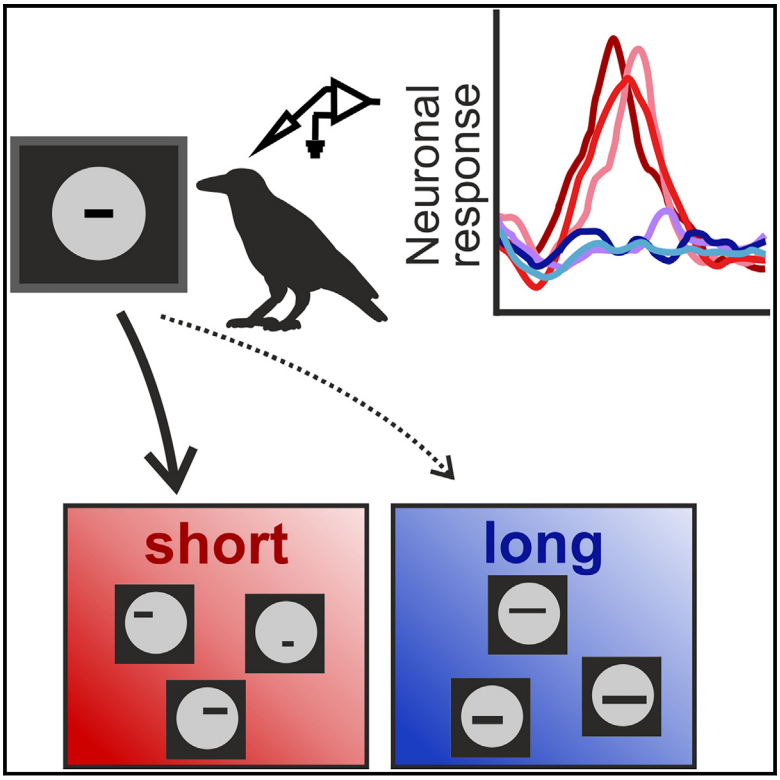
Article

Current Biology

Categorical representation of abstract spatial magnitudes in the executive telencephalon of crows

Graphical abstract



Authors

Lysann Wagener, Andreas Nieder

Correspondence

andreas.nieder@uni-tuebingen.de

In brief

Wagener and Nieder show that neurons in the NCL of crows trained to group lines into "short" and "long" categories reflected these magnitude categories in a behaviorally relevant way. Neuronal category representations changed flexibly after retraining a crow with identical stimuli to new categories "short", "medium", and "long".

Highlights

- Crows classified lines in a match-to-sample task into "short" and "long" categories
- NCL neurons encoded category information and category boundaries
- NCL activity changed with retraining to reflect new length categories
- Malleable categorization is mediated by the flexible networks of the crow NCL

Wagener & Nieder, 2023, Current Biology *33*, 2151–2162
June 5, 2023
https://doi.org/10.1016/j.cub.2023.04.013

CellPress

Current Biology

CellPress

Article

Categorical representation of abstract spatial magnitudes in the executive telencephalon of crows

Lysann Wagener[1] and Andreas Nieder[1,2,*]
[1]Animal Physiology Unit, Institute of Neurobiology, University of Tübingen, 72076 Tübingen, Germany
[2]Lead contact
*Correspondence: andreas.nieder@uni-tuebingen.de
https://doi.org/10.1016/j.cub.2023.04.013

SUMMARY

The ability to group abstract continuous magnitudes into meaningful categories is cognitively demanding but key to intelligent behavior. To explore its neuronal mechanisms, we trained carrion crows to categorize lines of variable lengths into arbitrary "short" and "long" categories. Single-neuron activity in the nidopallium caudolaterale (NCL) of behaving crows reflected the learned length categories of visual stimuli. The length categories could be reliably decoded from neuronal population activity to predict the crows' conceptual decisions. NCL activity changed with learning when a crow was retrained with the same stimuli assigned to more categories with new boundaries ("short", "medium," and "long"). Categorical neuronal representations emerged dynamically so that sensory length information at the beginning of the trial was transformed into behaviorally relevant categorical representations shortly before the crows' decision making. Our data show malleable categorization capabilities for abstract spatial magnitudes mediated by the flexible networks of the crow NCL.

INTRODUCTION

Perceptual categorization enables animals to group stimuli into behaviorally meaningful classes that can easily be generalized to new circumstances.[1] Variable stimuli are distinguished as belonging to the same category (within category) or to different categories (across category). Even if the sensory features of to be categorized stimuli change continuously, the classification judgment from one category to another is sudden, thus resulting in an abrupt category boundary.[2]

In some animals and domains, the categorical perception of stimuli can be largely innate. For example, female túngara frogs respond categorically to complex male mating calls,[3] crickets divide sound frequency categorically into attractive and repulsive sounds,[4] and lactating female house mice perceive the ultrasonic calls of their pups categorically.[5] In many other circumstances, however, perceptual categories need to be learned by trial-and-error based on behavioral feedback.[6] For instance, young vervet monkeys need to learn to identify the predator category alarm calls,[7] and songbirds learned to recognize new alarms by association with known alarms.[8] Evidently, the capability to categorize stimuli offers survival and reproduction benefits and therefore is widespread across the animal kingdom.[9]

Experience-dependent categorization is frequent in cognitively flexible vertebrates. It can be found in mammals[10,11] and birds such as pigeons[12–19] and crows.[20–24] Similar to the hierarchical processing pathway in the primate brain,[25] behaviorally relevant stimulus features supporting categorical neuronal responses seem to be extracted gradually along the two major visual forebrain pathways of birds[26]: the thalamofugal pathway (homolog to the mammalian geniculocortical pathway) and the tectofugal pathway (thought to be analogous to the mammalian extrastriate cortex[27]). In the avian telencephalon, rudimentary category representations emerge first via the thalamofugal pathway in the thalamorecipient structures of the visual Wulst and via the tectofugal pathway in the entopallium and the overlaying intercalated nidopallium (NI) and mesopallium ventrolaterale (MVL) layers.[28,29] From these layers, highly integrated visual information still lacking sufficient feature invariance is forwarded to the dominant associative cognitive control center of the avian brain, the nidopallium caudolaterale (NCL). Based on a variety of anatomical and functional criteria, the NCL is thought to be an avian equivalent of the primate prefrontal cortex (PFC),[30–36] a mammalian brain area of great importance in categorization.[11,37–40]

Neuronal responses that establish behaviorally relevant conditional stimulus-response contingencies have been reported several times in the avian NCL.[18,41–44] Clear categorical neuronal responses were observed in the realm of numerical quantity. NCL neurons are tuned to the number of items in visual displays, both in numerically trained crows[45–48] but also in numerically naive crows[49] and untrained 10-day-old domestic chicks.[50] The latter findings suggest that categorical responses to number emerge largely spontaneously based on mechanisms inherent to the visual system.[51,52] How learned magnitude categories emerge in the avian brain and the neuronal mechanisms underlying them is currently unknown.

Here, we explored crows' behavioral and neuronal representation of learned magnitude categories. We tested three assumptions: first, we hypothesized that neurons in the corvid NCL represented learned and abstract spatial categories in a stimulus feature invariant and behaviorally relevant manner. Therefore, we trained crows in a delayed match-to-category task to group

the lengths of parameter-controlled lines into the categories "short" vs. "long" by relying on learned and arbitrary rules while recording from neurons of the NCL during performance. Second, we assumed that NCL neurons can flexibly adapt to new category boundaries if categorization rules change. Therefore, we retrained one crow with the line lengths reassigned to three new categories short, medium, and long. Third, we predicted that crow NCL neurons, despite a distinct neuroanatomy, exhibit a similar code for categories as PFC neurons in monkeys. Similarities of crow NCL data with monkey PFC findings would lend support to the notion of a superior physiological solution to the same categorization challenge in convergently evolved telencephalic executive brain regions.

RESULTS

Two crows were trained in a delayed match-to-category task to categorize line stimuli according to their length into two groups (short and long categories). Six different line lengths were used that were assigned to the two length categories short (S1, S2, and S3) vs. long (L1, L2, and L3) (Figure 1B). To ensure that the crows categorized length rather than the area or thickness of the lines, we used two stimulus protocols ("standard," where line thickness varied pseudo-randomly across line lengths, and "control," where the area of each line was constant) in each session.

Behavior

Both crows were able to memorize and match the sample line length to the category-matching length in the test phase. The crows performed proficiently above the 50% chance level (crow 1: 87.2% ± 0.5% SEM, n = 52 sessions; crow 2: 87.7% ± 0.7% SEM, n = 55 sessions) in each session (all binomial tests, $p < 0.001$). The behavioral performance was a step function with similar responses for stimuli of the same category and a sharp change across the category boundary (Figures 1C and 1D). Both crows reliably categorized each of the six individual sample stimuli to the appropriate length category, irrespective of whether standard or control protocols were shown (Figures 1E and 1F).

As expected for parameterized length magnitude, both crows categorized the lengths most distant from the category boundary (S1 and L3, respectively) most proficient and the lengths near the category transition (S3 and L1, respectively) least proficient, resulting in performance differences of between 2.6% and 19.7% between the most distant and the closest line length to the category boundary (S1 vs. S3 and L3 vs. L1, respectively) (Kruskal-Wallis tests, $p < 0.001$; except for short stimuli in match conditions for crow 1 ($p = 0.16$), Figure 1C, left). However, this within-category performance drop was mild compared with the substantial across-category (short vs. long) difference of, on average, 76.9% for crow 1 and 77.3% for crow 2 (Kruskal-Wallis tests, $p < 0.001$ for both short vs. long and long vs. short categorizations in both crows).

Both crows were slightly better in discriminating the control protocol compared with the standard protocol (paired t test, $p < 0.001$, Figures 1E and 1F). The mean performance of crow 1 with standard and control stimuli was 84.9% and 89.8%, respectively. Crow 2 had a mean performance of 84.0% with standard stimuli and 92.1% with control stimuli. However, the performance of both crows in each session was clearly above the 50% chance level with either stimulus set (all binomial tests for individual sessions and both crows $p < 0.01$).

