* parameter will only activate the encountered exemplar in the psychological space.

For example, in a category learning task, a higher λ_α parameter generally leads to faster learning and better generalization, as the ability to selectively attend aids both processes. Similarly, a higher λ_w parameter leads to faster learning. However, because this parameter primarily strengthens the association weights between exemplars and responses, without a high λ_α parameter, learning is more likely achieved through rote memory, resulting in lower generalization. A higher ϕ parameter produces faster learning but lower generalization, as a high ϕ will allow the model to distinguish between the categories even with smaller activation values. Finally, a higher *c* parameter increases the distinctiveness of each exemplar. In this case, accessing an exemplar restricts the spread of activation to other similar exemplars in the psychological space. Consequently, a higher *c* parameter results in slower learning and lower generalization.

The simulation involved the same category structures and design used in the current study; the parameters were fixed to $\boldsymbol{\lambda}_{\alpha}$ = .001, and $\boldsymbol{\lambda}_{w}$ = .001, c = 5, and ϕ = 2.5 when they were not varied in the simulation[3]. We used the catlearn R package (Wills et al., 2017) to implement the simulation, and ran the simulation 100 times for each condition, where the order of the presented exemplars was randomized in each simulation.

Results

In Figure 4, each panel shows the simulation results from changing a single parameter while fixing the other parameters. If the parameter is crucial for explaining differences across groups when increasing the parameter values, then trials to criterion (top row) will decrease and generalization accuracy (bottom row) will increase. Results showed that only $\boldsymbol{\lambda}_{\alpha}$ follows this pattern (Figure 4A), whereas the c parameter shows an opposite pattern (Figure 4C), and the $\boldsymbol{\lambda}_{w}$ and ϕ parameters both result in a decrease in trials to criterion and generalization accuracy (Figure 4B, and 4D).

[3] We were not able to obtain stable estimates of the parameters by fitting the data using MLE and hierarchical Bayesian methods, so instead we used a simulation approach. The values were chosen within a range that were selected from previous studies. There were no qualitative differences when we used different values for the fixed parameters. The simulation code and results can be found in our OSF repository (https://osf.io/9dg6t/).

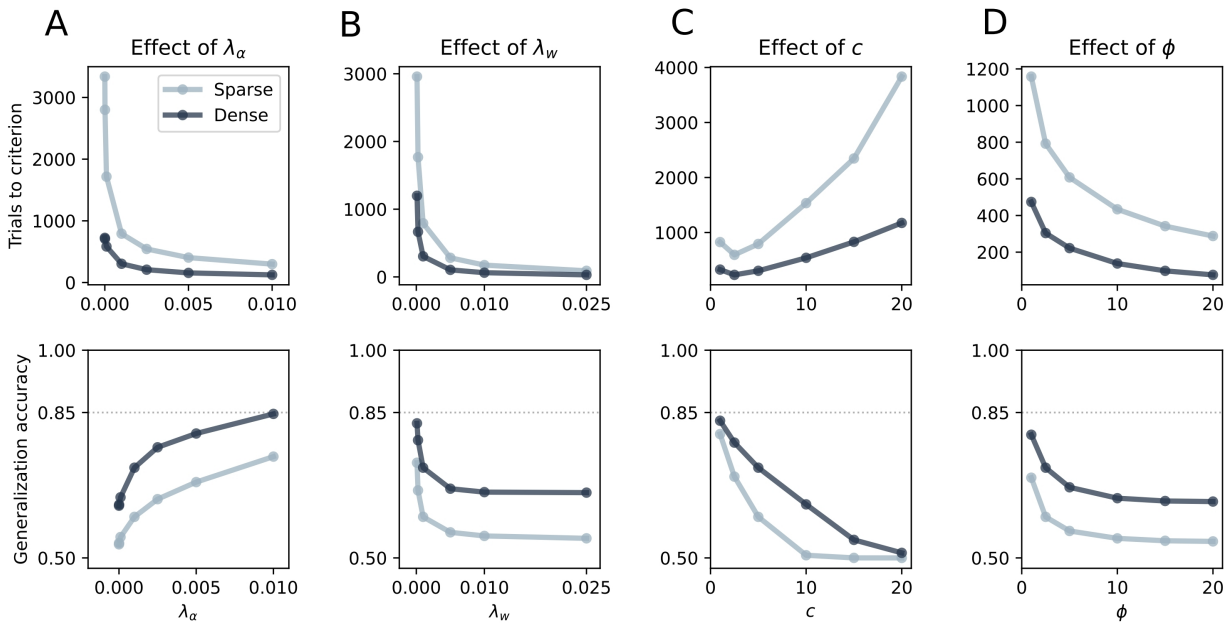

Figure 4. Simulation results by changing the four major parameters in ALCOVE. (A) $\boldsymbol{\lambda_\alpha}$, (B) $\boldsymbol{\lambda_w}$, (C) c, and (D) ϕ. For each column, the top row presents trials to reach a learning criterion of .85 for a given parameter, and the bottom row presents corresponding generalization accuracy at test. To examine the effect of these parameters, each was changed one at a time with the other parameters being fixed to $\boldsymbol{\lambda_\alpha}$ = .001, $\boldsymbol{\lambda_w}$ = .001, c = 5, and ϕ = 2.5. The dots in the plots represent the mean values across 100 simulations. Across the simulation results, only $\boldsymbol{\lambda_\alpha}$ shows a decrease in trials to criterion and an increase in generalization accuracy with increases in the parameter value.