Neuronal data

We recorded the single-cell activity of 449 NCL neurons (crow 1, 195 neurons; crow 2, 254 neurons) while the crows performed the length categorization task. Overall, 134 neurons (29.8% overall; 43% in crow 1 and 20% in crow 2) were found to be category selective in specific trial intervals and showed firing rate differences between the short vs. long categories (two-factor ANOVA, $p < 0.01$), but no differences within the two categories (Kruskal-Wallis tests, $p \geq 0.05$). In the sample phase, 65 neurons were category selective (Figure 2A), whereas 86 neurons were category selective during the delay phase (Figure 2B). The responses of four category-selective example neurons are shown in Figures 2C–2F. The neurons in Figures 2C and 2D increased their firing rates selectively to the line stimuli of the long category in the sample phase and during the delay, respectively. The other two example neurons responded selectively to category short in the early (Figure 2E) and later delay (Figure 2F), respectively.

The preferred category of a selective neuron was defined as the one eliciting the highest firing rate within the selective time window. In the sample phase, slightly more neurons preferred the short category (n = 41/65, binomial test, $p = 0.046$): in the delay, a similar number of 47 selective neurons preferred short, whereas 39 neurons preferred the long category (binomial test, $p = 0.45$)

Category-selective neurons robustly encoded the learned categories; their tuning fulfilled the hallmarks of categorical responses, i.e., a similar response to all members of the same category and a change in activity across the category boundary (Figures 2G and 2H). Both in the sample and the delay phase, a significant difference between the neurons' firing rates to the preferred vs. the non-preferred categories was observed (Wilcoxon signed-rank test, $p < 0.001$). No firing rate differences were found within the preferred and non-preferred categories for sample and delay (Friedman tests, $p > 0.05$), except for firing rates within the non-preferred category of sample-selective neurons (Friedman test, $p = 0.02$).

We calculated a category index to quantify the difference in the firing rates of the category-selective neurons (analogous to Freedman et al.[37,53]). We first derived the "within-category difference" (WCD) and "between-category difference" (BCD) from the neurons' firing rates (see STAR Methods). For the population of selective neurons, the BCD was significantly higher than the WCD, resulting in a shift of the data above the diagonal when plotted against each other (Figures 2I and 2J) (Wilcoxon signed-rank test, $p < 0.001$ for both sample [n = 65] and delay-selective neurons [n = 86]). The WCD and BCD were then used to calculate the category index—positive index values (max. +1) indicate higher firing rate differences for stimuli of different (across) categories, whereas negative values (min. −1) signify higher differences for stimuli of the same (within) category (Figures 2K and 2L). For both the sample- and delay-selective category neurons, the distributions were significantly shifted toward positive values with means of 0.36 and 0.45, respectively (both one-sample t tests, $p < 0.001$), indicating strong category coding of the selective

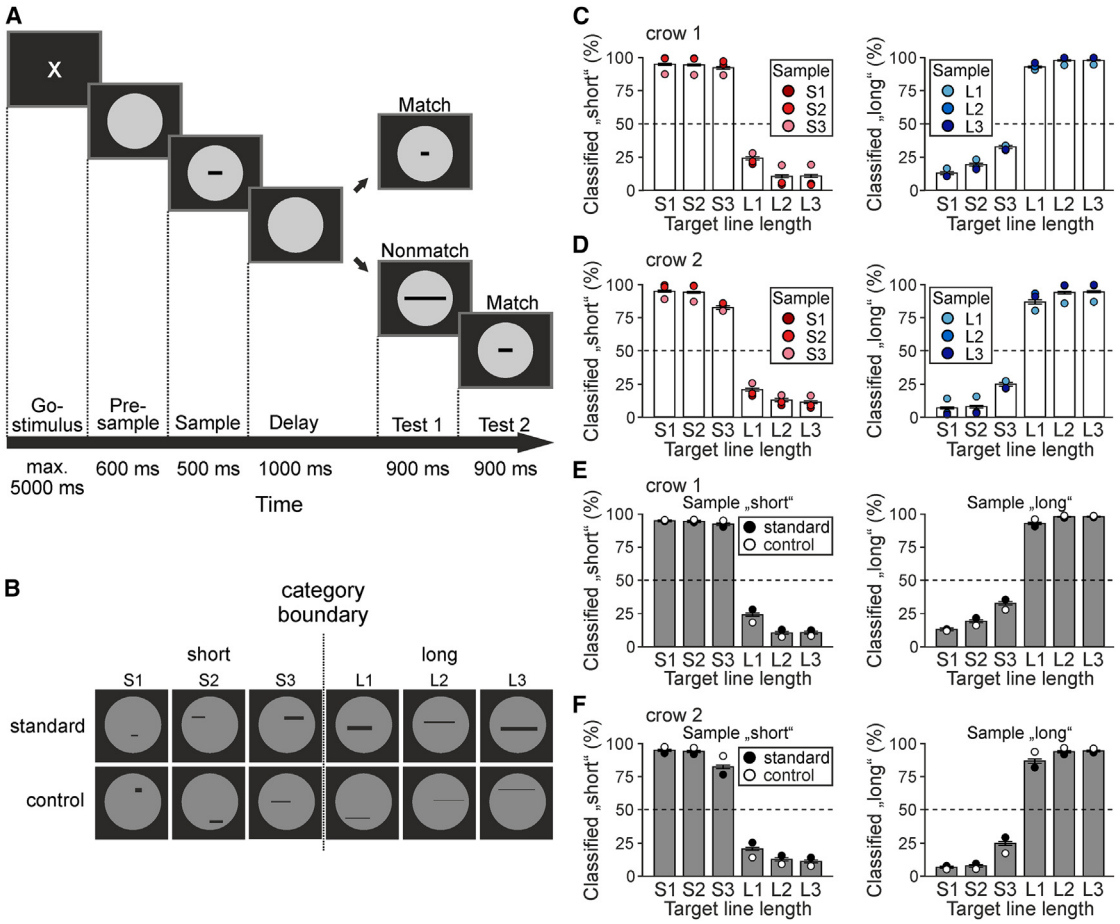


Figure 1. Task protocol, example stimuli, and behavioral performance in the two-category task

(A) Layout of the delayed match-to-category task with line-length stimuli. The crows had to respond whenever test 1 in 50% of the trials showed a line length that matched the short-vs.-long length category of the sample. In the other 50% of the trials, test 1 was a category "nonmatch"; here, the crow had to refrain from responding until the second test stimulus (test 2) was shown, which was always a category "match."

(B) Example stimulus displays of the two-category task. Two stimulus sets (standard and control) with six line lengths each were used. Category boundary divided the stimuli into short and long categories, with three line lengths each.

(C) Percent correct performance of crow 1 in the two-category task. Left: performance in trials with short sample stimuli. Right: performance in trials with long sample stimuli. The values depict the percentage of how often the crows correctly judged the length of either test 1 or test 2 as belonging to the same category as the line length of the sample stimulus. The circles indicate which exact line length was previously shown as the sample stimulus. Chance level is 50% (dashed lines). Error bars (very small) represent SEM across the sessions.

(D) Same as in (C) but for crow 2.

(E) Behavioral performance of crow 1 for the two different stimulus sets (standard and control) individually. Left: performance in trials with short sample stimuli. Right: performance in trials with long sample stimuli. Chance level is 50% (dashed lines). Error bars (very small) indicate SEM across the sessions.

(F) Same as in (E) but for crow 2.

neurons. Category indices had a tendency to be larger during the delay period (two-sample t test, $p = 0.067$).

Analysis of the population of category-selective neurons

To assess the activity of the category-selective neurons together across time, we transformed their activity into state space. Here, the activity of a population of neurons at every moment is represented as an *n*-dimensional vector in *n*-dimensional space. After dimensionality reduction to the three most informative dimensions (first three principle components [PCs]), the trajectories in three-dimensional space represent the time course of the neuronal activity to the different line lengths (see STAR Methods). Sample-selective and delay-selective category neurons were separately analyzed. Figures 3A and 3B depict the resulting activity trajectories across trial time in state space. While the absolute position of the color-coded trajectories representing the six line lengths is irrelevant, the distances between trajectories reveal differences in population activity. Visual inspection shows that similar line lengths are encoded in an orderly fashion by nearby trajectories. In addition, trajectories within a category seem to be closer, whereas trajectories across the two categories appear more distant.

We performed a cluster analysis to explore the potential clustering of population activity according to the categories. We calculated PC scores with average firing rates in the sample and delay period separately (see STAR Methods). The dispersion