We further examined how the model deployed attention to the category-relevant dimensions as the $\boldsymbol{\lambda_\alpha}$ parameter increased by re-plotting the results (see Figure 5). These re-plotted results showed that when the $\boldsymbol{\lambda_\alpha}$ parameter is small (Figure 5, first column), attention is distributed across all of the dimensions and is not focused on the category relevant dimension (i.e., red bar). This pattern resembles the results in the low-supervision/sparse condition with pigeons, which seem to exhibit little or no selective attention. However, when the $\boldsymbol{\lambda_\alpha}$ parameter is progressively increased (Figure 5, last column), attention is increasingly focused on the category relevant dimension depicted by the red bar.

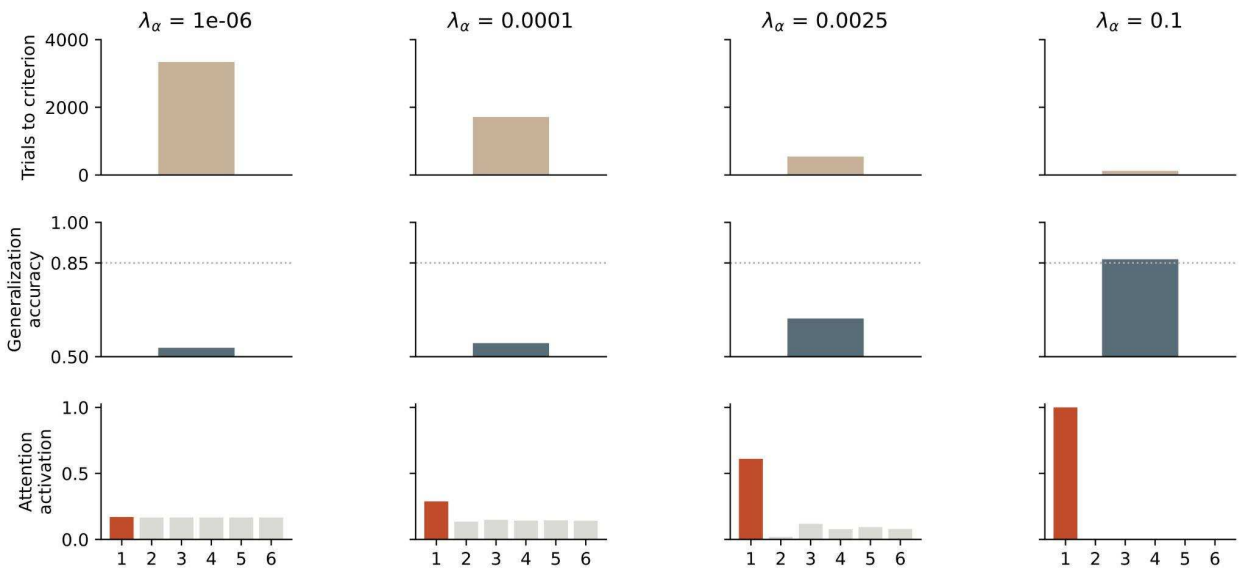

Figure 5. Simulation results using ALCOVE by changing the λ_α parameter in the sparse condition. Each column shows the simulation result of the model for a specific λ_α value presented on the top of each column while the other parameters were fixed (i.e., λ_w = .001, c = 5, and ϕ = 2.5). The top row presents the trials to criterion, where the criterion was .85. The middle row presents accuracy during the generalization phase after the criterion was reached. The bottom row presents attention weights after the criterion was reached, where the red bar represents the attention weight for the category-relevant dimension.

General Discussion

We investigated the effects of supervision, category structure, and selective attention on category learning. We examined two species that we suspected might exhibit different levels of selective attention, while manipulating category density and level of supervision during learning. Although humans learned the categories faster and generalized better to novel exemplars than pigeons, for both humans and pigeons (1) learning was faster overall when categories were densely structured than when they were sparsely structured, and (2) supervision provided greater benefits for learning sparse categories than dense ones.

Supervision aided categorization either through reducing the number of trials to learn

and/or improving generalization accuracy across all density conditions. These results differ from those of Kloos and Sloutsky (2008), who found that, although supervision facilitated learning of sparse categories, it hindered learning of dense categories. As noted earlier, we believe that this discrepancy stems from the way supervision was implemented. Kloos and Sloutsky provided supervision through verbal instructions describing all features relevant to category membership. Although useful with human participants, verbal instruction can increase cognitive load (e.g., Wan & Sloutsky, 2026; Zeithamova & Maddox, 2006). In the present experiment, supervision was provided through response feedback rather than verbal instruction, likely eliminating this additional cognitive load and thereby facilitating category learning across both category structures.