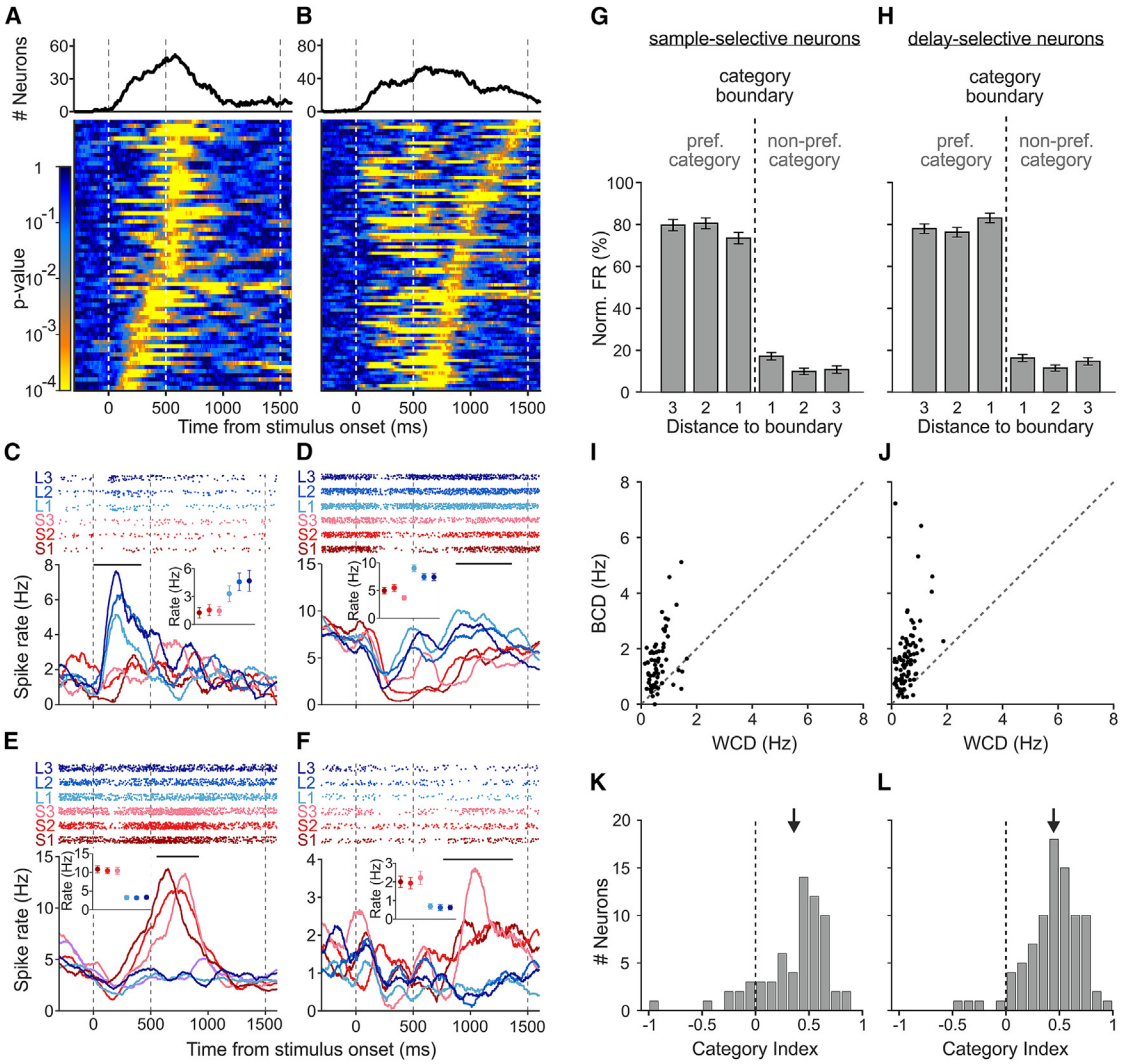


Figure 2. Single-neuron activity from NCL in the two-category task

(A and B) Pattern of task-selectivity of the neurons with a selective interval in the sample (A) and delay phase (B). Top: time-resolved histograms depicting the number of neurons for which factor "category" was significant at a given time point. Bottom: color-coded traces of the p values. Each line represents a neuron. Dashed lines separate the periods of a trial by indicating sample onset (at 0 ms), sample offset (at 500 ms), and end of delay (1,500 ms).

(C–F) Responses of four single neurons selective to category long in the sample (C) and delay phase (D), respectively, and selective to category short in the early delay (E) and later delay (F). Top panels depict dot-raster histograms (each line corresponds to a trial and each dot is an action potential). Bottom panels represent the corresponding averaged and smoothed (200 ms Gauss kernel, step size of 1 ms) spike-density functions. Each line shows the time course of the activity for the six different line lengths. Vertical dashed lines indicate sample onset, sample offset, and end of delay. The horizontal black line indicates the selective interval. Tuning function insets show the average firing rate to each line length during this interval (error bars indicate SEM across the trials).

(G and H) Average normalized activity of category-selective neurons in the sample (G) and delay phase (H) in response to the individual line lengths of their preferred and non-preferred category. The line lengths are arranged according to their distance from the category boundary. Error bars indicate SEM.

(I and J) Difference in firing rates in response to sample line lengths of the same (WCD) and different categories (BCD) for sample (I) and delay category-selective neurons (J).

(K and L) Frequency distribution of category indices for sample (K) and delay category-selective neurons (L). Arrows indicate respective means.

of the PC scores (only the first two PCs) for each trial (n = 180) in PC-space is shown in Figures 3C and 3D. We first determined the optimal number of clusters for the datasets by applying two measures: the Caliński-Harabasz index (also termed "variance ratio criterion [VRC]"),[54] and the "gap criterion" that determines the most dramatic decrease in error measurement (the "elbow" or "gap") of different cluster numbers (see STAR Methods).[55] In the sample period, the Caliński-Harabasz index (which is only defined for two or more clusters and thus less reliable) indicated two as the optimal cluster number, whereas the gap value indicated only one cluster as an optimal description of the population activity (Figure 3E). However, in the delay period, both measures indicated two clusters as the optimal cluster number (Figure 3F).

We then applied unsupervised k-means clustering to partition all trials in state space (n = 180) into the previously determined

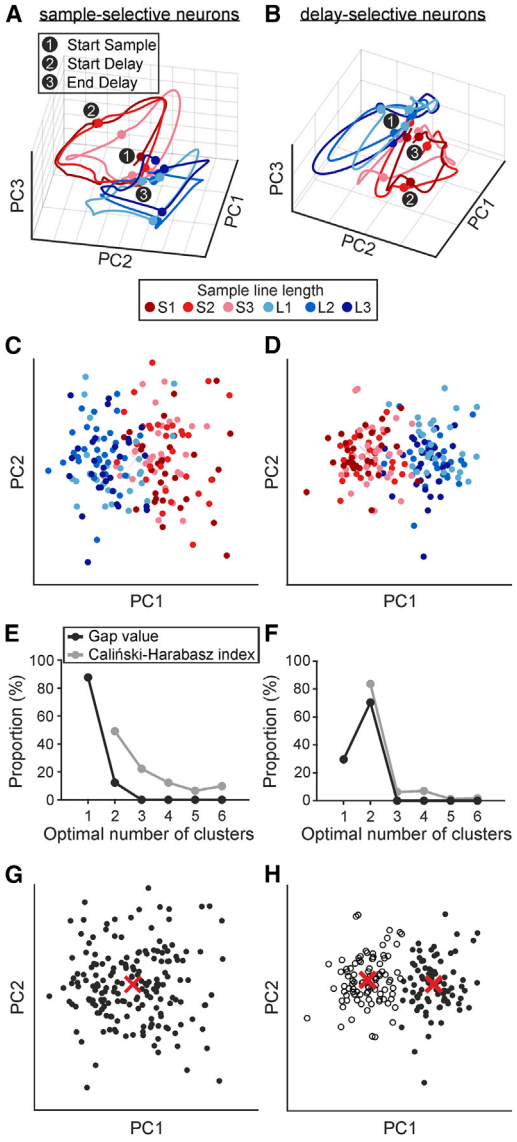


Figure 3. State space analysis of the selective neurons of the two-category task

(A and B) Time course of neuronal activity to the different line lengths throughout a trial (1, start sample phase; 2, start delay; 3, end of delay) of neurons that were category selective in the sample (A) and delay phase (B).

(C and D) Dispersion of the PC scores of an example clustering repetition during the sample (C) and delay phase (D). One dot corresponds to one trial, color-coded by the different sample line lengths.

(E and F) Proportion of the optimal number of clusters based on gap value and Caliński-Harabasz index, respectively, in the sample (E) and delay phase (F).

(G and H) Cluster assignment based on gap value of the same trials as in (C) and (D), respectively. The optimal number of clusters was "one" in the sample phase (G) and "two" in the delay phase (H). Red crosses indicate the position of the cluster's centroids.

optimal number of clusters.[56] In the sample period, the indifferent data comprised a single cluster (Figure 3G). In the delay period, however, the clustering algorithm detected one cluster for each of the two (short vs. long) length categories (Figure 3H). However, with trial progression, activity in state space encodes the relevant two length categories by two clusters that border between the length categories.

Analysis of the entire neuron population

In the next step, we explored the category coding capability of the entire population of recorded neurons (n = 348), irrespective of selectivity. We focused on the last 600 ms of the delay period in which the crows particularly relied on category information to solve the task. First, we calculated a correlation matrix to compare the responses of the neurons with pairs of stimuli (Figure 4A). A correlation coefficient was calculated for each stimulus combination, and its value is depicted as a color-coded tile in the correlation matrix. The emerging correlation pattern shows that the responses of all neurons were more similar to within-category stimuli than to across-category stimuli. The mean coefficient for correlations for within-category stimuli was 0.76 and thus higher compared with the mean coefficient of 0.63 for across-category stimuli (two-sample t test, $p < 0.001$) (Figure 4B).

To explore the behavioral relevance of population activity for the crows' categorization performance, we calculated the correlation coefficients for suitable neurons also in error trials in addition to correct trials over the last 600 ms of the delay period (Figures 4C and 4D). The correlation coefficients in correct trials differed significantly for this subset of neurons, with means of 0.77 and 0.62 for stimuli of the same and different categories, respectively (two-sample t test, $p < 0.001$). In error trials, however, no difference between within-category correlation coefficients (mean = 0.41) and between-category correlation coefficients (mean = 0.39) was found (two-sample t test, $p = 0.70$) (Figure 4E). This indicates that the activity differences between categories that are lacking in error trials are behaviorally relevant for the crows to group the sample stimuli into the learned categories during correct trials.