Critically, supervision interacted with both category density and the ability to selectively attend. For humans, supervision aided in decreasing the number of trials to learn, but did not affect the proportion of learners or the accuracy of category generalization. It seems that the absence of an effect for humans in the latter two measures was due to their performance being near ceiling in the dense condition. Humans learned rapidly and fully generalized the dense categories, so there was no room for supervision to afford additional assistance. On the other hand, there was at least a numerical difference in the sparse conditions in the proportion of learners (Low-Supervision: .72, High-Supervision: .84) and category generalization (Low-Supervision: .78, High-Supervision: .88). For pigeons, there was a benefit of supervision in the number of trials to learn[4] as well as in generalization performance. Although the magnitude of

[4] Note that the current experiment intended to train the pigeons until they learned. Therefore, comparing the proportion of learners across conditions in the pigeon group is misleading.

the benefit did not differ between the dense and sparse conditions (i.e., there was no interaction effect), it is noteworthy that pigeons' generalization performance was not statistically different from chance in the low-supervision sparse condition. Although the supervision manipulations were species-appropriate analogues, they were not perfectly equivalent, and this asymmetry should be considered when comparing supervision effects across species.

The finding that pigeons did not reliably generalize their learning of sparse categories under low supervision suggests that they failed to deploy selective attention in this condition. Instead, it is possible that pigeons stored all of the learned exemplars via rote memory, so that in the generalization phase, they failed to use the category-relevant dimension that was embedded within new features. This explanation is supported by the fact that pigeons failed to generalize even when they received many more training trials to learn the categories than did humans.

In ALCOVE, with more training trials, the model increases the association weights between the exemplars and the response output node. With a high attention learning parameter (i.e., λ_α), learning will be accelerated, as the model will focus only on the category-relevant dimensions in each exemplar. However, a low attention learning parameter will result in considering all exemplar dimensions equally, resulting in rote memorization of the exemplars. The ALCOVE simulation results show a similar pattern (see Figure 5, first column), where a low attention learning parameter requires longer training associated with distributed attention that lacks focus on the category-relevant dimension. However, as we employed a simulation-based approach, the simulation results obtained using ALCOVE should be

interpreted with caution. Specifically, the results do not provide a definitive mechanistic explanation, but rather serve as a proof of concept.

The advantage of the dense category structure was observed in both species, regardless of the amount of supervision. The dense category required fewer trials to learn and supported better generalization than did the sparse category. There are at least two possible explanations for this result. First, the dense category had more category-relevant dimensions (three out of six) than the sparse category (one out of six). Therefore, there is a greater chance for the learners to discover a relevant dimension in the dense condition. Additionally, as the total number of dimensions did not change between the dense and sparse structures, there were also fewer category-irrelevant dimensions in the dense structure; this too could decrease the level of distraction compared to the sparse structure. The former possibility is related to the ability to *focus* on the relevant information, whereas the latter possibility is related to the ability to *filter* out irrelevant information. Both abilities have been conceptualized under the selective attention term. However, studies show that these two abilities may not entail the same construct in category learning (e.g., Gao et al., 2024; Unger & Sloutsky, 2023). Theories of category learning have not effectively distinguished these two constructs (e.g., ALCOVE does not have a separate parameter for focusing and filtering), and it is not possible to disentangle which sub-construct may have been responsible for the differences between species in the current study.

A second, but not mutually exclusive, explanation for the dense category advantage stems from the fact that the relevant dimensions in the dense category had multiple features that were identical (i.e., three red triangles vs. three blue semi-circles). The repetitive

configurations may have increased feature salience and overall stimulus discriminability, perhaps also helping learners distinguish category-relevant from irrelevant information and optimize attentional allocation (e.g., Hammer, 2015; Hammer & Sloutsky, 2016). This argument is also supported by the fact that most of the humans and pigeons were able to learn and generalize the dense category without high supervision.

Despite the quantitative differences in the performance of humans and pigeons, the overall qualitative pattern of responding was quite similar in both species. It seems that substantially different forebrain anatomical organizations may nonetheless support similar behaviors. Birds lack a neocortex with its characteristic six-layered, laminar structure and do not possess a brain structure homologous to the mammalian PFC. Instead, the avian pallium is nucleated and organized differently from the mammalian cortex. This anatomical difference initially led researchers to assume that birds were unlikely candidates for higher-order cognition (e.g., Jarvis et al., 2005; Reiner et al., 2004).

However, recent decades of research have revealed that the avian nidopallium caudolaterale (NCL)—a multimodal telencephalic region in the posterior forebrain—shares numerous anatomical, physiological, and functional features with the mammalian PFC. Both regions serve as centers of multimodal integration (Jones & Powell, 1970; Moll & Nieder, 2015; Shimizu, Cox, & Karten, 1995), both connect sensory input to motor output (Herold et al., 2011; Kröner & Güntürkün, 1999), and both are densely innervated by dopaminergic fibers (Divac, Mogensen, & Bjorklund, 1985; Fuster, 2000; Waldmann & Güntürkün, 1993). Thus, the NCL is part of a modular forebrain network with a connective core of hub nodes that are centrally

555 located and richly connected (Shanahan et al., 2013), similar to the structural organization
556 found in mammalian brains.

557 In addition, the NCL has been implicated in characteristic prefrontal functions including
558 attention, executive control, working memory, and cognitive flexibility. Also, NCL lesions impair
559 performance on cognitive tasks that are typically disrupted by mammalian PFC damage. These
560 functional and structural similarities have led researchers to consider the NCL as the avian
561 neural structure most analogous to the mammalian PFC (e.g., Güntürkün, 2012; Güntürkün et
562 al., 2021; Pusch et al., 2023).