Next, we used a population decoding approach to explore categorical information contained in the neuronal responses. We trained a support vector machine (SVM) classifier with the firing rates of the neurons within the last 600 ms of the delay.[57] The classification performance was then tested with a subset of these firing rates which were not used for training (Figure 4F). The classifier grouped the firing rates with a high performance of 91.2% (±0.9% SEM) into the correct categories. Additionally, we trained an SVM classifier 3 times on different pairs of training stimuli to test whether the firing to each of the six individual stimuli was predictive of the short vs. long categorization. For each classifier training, we used the firing rates to one stimulus of each category (S1/L3, S2/L2, and S3/L1, respectively) and then predicted the category of the remaining four stimuli. The classifier was able to predict the correct category at a mean performance of 77.9% (±3.2% SEM) (Figure 4G). All training sets resulted in similarly high classifier performance without performance differences between the tested cross-category line pairs (two-factor ANOVA, $p = 0.58$; mean classification performance with S1/L3 as the training set, 76.0%; with S2/L2, 83.0%; with S3/L1, 74.8%). Importantly, there was no difference between the performance of stimuli within the short and the long category (two-factor ANOVA, $p = 0.52$; mean prediction performance for stimuli of the short category: 75.7% and long category: 80.2%). These decoding results show that the neurons respond in a similar manner to all stimuli of the same category but differently to members of the other category, thus allowing a classifier to predict category membership.

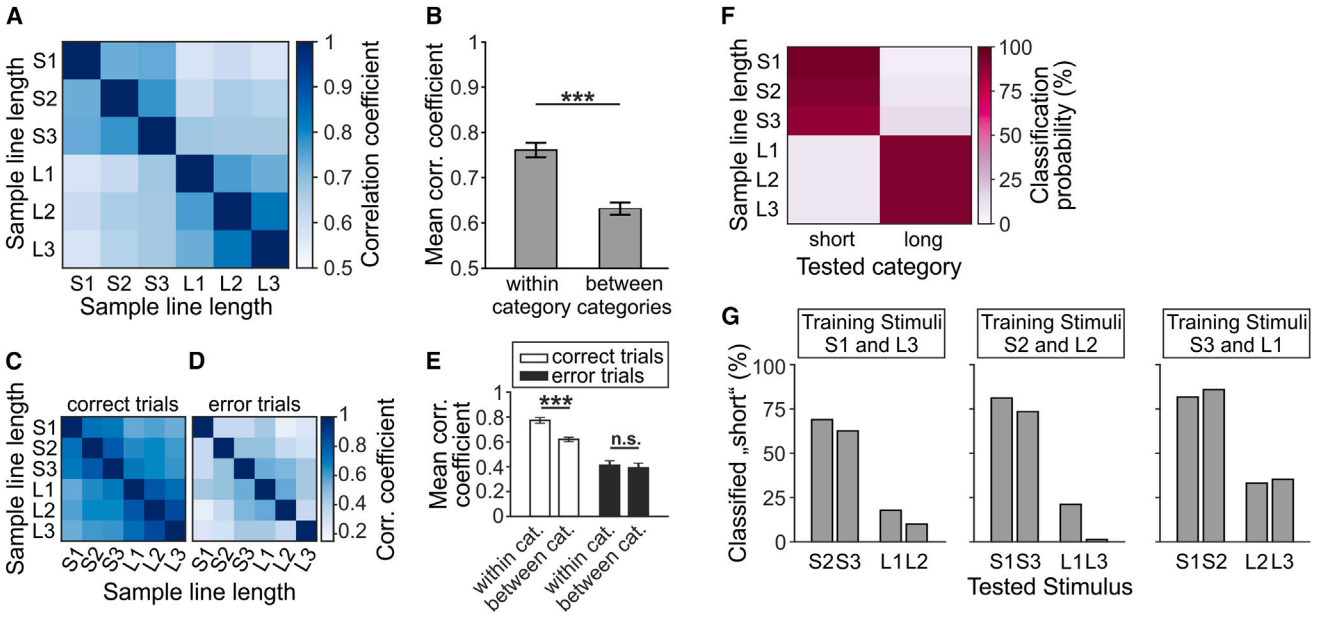


Figure 4. Correlated activity to pairs of stimuli and classification probability of an SVM classifier for the entire neuronal population in the two-category task

(A) Correlation matrix to pairs of stimuli comparing the neuronal activity during the last 600 ms of the delay. Each tile represents the correlation coefficient via color code. Darker colors indicate higher correlation. The tiles along the diagonal represent the maximum correlation when comparing stimuli with themselves ($r = 1.0$).

(B) Mean correlation coefficient across all comparisons of stimuli within the same category and of different categories, respectively. Error bars represent SEM. ***: $p < 0.001$.

(C and D) Correlation coefficients in correct (C) and error trials (D) for a subset of the neuronal population for which error trials could be analyzed.

(E) Mean correlation coefficients across comparisons of stimuli within the same category and of different categories in correct trials (left two bars) and error trials (right two bars). Error bars represent SEM. ***: $p < 0.001$.

(F) Performance of an SVM classifier classifying the category of the sample stimulus after being trained on the firing rates of the entire neuronal population during the last 600 ms of the delay.

(G) SVM classifier predictive performance to novel stimuli after classifier training on the firing rates to two other stimuli. Data for classifier training with S1 and L3 (left), S2 and L2 (middle), and S3 and L1 (right). Columns represent the proportion of how often a stimulus was assigned to the short category.

Behavior in the three-category task

After collecting data in the two-category task, we retrained crow 1 on a three-category task to explore learning-related categorization changes. We used the same line-length stimuli but applied two category boundaries that resulted in the three length categories short, medium, and long (Figure 5A). The crow was able to learn the new categories and performed above the 50% chance level in each session (all binomial tests, $p < 0.001$). The mean correct performance across all sessions was 83.1% (±3.7% SEM, $n = 58$ sessions).

The crow showed similarly high performances for either stimulus of each category and a sharp drop-off across the two category boundaries (Figure 5B). As with the two-category task, the crow performed best when the sample stimulus was S1 (95.0% ± 0.5% SEM, Figure 5B, left). However, performance was also high for the medium category which is the most difficult category because both within-category stimuli are adjacent to a category boundary (M1, 81.3% ± 0.9% SEM; M2, 75.1% ± 1.0% SEM; Figure 5B, middle). The mean performance with standard and control stimuli was 81.4% (±0.7% SEM) and 85.2% (±0.7% SEM), respectively, and thus slightly better with stimuli of the control set (paired t test, $p < 0.001$, Figure 5C). However, the crow's performance in each session was well above chance level with both stimulus sets (all binomial tests, $p < 0.001$).

Neuronal data in the three-category task

We recorded 336 single neurons while crow 1 was performing the three-category task. Of these, 128 neurons (38.1%) were category selective (47 neurons in the sample phase and 93 neurons during the delay). Three category-selective neurons are shown in Figures 5D–5F. The sample-selective neuron in Figure 5D was tuned to the long category. The other two delay-selective example neurons preferred the medium category in the middle of the delay (Figure 5E) and the long category toward the end of the delay (Figure 5F), respectively. In the sample phase, 14 neurons preferred a stimulus of the short category, 11 of the medium, and 22 of the long category. During the delay, 17 preferred short, 53 preferred medium, and 23 preferred long.

Analogous to the analysis of the two-category data, we assess the activity of all selective neurons in the three-category task together in state space (Figures 6A and 6B). In addition to the orderly representation of adjacent line lengths, trajectories within one of the categories seem to be closer, whereas trajectories across categories appear more distant. We again performed cluster analysis with PC scores ($n = 180$ trials) and separately for the sample and delay periods (Figures 6C and 6D).

In the sample period, the Caliński-Harabasz index indicated three as the optimal cluster number, whereas the gap value indicated only two clusters as an optimal description of population

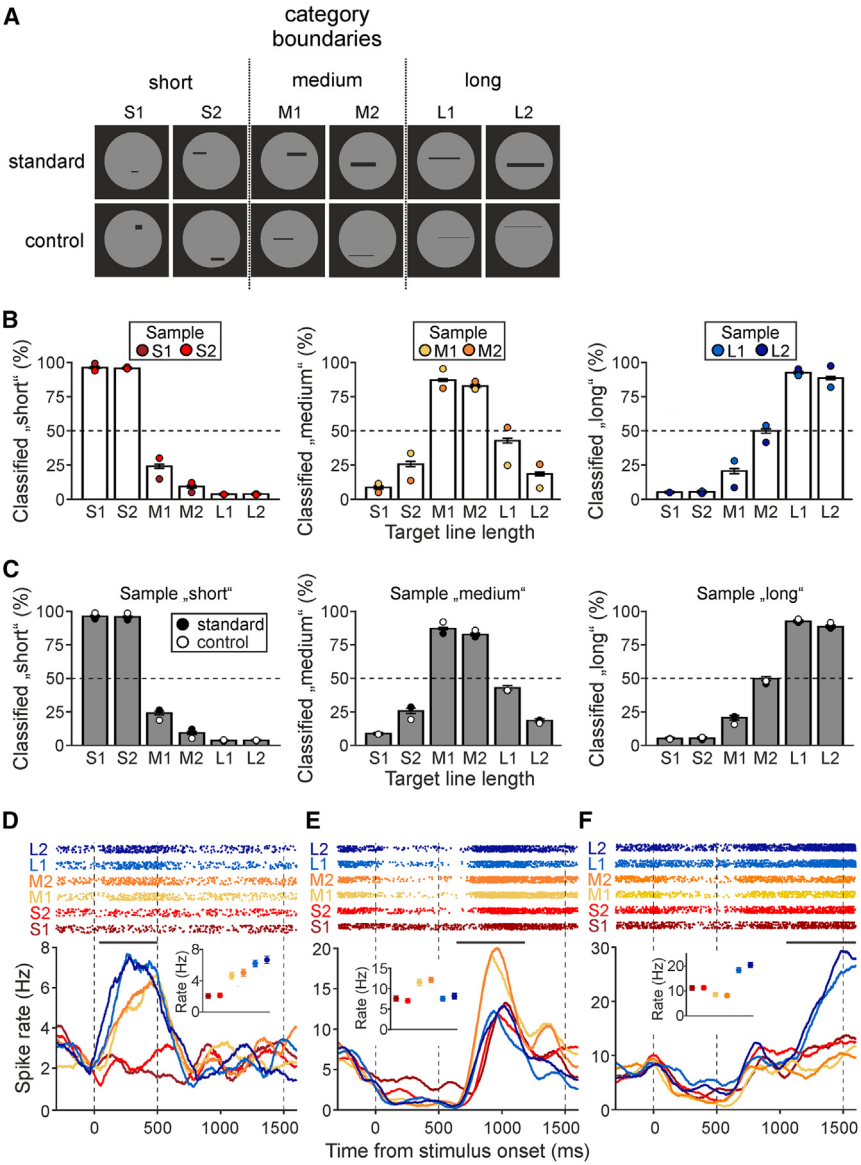


Figure 5. Example stimuli, behavioral performance, and single-neuron activity in the three-category task

(A) Example stimulus displays of the three-category task. Two different stimulus sets (standard and control) with six line lengths each were used. The category boundaries divided the stimuli into three categories (short, medium, and long), with two line lengths each.