563 Despite these findings, relatively little research has investigated how the specific
564 differences between the NCL and PFC—such as their distinct thalamic projections (Waldmann &
565 Güntürkün, 1993)—might affect the deployment of attention in birds. As we have discussed
566 above, selective attention involves not only focusing on relevant task information, but also
567 filtering out irrelevant information, ensuring that less predictive stimulus dimensions are
568 excluded from the learning process (e.g., Corbetta & Shulman, 2002; Gulbinaite et al., 2014;
569 Lennert & Martinez-Trujillo, 2011). It could be that pigeons do focus on relevant information
570 (e.g., Castro & Wasserman, 2014), but have difficulties filtering irrelevant information. If that's
571 the case, then the pattern of distributed attention we have reported in the past (Castro et al.,
572 2020) and that our current ALCOVE simulations indicate, may be due to a filtering deficit. Thus,
573 the mature mammalian PFC may be crucial for suppressing attention to irrelevant or less
574 reliable environmental cues. The avian NCL and associated forebrain structures may lack
575 comparable filtering capabilities; without robust mechanisms for both focusing attention on

critical information and filtering out irrelevant information, pigeons may be unable to achieve the same degree of selective processing observed in humans.

Understanding the differences between the mammalian PFC and the avian NCL, and their functional consequences remains an important area for future research, particularly for clarifying how different brain architectures can support higher-order cognitive abilities.

Declarations

Conflicts of interest/Competing interests: The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' contributions: All authors contributed to conceptualization, methodology, and writing the original draft as well as manuscript review and editing. HY and LC contributed to investigation and data curation. HY contributed to formal data analysis, and visualization.

References

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5), 204–210.

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology*, *116*(2), 105–119. https://doi.org/10.1016/j.jecp.2013.05.002

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. http://dx.doi.org/10.1037/a0030404

Bruner, J., Goodnow, J. J., & Austin, G. (1956). *A study of thinking*. Wiley.

Castro, L., & Wasserman, E. A. (2014). Pigeons' tracking of relevant attributes in categorization learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 40*, 195-211. doi:10.1037/xan0000022

Castro, L., Albrecht-Buehler, A., & Wasserman, E.A. (2026). Focused and distributed: Diverse attentional profiles in pigeon category learning. *Learning & Behavior.* https://doi.org/10.3758/s13420-026-00706-6

Castro, L., Savic, O., Navarro, V., Sloutsky, V. M., & Wasserman, E. A. (2020). Selective and distributed attention in human and pigeon category learning. *Cognition*, *204*, 104350. https://doi.org/10.1016/j.cognition.2020.104350

Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to

613 categorize via feedback and via observation. *Journal of Cognitive Neuroscience, 19,* 249–
614 265. doi:10.1162/jocn.2007.19.2.249
615 Colreavy, E., & Lewandowsky, S. (2008). Strategy development and learning differences in
616 supervised and unsupervised categorization. *Memory & Cognition*, *36*(4), 762–775.
617 https://doi.org/10.3758/MC.36.4.762
618 Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in
619 the brain. *Nature Reviews Neuroscience, 3,* 201–215. doi:10.1038/nrn755
620 Deng, W. (Sophia), & Sloutsky, V. M. (2015). The development of categorization: Effects of
621 classification and inference training on category representation. *Developmental*
622 *Psychology*, *51*(3), 392–405. https://doi.org/10.1037/a0038749
623 Deng, W. (Sophia), & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the
624 development of categorization. *Cognitive Psychology*, *91*, 24–62.
625 https://doi.org/10.1016/j.cogpsych.2016.09.002
626 Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood:
627 Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.),
628 *Principles of frontal lobe function* (pp. 466–503). Oxford University Press.
629 Divac, I., Mogensen, J., & Björklund, A. (1985). The prefrontal "cortex" in the pigeon.
630 Biochemical evidence. *Brain Research, 332,* 365–368. doi:10.1016/0006-8993(85)90606-
631 7
632 Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-
633 dimension stimuli. *Quarterly Journal of Experimental Psychology*, *65*(8), 1537–1562.
634 https://doi.org/10.1080/17470218.2012.658821

635 Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the
636 primate prefrontal cortex: neurophysiology and behavior. *Journal of*
637 *neurophysiology*, *88*(2), 929-941.
638 Fuster, J. M. (2000). The prefrontal cortex of the primate: A synopsis. *Psychobiology, 28,* 125–
639 131. doi:10.3758/BF03331972
640 Gao, M., Turner, B. M., & Sloutsky, V. M. (2024). The Role of Attention in Category
641 Representation. *Cognitive Science*, *48*(4), e13438. https://doi.org/10.1111/cogs.13438
642 Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain*
643 *Theory*, *4*, 161–178.
644 Gibson, B. M., Wasserman, E. A., Frei, L., & Miller, K. (2004). Recent advances in operant
645 conditioning technology: A versatile and affordable computerized touchscreen system.
646 *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic*
647 *Society, Inc*, *36*(2), 355–362.
648 Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork.
649 *Cognition*, *52*(2), 125–157. https://doi.org/10.1016/0010-0277(94)90065-5
650 Goldstone, R. L., & Kersten, A. (2003). Concepts and Categories. In A. F. Healy & R. W. Proctor
651 (Eds.), *Comprehensive handbook of psychology, Volume 4: Experimental psychology.* (pp.
652 591–621). Wiley.
653 Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions
654 during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116–139.
655 https://doi.org/10.1037/0096-3445.130.1.116
656 Gulbinaite, R., Johnson, A., de Jong, R., Morey, C. C., & van Rijn, H. (2014). Dissociable

mechanisms underlying individual differences in visual working memory capacity. *NeuroImage, 99,* 197–206. doi:10.1016/j.neuroimage.2014.05.060