(B) Behavioral performance of crow 1 in the three-category task. Left to right: performance in trials with short, medium, and long sample stimuli, respectively. Circles indicate which line length was shown as the sample stimulus. Dashed lines represent chance. Error bars (very small) represent SEM across the sessions.

(C) Behavioral performance of crow 1 for the two different stimulus sets (standard and control) individually. Left to right: performance in trials with short, medium, and long sample stimuli, respectively. Layout as in (B).

(D) Category-selective neuron encoding the three categories in the sample phase and preferring category long. Layout as in Figures 2C–2F.

(E and F) Example category-selective neurons during the delay, preferring the medium (E) and long categories (F), respectively. Layout as in Figures 2C–2F.

activity (Figure 6E). However, in the delay period, both measures indicated three as the optimal cluster number (Figure 6F). In the sample period, k-means clustering partitioned the trials into the previously determined two optimal clusters which mainly consisted of the first shortest vs. the three longest line stimuli (Figure 6G). In the delay period, however, the clustering algorithm detected three clusters correlating with the line lengths of the short, medium, and long categories (Figure 6H). Thus, state space activity later in the delay period encodes the relevant three length categories by three clusters that border between the trained length categories.

As before, we analyzed the category coding of the entire population of recorded neurons (n = 278) and used again the activity in the last 600 ms of the delay period to calculate a correlation matrix. After retraining crow 1, the correlation pattern now reflected the new three categories (Figure 7A). The difference between the mean correlation coefficient for stimuli within the same category (0.80) and for correlations of stimuli across categories (0.59) was significant (Figure 7B) (two-sample t test, $p < 0.001$). By contrast, no differences in correlation coefficients were observed for stimuli that belonged to adjacent categories (short vs. medium and medium vs. long; mean = 0.57) or had a greater distance (short vs. long; mean = 0.62) (two-sample t test, $p = 0.06$).

We tested whether the neuronal population recorded during the three-category task may still encode the now invalid two categories of the original task. The mean correlation coefficients for stimuli within the original categories of the two-category task compared with stimuli between the original two categories were 0.65 and 0.61, respectively, and were indifferent (two-sample t test, $p = 0.44$). This indicates that the population of neurons no longer encoded the categories of the two-category task. As a control, we explored three-category coding in the original two-category task and calculated the correlation analysis with the categories of the three-category task for the data recorded during the two-category task (see Figure 4A). Here as well, the mean correlation coefficients for stimuli within the same category (0.74) and those between categories (0.67) were indifferent (two-sample t test, $p = 0.12$). Thus, the correlation differences within and between categories of the three-category task were not a chance event but resulted from retraining with the new categories.

A subset of the neuronal population (n = 120 suitable neurons; see STAR Methods) was used to analyze the correlation coefficients in error trials (Figures 7C and 7D). Although a higher correlation coefficient for stimuli within the same category than for stimuli across categories (0.82 and 0.56, respectively) was

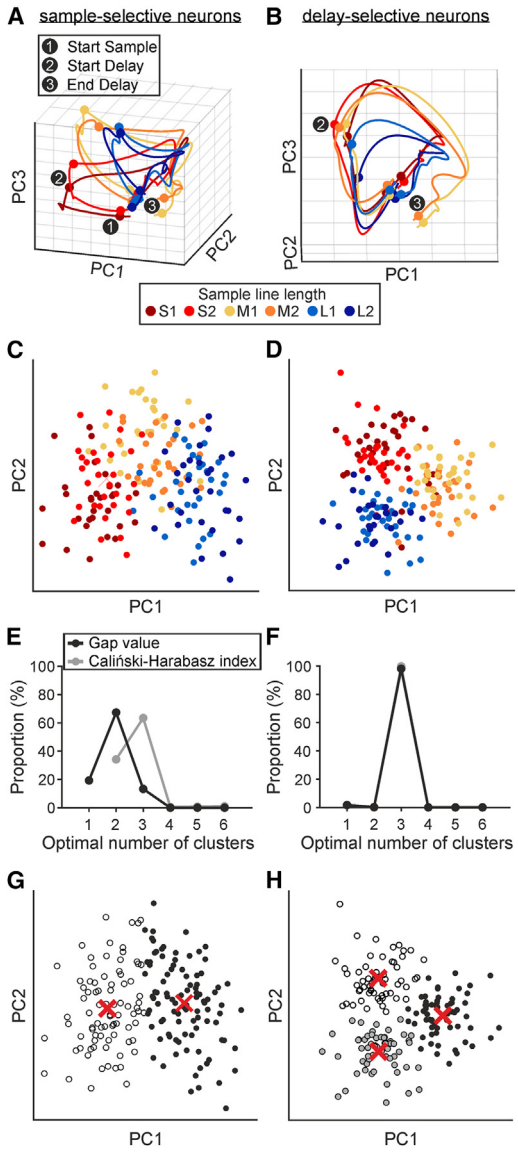


Figure 6. State space analysis of the selective neurons of the three-category task

(A and B) Course of neuronal activity in response to the different line lengths throughout a trial (1, start sample phase; 2, start delay; 3, end of delay) for neurons that were category selective in the sample phase (A) and during the delay (B).

(C and D) Dispersion of the PC scores of an example clustering repetition during the sample period (C) and during the delay (D). One dot corresponds to one trial, color-coded by the different sample line lengths.

(E and F) Proportion of the optimal number of clusters based on gap value and Caliński-Harabasz index, respectively, in the sample phase (E) and during the delay (F).

(G and H) Cluster assignment based on gap value of the same trials as in (C) and (D), respectively. Here, the optimal number of clusters was two in the sample phase (G) and three in the delay (H). Red crosses indicate the position of the cluster's centroids.

observed in correct trials (two-sample t test, $p < 0.001$, Figure 7E), in error trials, no difference between the correlation coefficients for stimuli of the same category and stimuli across categories was detected (both 0.32, two-sample t test, $p = 0.99$). This again indicates that category-specific signals in the three-category task had vanished during error trials, supporting the behavioral relevance of the neurons' responses.

Finally, we employed decoding population analyses irrespective of single neurons' category selectivity. We again trained an SVM classifier using the firing rates of the last 600 ms of the delay and tested its classification performance with a subset of firing rates that were not used for training (Figure 7F). The classifier grouped the firing rates correctly into the three categories with a high probability of 90.1% (±1.5% SEM). This indicates a robust representation of the three length categories by the random population of NCL neurons.

DISCUSSION

We report that crows efficiently learned to apply a matching to category rule based on short or long line length. We report three major findings from recordings during task performance. First, a substantial proportion of neurons encoded the category information by showing large activity differences between length categories but similar responses to stimuli within each length category. Second, after the retraining of a crow and testing with new and more length categories (short, medium, and long), NCL neuron activity had flexibly changed to now reflect these new categories. Our data show malleable categorization capability mediated by the flexible networks of the crow NCL that are reminiscent of findings in the PFC of monkeys.

Temporal dynamics of behaviorally relevant category activity with trial time

The task design allowed us to compare length category selectivity during the visual encoding phase (sample phase) and during memorization (delay phase). During the two-category task, more category-selective neurons, with stronger selectivity, were observed during the delay compared with the sample period. Moreover, behaviorally relevant activity clusters (derived from state space) based on the population of category-selective neurons only emerged during the delay period but were largely absent during the sample period in both two-category and three-category tasks. In addition, activity differences between length categories (as measured by correlation measures) collapsed during the delay period in error trials in both two-category and three-category tasks. These results suggest that category-indifferent neural responses rendered the crows' error-proneness.

Collectively, our data argue that the activity of NCL neurons is relevant for the crows' categorization performance, confirming previous findings during numerical categorization.[47,48] Moreover, the observed time course of category selectivity indicates that the initial sensory and largely category-void encoding of length stimuli became dynamically restructured along trial time. Direct sensory length information in conjunction with conceptual information (i.e., short, medium, and long categories) retrieved from long-term memory sculptured NCL activity until the crows have access to categorical representations at the end of the delay period when they need this data to solve the task.

Mechanism of categorization

Mechanistically, the emergence and shaping of categorical tuning could be implemented via broad inhibitory mechanisms.[58] To pin

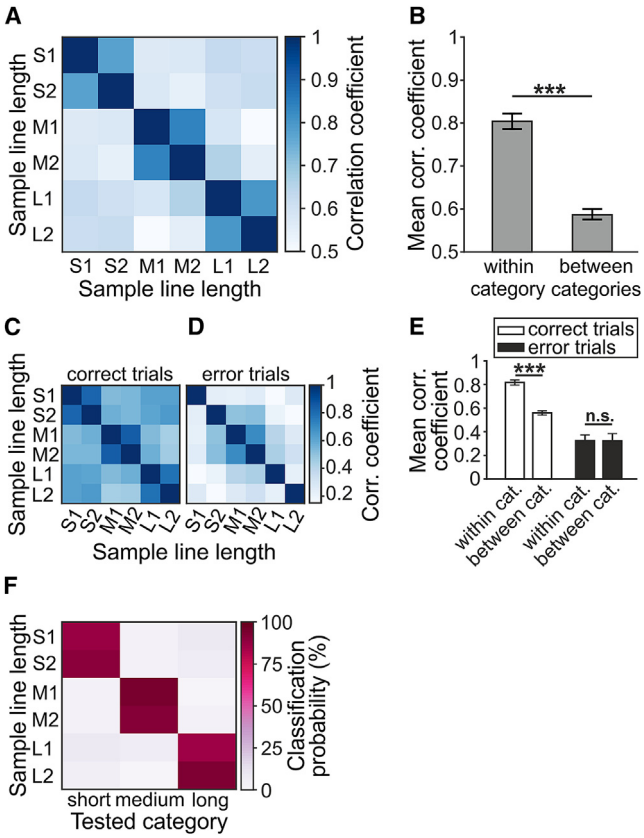


Figure 7. Correlated activity to pairs of stimuli and classification probability of an SVM classifier for the entire neuronal population in the three-category task

(A) Correlation matrix comparing the neuronal activity during the last 600 ms of the delay. Layout as in Figure 4A.