Güntürkün, O. (2005). The avian 'prefrontal cortex' and cognition. *Current Opinion in Neurobiology, 15,* 686–693.

Güntürkün, O. (2012). The convergent evolution of neural substrates for cognition. *Psychological Research, 76,* 212–219. doi:10.1007/s00426-011-0377-9

Güntürkün, O., von Eugen, K., Packheiser, J., & Pusch, R. (2021). Avian pallial circuits and cognition: A comparison to mammals. *Current opinion in neurobiology*, *71*, 29–36. https://doi.org/10.1016/j.conb.2021.08.007

Hammer, R. (2015). Impact of feature saliency on visual category learning. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00451

Hammer, R., & Sloutsky, V. (2016). Visual Category Learning Results in Rapid Changes in Brain Activation Reflecting Sensitivity to the Category Relation between Perceived Objects and to Decision Correctness. *Journal of Cognitive Neuroscience*, *28*(11), 1804–1819. https://doi.org/10.1162/jocn_a_01008

Herold, C., Palomero-Gallagher, N., Hellmann, B., Kröner, S., Theiss, C., Güntürkün, O., & Zilles, K. (2011). The receptor architecture of the pigeons' nidopallium caudolaterale: an avian analogue to the mammalian prefrontal cortex. *Brain Structure and Function, 216,* 239–254. doi:10.1007/s00429-011-0301-5

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

679 Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., … Butler, A. B. (2005).
680 Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews*
681 *Neuroscience, 6,* 151–159. doi:10.1038/nrn1606
682 JASP Team (2024). JASP (Version 0.19)[Computer software].
683 Jones, E. G., & Powell, T. P. S. (1970). An anatomical study of converging sensory pathways
684 within the cerebral cortex of the monkey. *Brain, 93,* 793–820.
685 doi:10.1093/brain/93.4.793
686 Kloos, H., & Sloutsky, V. M. (2008). What's Behind Different Kinds of Kinds: Effects of Statistical
687 Density on Learning and Representation of Categories. *Journal of Experimental*
688 *Psychology: General*, *137*(1), 52–72.
689 Kröner, S., & Güntürkün, O. (1999). Afferent and efferent connections of the caudolateral
690 neostriatum in the pigeon (*Columba livia*): A retro- and anterograde pathway tracing
691 study. *The Journal of Comparative Neurology*, *407,* 228–260.
692 Kruschke, J. K. (1992). ALCOVE: An Exemplar-Based Connectionist Model of Category Learning.
693 *Psychological Review*, *99*(1), 22–44.
694 Kuenzel, W. J., Medina, L., Csillag, A., Perkel, D. J., & Reiner, A. (2011). The avian subpallium:
695 new insights into structural and functional subdivisions occupying the lateral subpallial
696 wall and their embryological origins. *Brain research*, *1424*, 67–101.
697 doi:10.1016/j.brainres.2011.09.037
698 Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural
699 stimulus domains. *Psychonomic Bulletin & Review*, *9*(1), 43–58.
700 Lennert, T., & Martinez-Trujillo, J. (2011). Strength of Response Suppression to Distracter

Stimuli Determines Attentional-Filtering Performance in Primate Prefrontal Neurons. *Neuron, 70,* 141-152. doi:10.1016/j.neuron.2011.04.011

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, *9*(4), 829–835.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Mackintosh, N. J. (1965). Selective Attention in Animal Discrimination Learning. *Psychological Bulletin*, *64*(2), 124~150.

Medin, D. L., & Schaffer, M. M. (1978). Context Theory of Classification Learning. *Psychological Review*, *85*(3), 207–238.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202. doi:10.1146/annurev.neuro.24.1.167

Moll, F.W., & Nieder, A. (2015). Cross-modal associative mnemonic signals in crow endbrain neurons. *Current Biology, 25,* 2196–2201. doi: 10.1016/j.cub.2015.07.013

Nicholls, E., Ryan, C. M. E., Bryant, C. M. L., & Lea, S. E. G. (2011). Labeling and family resemblance in the discrimination of polymorphous categories by pigeons. *Animal Cognition*, *14*(1), 21–34. https://doi.org/10.1007/s10071-010-0339-4

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, *115*(1), 39~57.