(B) Mean correlation coefficients across all comparisons of stimuli within the same category and of different categories, respectively. Error bars represent SEM. ***: $p < 0.001$.

(C and D) Correlation coefficients in correct (C) and error trials (D) for a subset of the neuronal population for which error trials could be analyzed.

(E) Mean correlation coefficients across comparisons of stimuli within and between categories in correct (left columns) and error trials (right columns). Error bars represent SEM. ***: $p < 0.001$.

(F) Performance of SVM classifier trained on firing rates of the entire neuron population during the last 600 ms of the delay. Layout as in Figure 4F.

down inhibitory mechanisms, the major pallial cell types, putative excitatory projection cells and inhibitory interneurons, have been identified by means of waveform analyses of intra- and extracellularly recorded action potentials.[59–62] Waveform analyses and segregation of putative excitatory projection and inhibitory interneurons were recently also accomplished in NCL neurons of crows discriminating numerosities. It turned out that putative inhibitory interneurons showed stronger stimulus-evoked responses, shorter response latencies, and broader numerosity tuning compared with putative projection neurons.[24] In addition, nearby and functionally coupled putative excitatory projection neurons were synchronously excited and exhibited similar numerosity tuning, whereas coupled putative inhibitory interneurons and projection neurons inhibited each other's firing and showed inverse tuning relative to each other.[24] These data suggest an inhibitory feedforward mechanism for the shaping of neurons tuned to numerical categories in the crow NCL.[24] Such a microcircuit ensures that only projection cells that respond to the correct category remain active and control the animal's response.

Category selectivity arising through reinforcement learning

The length categories applied in the current task—first, short vs. long, and later, short, medium, and long—had no congenital origin and needed to be learned by the crows over time as the result of trial-and-error reinforcement learning. Reinforcement learning based on reward can refine functional connectivity between neurons[63] and typically relies on dopamine signals[64,65]—reward prediction error signals arising from the dopamine system modulate reward-dependent plasticity in primates.[66] Similar processes may be at work in birds learning to categorize, as reward prediction errors have also been observed in the avian NCL, a pallial brain area that is characterized by strong dopaminergic innervation.[33,35,67] According to a cortical circuit model designed for neuronal category learning, weak but systematic correlations between trial-to-trial fluctuations of the firing rates and the accompanying reward after appropriate behavioral choices generate neurons that gradually become category selective.[68] In this model, initially nonselective neurons that show fluctuations that correlate with behavioral outcome developed categorical tuning. Therefore, when a crow learns to respond appropriately to length categories in order to receive a reward, such a mechanism might suffice to produce category-selective NCL neurons from originally untuned neurons. Alternatively, the NCL may contain a special set of malleable category-tuned neurons that change their boundaries with experience. In a previous study, we reported that association learning exclusively recruited NCL neurons that already represented previously established associations.[69] Translated to categories, learning could cause the same pool of neurons to respond to new category boundaries applied to the same length stimulus space.

Category representations in crow NCL vs. primate PFC

The avian NCL is often said to be a functional equivalent of the primate PFC. A comparison of the current data in the crow NCL adds to this functional resemblance in the realm of learned categorization. In primates, behaviorally relevant representations of learned categories have been studied extensively in the PFC using delayed match-to-category tasks. In a seminal series of experiments, macaques were trained to categorize morphed visual stimuli into arbitrary cat and dog categories.[37–40,53,70] As expected for category selectivity, a large proportion of PFC neurons encoded category information by exhibiting significant activity differences between cat and dog categories but similar responses to stimuli within each category.[37–40,53,70] Moreover, although the monkeys learned new category boundaries within the same stimulus space, PFC neurons changed selectivity to now encode the new category boundary, indicating the malleability of the PFC in representing acquired categories.[37–40,53,70]

Beyond perceptual categories (such as cats and dogs), the primate PFC is also equipped to represent more abstract learned spatial categories comparable to those tested here in crows. In the past, representations of abstract magnitude, such as the

absolute[71,72] and relative line length,[73,74] the absolute[75] and relative spatial distance,[76,77] or numerical quantity[78,79] have been reported in the macaque PFC. In one study, monkeys had to learn to classify spatial proportions, i.e., the relation between the variable lengths of two horizontal lines, with proportions ranging from 1:4 and 2:4 to 3:4 and 4:4.[80] Here, PFC neurons showed categorical proportion tuning to the four different proportion categories, very similar to the three length categories short, medium, and long reported here in crows.

These similarities in the flexibility of telencephalic associative brain areas to represent abstract learned categories are remarkable in the face of independent evolution of these brain areas in mammalian and avian lineages.[81] Compared with the mammalian neocortex, the avian telencephalic integration centers originate from different pallial territories during embryology,[82] show distinct neural architectures,[83] and have evolved classes of excitatory and inhibitory pallial neurons that have no counterpart in the mammalian neocortex.[84–86] Despite all this independent brain evolution, crows and monkeys seem to be equipped with equivalent neuronal circuits that can flexibly represent abstract learned magnitude categories.[19,87]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Subjects
- METHOD DETAILS
  - Apparatus
  - Behavioral protocol
  - Stimuli
  - Surgery and neurophysiological recordings
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral analysis
  - Neuronal analysis

ACKNOWLEDGMENTS

This work was supported by DFG grants NI 618/7-1 and NI 618/11-1 to A.N.

AUTHOR CONTRIBUTIONS

A.N. and L.W. designed the study, interpreted the data, and wrote the manuscript. L.W. performed experiments and analyzed the data. All authors gave final approval for publication.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 21, 2022
Revised: April 3, 2023
Accepted: April 7, 2023
Published: May 2, 2023

REFERENCES

1. Miller, E.K., Nieder, A., Freedman, D.J., and Wallis, J.D. (2003). Neural correlates of categories and concepts. Curr. Opin. Neurobiol. *13*, 198–203.
2. Seger, C.A., and Miller, E.K. (2010). Category learning in the brain. Annu. Rev. Neurosci. *33*, 203–219.
3. Baugh, A.T., Akre, K.L., and Ryan, M.J. (2008). Categorical perception of a natural, multivariate signal: mating call recognition in túngara frogs. Proc. Natl. Acad. Sci. USA *105*, 8985–8988.
4. Wyttenbach, R.A., May, M.L., and Hoy, R.R. (1996). Categorical perception of sound frequency by crickets. Science *273*, 1542–1544.
5. Ehret, G., and Haack, B. (1981). Categorical perception of mouse pup ultrasound by lactating females. Naturwissenschaften *68*, 208–209.
6. Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. Eur. J. Neurosci. *11*, 1223–1238.
7. Cheney, D.L., and Seyfarth, R.M. (1990). How Monkeys See the World (University of Chicago Press).
8. Potvin, D.A., Ratnayake, C.P., Radford, A.N., and Magrath, R.D. (2018). Birds learn socially to recognize heterospecific alarm calls by acoustic association. Curr. Biol. *28*, 2632–2637.e4.
9. Lazareva, O.F., and Wasserman, E.A. (2017). Categories and concepts in animals. In Learning and Memory: a Comprehensive Reference, *119* (Elsevier), pp. 111–139.
10. Mansouri, F.A., Freedman, D.J., and Buckley, M.J. (2020). Emergence of abstract rules in the primate brain. Nat. Rev. Neurosci. *21*, 595–610.
11. Yin, P., Strait, D.L., Radtke-Schuller, S., Fritz, J.B., and Shamma, S.A. (2020). Dynamics and hierarchical encoding of non-compact acoustic categories in auditory and frontal cortex. Curr. Biol. *30*, 1649–1663.e5.
12. Herrnstein, R.J., and Loveland, D.H. (1964). Complex visual concept in the pigeon. Science *146*, 549–551.
13. Watanabe, S., Sakamoto, J., and Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. J. Exp. Anal. Behav. *63*, 165–174.
14. Soto, F.A., and Wasserman, E.A. (2014). Mechanisms of object recognition: what we have learned from pigeons. Front. Neural Circuits *8*, 122.
15. Scarf, D., Boy, K., Uber Reinert, A., Devine, J., Güntürkün, O., and Colombo, M. (2016). Orthographic processing in pigeons (Columba livia). Proc. Natl. Acad. Sci. USA *113*, 11272–11276.
16. Huber, L., and Aust, U. (2017). Mechanisms of perceptual categorization in birds. In Avian Cognition, C. ten Cate, and S. Healy, eds. (Cambridge University Press), pp. 208–228.
17. Güntürkün, O., Koenen, C., Iovine, F., Garland, A., and Pusch, R. (2018). The neuroscience of perceptual categorization in pigeons: A mechanistic hypothesis. Learn. Behav. *46*, 229–241.
18. Anderson, C., Parra, R.S., Chapman, H., Steinemer, A., Porter, B., and Colombo, M. (2020). Pigeon nidopallium caudolaterale, entopallium, and mesopallium ventrolaterale neural responses during categorisation of Monet and Picasso paintings. Sci. Rep. *10*, 15971.
19. Pusch, R., Clark, W., Rose, J., and Güntürkün, O. (2023). Visual categories and concepts in the avian brain. Anim. Cogn. *26*, 153–173.
20. Moll, F.W., and Nieder, A. (2014). The long and the short of it: rule-based relative length discrimination in carrion crows, Corvus corone. Behav. Proc. *107*, 142–149.
21. Brecht, K.F., Wagener, L., Ostojić, L., Clayton, N.S., and Nieder, A. (2017). Comparing the face inversion effect in crows and humans. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *203*, 1017–1027.
22. Wagener, L., and Nieder, A. (2020). Categorical auditory working memory in crows. iScience *23*, 101737.
23. Rinnert, P., and Nieder, A. (2021). Neural code of motor planning and execution during goal-directed movements in crows. J. Neurosci. *41*, 4060–4072.
24. Ditz, H.M., Fechner, J., and Nieder, A. (2022). Cell-type specific pallial circuits shape categorical tuning responses in the crow telencephalon. Commun. Biol. *5*, 269.

25. Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. Nat. Neurosci. *3*, 1199–1204.

26. Clark, W.J., and Colombo, M. (2020). The functional architecture, receptive field characteristics, and representation of objects in the visual network of the pigeon brain. Prog. Neurobiol. *195*, 101781.

27. Butler, A.B., Reiner, A., and Karten, H.J. (2011). Evolution of the amniote pallium and the origins of mammalian neocortex. Ann. N. Y. Acad. Sci. *1225*, 14–27.

28. Azizi, A.H., Pusch, R., Koenen, C., Klatt, S., Bröker, F., Thiele, S., Kellermann, J., Güntürkün, O., and Cheng, S. (2019). Emerging category representation in the visual forebrain hierarchy of pigeons (*Columba livia*). Behav. Brain Res. *356*, 423–434.

29. Clark, W., Chilcott, M., Azizi, A., Pusch, R., Perry, K., and Colombo, M. (2022). Neurons in the pigeon visual network discriminate between faces, scrambled faces, and sine grating images. Sci. Rep. *12*, 589.

30. Mogensen, J., and Divac, I. (1982). The prefrontal "cortex" in the pigeon. Behavioral evidence. Brain Behav. Evol. *21*, 60–66.

31. Güntürkün, O. (2005). The avian "prefrontal cortex" and cognition. Curr. Opin. Neurobiol. *15*, 686–693.

32. Nieder, A. (2017). Inside the corvid brain—probing the physiology of cognition in crows. Curr. Opin. Behav. Sci. *16*, 8–14.

33. von Eugen, K., Tabrik, S., Güntürkün, O., and Ströckens, F. (2020). A comparative analysis of the dopaminergic innervation of the executive caudal nidopallium in pigeon, chicken, zebra finch, and carrion crow. J. Comp. Neurol. *528*, 2929–2955.

34. Güntürkün, O., von Eugen, K., Packheiser, J., and Pusch, R. (2021). Avian pallial circuits and cognition: a comparison to mammals. Curr. Opin. Neurobiol. *71*, 29–36.

35. Kersten, Y., Friedrich-Müller, B., and Nieder, A. (2021). A histological study of the song system of the carrion crow (Corvus corone). J. Comp. Neurol. *529*, 2576–2595.

36. Kersten, Y., Friedrich-Müller, B., and Nieder, A. (2022). A brain atlas of the carrion crow (Corvus corone). J. Comp. Neurol. *530*, 3011–3038.

37. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science *291*, 312–316.

38. Roy, J.E., Riesenhuber, M., Poggio, T., and Miller, E.K. (2010). Prefrontal cortex activity during flexible categorization. J. Neurosci. *30*, 8519–8528.

39. Roy, J.E., Buschman, T.J., and Miller, E.K. (2014). PFC neurons reflect categorical decisions about ambiguous stimuli. J. Cogn. Neurosci. *26*, 1283–1291.

40. Cromer, J.A., Roy, J.E., Buschman, T.J., and Miller, E.K. (2011). Comparison of primate prefrontal and premotor cortex neuronal activity during visual categorization. J. Cogn. Neurosci. *23*, 3355–3365.

41. Kirsch, J.A., Vlachos, I., Hausmann, M., Rose, J., Yim, M.Y., Aertsen, A., and Güntürkün, O. (2009). Neuronal encoding of meaning: establishing category-selective response patterns in the avian "prefrontal cortex.". Behav. Brain Res. *198*, 214–223.

42. Veit, L., and Nieder, A. (2013). Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. Nat. Commun. *4*, 2878.

43. Moll, F.W., and Nieder, A. (2015). Cross-modal associative mnemonic signals in crow endbrain neurons. Curr. Biol. *25*, 2196–2201.

44. Veit, L., Pidpruzhnykova, G., and Nieder, A. (2015). Associative learning rapidly establishes neuronal representations of upcoming behavioral choices in crows. Proc. Natl. Acad. Sci. USA *112*, 15208–15213.

45. Ditz, H.M., and Nieder, A. (2015). Neurons selective to the number of visual items in the corvid songbird endbrain. Proc. Natl. Acad. Sci. USA *112*, 7827–7832.

46. Ditz, H.M., and Nieder, A. (2016). Sensory and working memory representations of small and large numerosities in the crow endbrain. J. Neurosci. *36*, 12044–12052.

47. Ditz, H.M., and Nieder, A. (2020). Format-dependent and format-independent representation of sequential and simultaneous numerosity in the crow endbrain. Nat. Commun. *11*, 686.

48. Kirschhock, M.E., Ditz, H.M., and Nieder, A. (2021). Behavioral and neuronal representation of numerosity zero in the crow. J. Neurosci. *41*, 4889–4896.

49. Wagener, L., Loconsole, M., Ditz, H.M., and Nieder, A. (2018). Neurons in the endbrain of numerically naive crows spontaneously encode visual numerosity. Curr. Biol. *28*, 1090–1094.e4.

50. Kobylkov, D., Mayer, U., Zanon, M., and Vallortigara, G. (2022). Number neurons in the nidopallium of young domestic chicks. Proc. Natl. Acad. Sci. USA *119*, e2201039119.

51. Nasr, K., Viswanathan, P., and Nieder, A. (2019). Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. Sci. Adv. *5*, eaav7903.

52. Nasr, K., and Nieder, A. (2021). Spontaneous representation of numerosity zero in a deep neural network for visual object recognition. Iscience *24*, 103301.

53. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2002). Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. J. Neurophysiol. *88*, 929–941.

54. Caliński, T., and Harabasz, J. (1974). A dendrite method for cluster analysis. Commun. Stat. Theor. Methods *3*, 1–27.

55. Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. B *63*, 411–423.

56. Lloyd, S. (1982). Least squares quantization in PCM. IEEE Trans. Inf. Theor. *28*, 129–137.

57. Chang, C.C., and Lin, C.J. (2011). LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. *2*, 1–27.

58. Mahajan, N.R., and Mysore, S.P. (2022). Donut-like organization of inhibition underlies categorical neural responses in the midbrain. Nat. Commun. *13*, 1680.

59. Spiro, J.E., Dalva, M.B., and Mooney, R. (1999). Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. J. Neurophysiol. *81*, 3007–3020.

60. Kosche, G., Vallentin, D., and Long, M.A. (2015). Interplay of inhibition and excitation shapes a premotor neural sequence. J. Neurosci. *35*, 1217–1227.

61. Calabrese, A., and Woolley, S.M.N. (2015). Coding principles of the canonical cortical microcircuit in the avian brain. Proc. Natl. Acad. Sci. USA *112*, 3517–3522.

62. Yanagihara, S., and Yazaki-Sugiyama, Y. (2016). Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. Nat. Commun. *7*, 11946.

63. Law, C.-T., and Gold, J.I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. Nat. Neurosci. *12*, 655–663.

64. Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. Physiol. Rev. *95*, 853–951.

65. Herpers, J., Arsenault, J.T., Vanduffel, W., and Vogels, R. (2021). Stimulation of the ventral tegmental area induces visual cortical plasticity at the neuronal level. Cell Rep. *37*, 109998.

66. Stauffer, W.R., Lak, A., Yang, A., Borel, M., Paulsen, O., Boyden, E.S., and Schultz, W. (2016). Dopamine neuron-specific optogenetic stimulation in rhesus macaques. Cell *166*, 1564–1571.e6.

67. Wynne, B., and Güntürkün, O. (1995). Dopaminergic innervation of the telencephalon of the pigeon (Columba livia): A study with antibodies against tyrosine hydroxylase and dopamine. J. Comp. Neurol. *357*, 446–464.

68. Engel, T.A., Chaisangmongkon, W., Freedman, D.J., and Wang, X.J. (2015). Choice-correlated activity fluctuations underlie learning of neuronal category representation. Nat. Commun. *6*, 6454.

69. Veit, L., Pidpruzhnykova, G., and Nieder, A. (2017). Learning recruits neurons representing previously established associations in the corvid endbrain. J. Cogn. Neurosci. *29*, 1712–1724.

70. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. *23*, 5235–5246.

71. Tudusciuc, O., and Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. Proc. Natl. Acad. Sci. USA *104*, 14513–14518.

72. Tudusciuc, O., and Nieder, A. (2009). Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. J. Neurophysiol. *101*, 2984–2994.

73. Eiselt, A.K., and Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. J. Neurosci. *33*, 7526–7534.

74. Eiselt, A.K., and Nieder, A. (2016). Single-cell coding of sensory, spatial and numerical magnitudes in primate prefrontal, premotor and cingulate motor cortices. Exp. Brain Res. *234*, 241–254.

75. Marcos, E., Tsujimoto, S., and Genovesio, A. (2017). Independent coding of absolute duration and distance magnitudes in the prefrontal cortex. J. Neurophysiol. *117*, 195–203.

76. Genovesio, A., Tsujimoto, S., and Wise, S.P. (2011). Prefrontal cortex activity during the discrimination of relative distance. J. Neurosci. *31*, 3968–3980.

77. Genovesio, A., Tsujimoto, S., and Wise, S.P. (2012). Encoding goals but not abstract magnitude in the primate prefrontal cortex. Neuron *74*, 656–662.