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*,

723 *9*(1), 169–174. https://doi.org/10.3758/BF03196274
724 O'Reilly, R. C. (2006). Biologically based computational models of high level cognition. *Science,*
725 *314*, 91–94. http://dx.doi.org/10.1126/science.1127242
726 Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004). Computational approaches to the
727 development of perceptual expertise. *Trends in Cognitive Sciences*, *8*(8), 378–386.
728 https://doi.org/10.1016/j.tics.2004.06.001
729 Pothos, E. M., Edwards, D. J., & Perlman, A. (2011). Supervised versus Unsupervised
730 Categorization: Two Sides of the Same Coin? *Quarterly Journal of Experimental*
731 *Psychology*, *64*(9), 1692–1713. https://doi.org/10.1080/17470218.2011.554990
732 Pusch, R., Clark, W., Rose, J., & Güntürkün, O. (2023). Visual categories and concepts in the
733 avian brain. *Animal Cognition, 26(1),* 153–173. https://doi.org/10.1007/s10071-022-
734 01711-8
735 Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of
736 perceptually similar natural categories by 3-month-old and 4-month-old infants.
737 *Perception*, *22*(4), 463–475.
738 Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning.
739 *Cognitive Psychology*, *51*(1), 1–41.
740 Reiner, A., Perkel, D. J., Bruce, L. L., Butler, A. B., Csillag, A., Kuenzel, W., … Jarvis, E. D. (2004).
741 Revised nomenclature for avian telencephalon and some related brainstem nuclei. *The*
742 *Journal of Comparative Neurology, 473,* 377–414. doi:10.1002/cne.20118
743 Shanahan, M., Bingman, V. P., Shimizu, T., Wild, M., & Güntürkün, O. (2013). Large-scale
744 network organization in the avian forebrain: a connectivity matrix and theoretical

analysis. *Frontiers in Computational Neuroscience, 7.* doi:10.3389/fncom.2013.00089

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* (pp. vi, 117). University of Illinois Press.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*(13), 1–41.

Shimizu, T., Cox, K., & Karten, H. J. (1995). Intratelencephalic projections of the visual wulst in pigeons (*Columba livia*). *The Journal of Comparative Neurology, 359,* 551–572. doi: 10.1002/cne.903590404

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*, 1244–1286.

Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., Spiering, B., Beran, M. J., Church, B. A., Ashby, F. G., & Grace, R. C. (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2355–2369. https://doi.org/10.1016/j.neubiorev.2012.09.003

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436. https://doi.org/10.1037/0278-7393.24.6.1411

Squire, R. F., Noudoost, B., Schafer, R. J., & Moore, T. (2013). Prefrontal Contributions to Visual Selective Attention. *Annual Review of Neuroscience*, *36*(1), 451–466. https://doi.org/10.1146/annurev-neuro-062111-150439

Talwar, A., Cormack, F., Huys, Q. J. M., & Roiser, J. P. (2024). A hierarchical reinforcement learning model explains individual differences in attentional set shifting. *Cognitive,*

Affective, & Behavioral Neuroscience, 24(6), 1008–1022.

Unger, L., & Sloutsky, V. M. (2023). Category learning is shaped by the multifaceted development of selective attention. *Journal of Experimental Child Psychology*, *226*, 105549. https://doi.org/10.1016/j.jecp.2022.105549

Vandist, K., De Schryver, M., & Rosseel, Y. (2009). Semisupervised category learning: The impact of feedback in learning the information-integration task. *Attention, Perception, & Psychophysics*, *71*(2), 328–341. https://doi.org/10.3758/APP.71.2.328

Vong, W. K., Hendrickson, A. T., Navarro, D. J., & Perfors, A. (2019). Do Additional Features Help or Hurt Category Learning? The Curse of Dimensionality in Human Learners. *Cognitive Science*, *43*(3), e12724. https://doi.org/10.1111/cogs.12724

Waldmann, C., & Güntürkün, O. (1993). The dopaminergic innervation of the pigeon caudolateral forebrain: immunocytochemical evidence for a “prefrontal cortex” in birds? *Brain Research, 600,* 225–234. doi:10.1016/0006-8993(93)91377-5

Wan, Q., & Sloutsky, V.M. (2026). Working Memory Shapes Information Sampling and Attention Allocation Across Development. *Journal of Experimental Psychology: General*, 155(2), 479–498.

Wasserman, E. A., & Castro, L. (2021). Assessing attention in category learning by animals. *Current Directions in Psychological Science, 30*(6), 495–502.

Wills, A. J., O’Connell, G., Edmunds, C. E. R., & Inkster, A. B. (2017). Progress in Modeling Through Distributed Collaboration. In *The Psychology of Learning and Motivation* (Vol. 66). Elsevier Ltd. https://doi.org/10.1016/bs.plm.2016.11.007

Yim, H., Best, C. A., & Sloutsky, V. M. (2012). *The Role of Supervision in Infants’ Category*

Learning. XVIII Biennial International Conference on Infant Studies (ICIS).

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387–398. https://doi.org/10.3758/BF03193416

Supplementary Material

Experiment examining children. We have tested 4-5-year-olds using a similar paradigm as in Experiment 1. However, many children did not learn the sparse category after two days of training, which made the interpretation of the results unclear. Here we have included the experiment for full disclosure.

Experiment SM

In Experiment SM, 4-5-year-olds were examined in a category learning task similar to Experiment 1, where the procedures were slightly modified to accommodate the children.

Method

Participants

One hundred twenty-five 4-5-year-olds (66 females, *M* = 4.97 years, *SD* = .40) were recruited from preschools and daycare centers located in middle-class suburbs of Columbus, Ohio, USA. Each participant was randomly assigned to one of four conditions: Density (Dense vs. Sparse) × Supervision (Low vs. High); this assignment resulted in 31 participants except for the Low-supervision sparse condition (32). There were 22 additional participants who did not finish the experiment (1 from the High-supervision Dense condition, 4 from the Low-supervision Dense condition, 9 from the High-supervision Sparse condition, and 7 from the Low-supervision Sparse condition). The experiment was approved by the Institutional Review Board at The Ohio State University.

Materials, Design, and Procedure

Materials, design, and procedures were identical to Experiment 1, except that the experiment was conducted across two consecutive days with six blocks being administered each day. Also, the experimenter guided the participant for the entire experiment, where verbal feedback was provided by the experimenter. Moreover, the participants used a touch screen to make decisions instead of a computer mouse.

Results

Learners were defined as in Experiment 1. Overall, there were more learners in the Dense conditions and in the High-supervision conditions (see Table SM1). A binary logistic regression with Supervision and Density as covariates disclosed that there was a statistically significant effect for Supervision (Z = 3.07, p = .002, 95% CI [.86, 3.91]), and Density (Z = 5.78, p < .001, 95% CI [3.00, 6.08]). The non-learners were excluded from all following analyses.

Table SM1.

Proportion of learners in Experiment SM. The numbers in the parentheses represent the number of learners and the total number of participants.

		Sparse	*Dense*
Children	*Low-Supervision*	**0.06** (2/32)	**0.77** (24/31)
	High-Supervision	**0.29** (9/31)	**1.0** (31/31)

Trials-to-criterion were calculated as in Experiment 1 and are shown in Figure SM1A. The two Dense conditions exhibited the fastest learning (High-Supervision d: M = 42.19, SD = 30.50; Low-Supervision: M = 41.5, SD = 20.63), followed by the Sparse-Low-Supervision condition (M = 72.0, SD = 50.91), and the Sparse-High-Supervision condition (M = 90.67, SD = 42.47) condition. A 2 × 2 (Density × Supervision) between-subjects analysis of variance (ANOVA) showed a statistically significant main effect for Density (F = 10.32, p = .002, η^2 = .14, 95% CI [-2.19, -.47]), but not for Supervision, nor an interaction (Fs < .62, ps > .43, η^2 < .008).

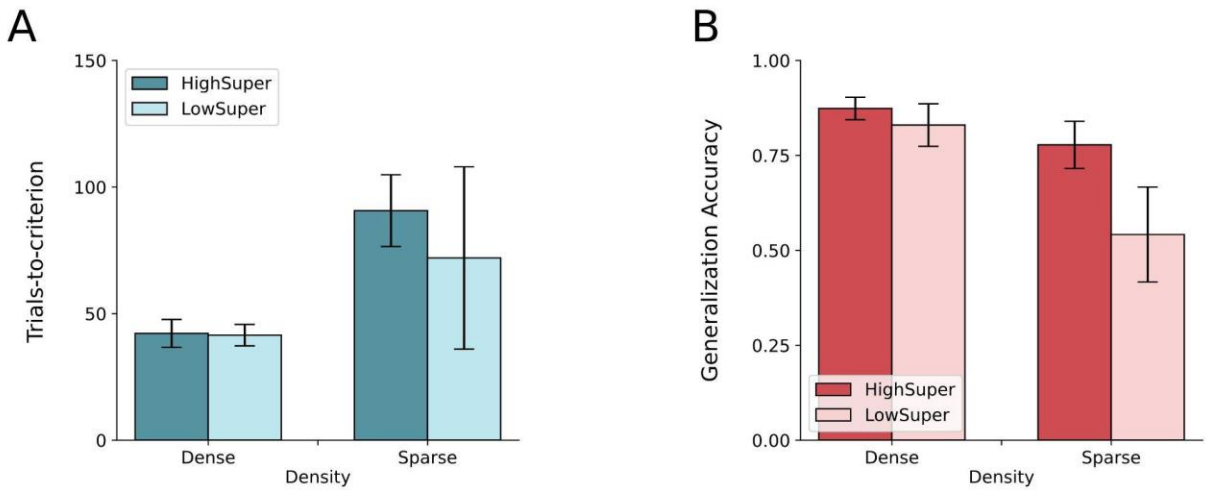

Figure SM1. Results from Experiment SM with children. (A) Trials-to-criterion, and (B) Generalization accuracy. (*Note that error bars represent +/- 1 SEM*)

Generalization showed the highest accuracy in the Dense-High-Supervision condition (M = .87, SD = .17) followed by the Dense-Low-Supervision condition (M = .83, SD = .27), the Sparse-High-Supervision condition (M = .78, SD = .19), and the Sparse-Low-Supervision condition (M = .54, SD = .18). A Density × Supervision between-subjects ANOVA showed a statistically significant main effect for Density (F = 4.68, p = .034, η^2 = .07, 95% CI [.05, 1.74]), but not for Supervision or an interaction (Fs < 2.48, ps > .12, η^2 = .04). Additionally, generalization performances were above chance (.5) for all conditions (ts > 4.47, ps < .005, Cohen's ds > 1.20), except for the Low-Supervision Sparse condition (t = .33, p = .80, Cohen's d = .24).

As there were not many learners in the two Sparse conditions (i.e., nine in the High-Supervision and two in the Low-Supervision condition), the above analyses on trials-to-criterion and generalization may have shown unreliable results (e.g., more trials to learn the High Supervision conditions). One way to re-analyze the data is to include the non-learners and assume that the participants learned the categories after experiencing all 144 trials (i.e., 12 blocks × 12 trials), and showed a chance level generalization (see Figure SM2). This would be a reasonable estimation for the generalization

performance, but it can be an underestimation for the trials-to-criterion as children did not learn even after 144 trials. However, we believe that the assumption provides a useful perspective on examining the data. Results showed the fastest learning for the Dense conditions (High Supervision: M = 42.19, SD = 30.50; Low Supervision: M = 64.65, SD = 47.16), than the Sparse conditions (High Supervision: M = 128.52, SD = 32.96; Low Supervision: M = 139.5, SD = 19.93). A 2 × 2 (Density × Supervision) between-subjects ANOVA showed a statistically significant main effect for Supervision (F = 7.57, p = .007, η^2 = .025, 95% CI [.13, .85]), and Density (F = 175.96, p < .001, η^2 = .58, 95% CI [1.91, 2.84]), but not for an interaction (F = .89, p = .35, η^2 = .003). Regarding generalization, the Dense conditions showed higher accuracy (High Supervision: M = .87, SD = .17; Low Supervision: M = .76, SD =.28), than the Sparse conditions (High Supervision: M = .58, SD = .16; Low Supervision: M = .50, SD = .03). A 2 × 2 (Density × Supervision) between-subjects ANOVA showed a statistically significant main effect for Supervision (F = 9.24, p = .003, η^2 = .046, 95% CI [-.91, -.18]), and Density (F = 71.43, p < .001, η^2 = .35, 95% CI [-1.92, -1.11]), but not for an interaction (F = .39, p = .53, η^2 = .002). Additionally, generalization performances were above chance (.50) for all conditions (ts > 2.80, ps < .01, Cohen's ds > .50), except for the Low-Supervision Sparse condition (t = .44, p = .66, Cohen's d = .08).

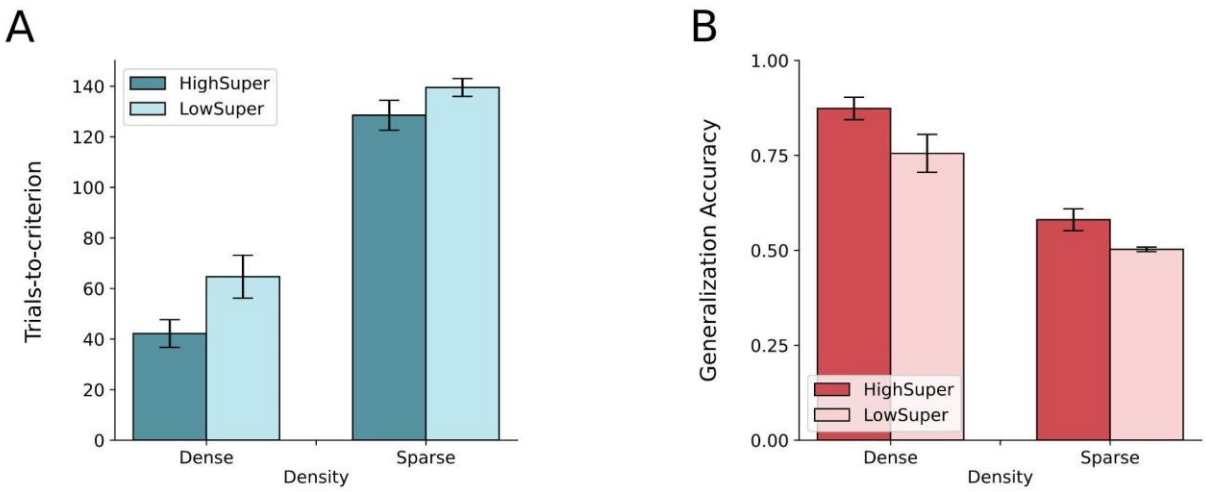

Figure SM2. Results from reanalyzing Experiment SM with children including the non-learners. It was assumed that the non-learners learned the categories after 144 trials and showed a chance level generalization. (A) Trials-to-criterion, and (B) Generalization accuracy. (*Note that error bars represent +/- 1 SEM*)

Results from the children showed an effect of category density and supervision during learning as well as during generalization. Compared to the results from adults (Experiment 1), there was evidence for the effect of supervision. The results imply that adults did not benefit from supervision as their selective attention ability was sufficient for learning both category structures. On the other hand, as the ability to selectively attend is still developing for children, the benefit of supervision was prominent for children during learning.

One caveat in interpreting the current results is that there were many children in the sparse condition that did not learn the task. As there is a practical limit on the number of learning trials that can be administered to children, we made a reasonable assumption about the non-learners when re-analyzing the data. Although the method provided a useful insight into the data, it would be optimal to actually measure the number of trials that are required to learn the categories by providing a full opportunity to the participants.