78. Viswanathan, P., and Nieder, A. (2015). Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. *25*, 1259–1269.

79. Ramirez-Cardenas, A., Moskaleva, M., and Nieder, A. (2016). Neuronal representation of numerosity zero in the primate parieto-frontal number network. Curr. Biol. *26*, 1285–1294.

80. Vallentin, D., and Nieder, A. (2008). Behavioral and prefrontal representation of spatial proportions in the monkey. Curr. Biol. *18*, 1420–1425.

81. Nieder, A. (2021). The evolutionary history of brains for numbers. Trends Cogn. Sci. *25*, 608–621.

82. Briscoe, S.D., and Ragsdale, C.W. (2019). Evolution of the chordate telencephalon. Curr. Biol. *29*, R647–R662.

83. Jarvis, E.D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D.J., Shimizu, T., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. Nat. Rev. Neurosci. *6*, 151–159.

84. Nomura, T., Ohtaka-Maruyama, C., Kiyonari, H., Gotoh, H., and Ono, K. (2020). Changes in Wnt-dependent neuronal morphology underlie the anatomical diversification of neocortical homologs in amniotes. Cell Rep. *31*, 107592.

85. Tosches, M.A., Yamawaki, T.M., Naumann, R.K., Jacobi, A.A., Tushev, G., and Laurent, G. (2018). Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. Science *360*, 881–888.

86. Colquitt, B.M., Merullo, D.P., Konopka, G., Roberts, T.F., and Brainard, M.S. (2021). Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. Science *371*, eabd9704.

87. Nieder, A. (2021). Consciousness without cortex. Curr. Opin. Neurobiol. *71*, 69–76.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Corvus corone	University of Tübingen, Institute of Neurobiology	crow 1, crow 2
Software and algorithms		
NIMH Cortex	National Institute of Mental Health	c595; https://www.nimh.nih.gov/labs-at-nimh/research-areas/clinics-and-labs/ln/shn/software-projects.shtml
MAP Data Acquisition System	Plexon	https://plexon.com/
MATLAB R2019a	MathWorks	https://www.mathworks.com
Other		
Dental Cement	Heraeus	Paladur, ISO 20795, CE 0197
Microdrives	Animal Physiology Unit	Custom fabrication
Electrodes	Alpha Omega LTD	Cat.#: 366-130620-00 www.alphaomega-eng.com

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Andreas Nieder (andreas.nieder@uni-tuebingen.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data reported in this paper will be shared by the lead contact upon request. This paper does not report original code. Any additional information required to re-analyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects

Two hand-raised adult male carrion crows (*Corvus corone*) from the institute's breeding facility were used. The crows were 3 and 6 years old. They were housed in an indoor aviary in social groups. During the experiment, the crows were on a controlled feeding protocol and received their daily amount of food as reward during training and recording or, if necessary, after the sessions. Water was available *ad libitum* during the experiments and in the aviary. All procedures were carried out according to the guidelines for animal experimentation and approved by the responsible national authorities, the Regierungspräsidium Tübingen, Germany.

METHOD DETAILS

Apparatus

The experiment was conducted in a darkened operant conditioning chamber. The crows were placed on a perch in front of a 15" touchscreen monitor (ART development MT1599-BS and ART development PS-150, respectively). Viewing distance to the screen was 14 cm. The touchscreen was used only for stimulus presentation, as the crows responded by head movements.

Behavior and response of the crows were controlled by an infrared reflexive light system which was located above the crows and registered the position of a reflector foil attached on top of the crows' head. The crows initiated trials by keeping their heads still in the center position in front of the touchscreen monitor and were required to keep the head within this position throughout the trial until the target stimulus appeared.

The crows reported the detection of the target stimulus by briefly moving their heads ('nodding'), which was again automatically detected by the infrared reflexive light system. With every correct answer, a food reward (either birdseed pellets or mealworms

(*Tenebrio molitor* larvae)) was given by the briefly illuminated feeder below the touchscreen monitor. Auditory feedback was provided by speakers (Lasmex S-03) located behind the touchscreen monitor. We used the CORTEX system (National Institute of Mental Health) to run the experiment and collect behavioral data.

Behavioral protocol

The crows were trained to group horizontal line stimuli into learned categories according to their length (Figure 1A). The crows initiated a trial by positioning their heads facing the screen whenever the go-stimulus (small white cross, 2x2 dva (degree of visual angle)) was shown. A click sound indicated that the correct position had been entered and the go-stimulus turned briefly (for 60ms) into a circle before it vanished. This head position had to be maintained throughout the trial until the test phase. Premature head movements aborted the ongoing trial which was then discarded.

After a 600 ms pre-sample phase in which only the grey background circle was shown, the sample stimulus was presented for 500 ms. Then the screen returned to the grey background circle for a delay of 1000 ms. In the subsequent test phase, the first test stimulus (Test 1) appeared for max. 900 ms. In 50% of the trials, Test 1 displayed a line length that belonged to the same category as the sample stimulus (i.e. "match"). In the other half of the trials, Test 1 was not member of the same category as the sample stimulus (i.e. "nonmatch"). The chance level of Test 1 or Test 2 being a match was therefore 50%. The crow indicated a category "match" by instantaneously nodding, i.e. moving its head out of the monitored center position.

A "nonmatch" stimulus required the crow to maintain head position and to refrain from responding until the subsequent second test stimulus (Test 2) appeared which always belonged to the same category as the sample (a "match"). A correct response to a "match" stimulus (either Test 1 or Test 2) led to a reward for the crow. A response to a "nonmatch" stimulus or no response to either Test 1 or Test 2 aborted the trial and was considered as error trial and not rewarded. Within each session, all behaviorally relevant parameters (i.e. sample line length, stimulus sets and match/nonmatch trials) were balanced and pseudo-randomly interleaved.

Stimuli

The line length stimuli were generated using MATLAB software. They consisted of a horizontal black line at random position within a grey background circle (Figure 1B). Six different line lengths were used that were assigned to either two (first experiment) or three different length categories (second experiment). The lengths were consecutive multiples of lengths of 2.6 dva ranging from 3.3 dva (shortest line) to 16.3 dva (longest line).

To ensure that the crows categorized length rather than the area or thickness of the lines, we used two interleaved sets of stimuli in each session, a standard stimulus protocol and a control protocol. In the standard protocol, the thickness of the lines varied randomly between 0.4 and 2.0 dva. In the control protocol, the black area of each line was kept constant to 6.5 dva^2 across the different lengths, with the thickness of the shortest line being always 2 dva and thickness of the longest line always 0.4 dva. In addition, the sample and the test images within a trial were never identical. New stimulus sets were generated for each session to prevent the crows from memorizing visual patterns.

First, we trained both crows on the two-category task (Experiment 1). For that, one category boundary divided the six different line lengths into two groups of three line lengths each. The "short" category included the lengths S1, S2 and S3; the "long" category included L1, L2 and L3. Thereafter, we retrained crow 1 on the three length categories "short", "medium", and "long" (Experiment 2). To that aim, we divided the line lengths into three categories of two line lengths each. The absolute lengths of the lines remained unaffected, i.e. the physical appearance of the stimuli stayed the same, only the category membership changed. The "short" category still included the lengths S1 and S2, and the "long" category still contained the lengths L2 and L3 (now renamed as L1 and L2, respectively). The former lengths S3 and L1 constituted the new "medium" category and in this context were renamed as M1 and M2, respectively.

Surgery and neurophysiological recordings

The surgeries were performed while the animal was under general anesthesia with a mixture of ketamine (50 mg/kg) and xylazine (5 mg/kg). The animals were placed in a stereotaxic holder. We targeted the dorsal part of the *nidopallium caudolaterale* (NCLd)[33,35,36] by performing a craniotomy at 5 mm anterior-posterior and 13 mm medio-lateral on the right hemisphere. Two manual micro drives containing four electrodes each (2 MΩ, Alpha Omega Co.) and a miniature connector for the head stage were implanted. After the surgery, the crows received an analgesic. A small holder for attaching the reflector of the light barrier and head-tracking system, respectively, had been already implanted under the same conditions.

Each recording session started with adjusting the electrodes until a proper neuronal signal (of at least 3:1 signal to noise) was detected on at least one channel (see also Figures 4A and 4B in Veit and Nieder,[42] for an example recording trace). Neurons were not preselected in the involvement of the task. Signal amplification, filtering, and digitizing of spike waveforms was performed using the Plexon MAP system (Plexon Inc., Dallas, Texas). Spectral filtering of recordings was accomplished by a combined preamplifier filter (150 Hz–8kHz, 1 pole low-cut, 3 pole high-cut) and main filter (250 Hz, 2-pole, low-cut filter). Amplitude amplifications were set individually for different channels in the range of ca. 20,000x gain. Spike waveforms were sampled at a frequency of 40 kHz (one entry every 25 μs) for a duration of 800 μs. Plexon's offline Sorter was used to manually offline sort spikes into single-unit waveforms by applying mainly principal component analysis. We recorded 52 sessions in crow 1 and 55 sessions in crow 2 performing the two-category task, and 58 sessions in crow 1 performing the three-category task.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analysis

The behavioral performance was measured as the percentage correct categorization of the sample line lengths, i.e. of how often the crows correctly judged that the line lengths of the test stimuli (either Test 1 or Test 2) belonged to the same category as the length of the sample stimulus. For each session we used a binomial test to verify that the ratio of correct answers was above the 50% chance level (for both stimulus sets separately and combined).

Neuronal analysis

Analyses of category-selective neurons

For the neuronal analyses, we included all neurons which had an average firing rate of at least 1 Hz during the overall trial and were recorded for at least 10 correct trials for each sample line length. First, we identified category-selective neurons. To that aim, we analyzed the activity of the neurons in sliding windows of 200 ms length which were advanced by 10 ms steps, starting at sample onset and ending 100 ms after delay offset (to account for the neurons' response latency). In each window, we performed a combination of two statistical tests on the neurons' firing rates to determine category-selective neurons: First, we calculated a two-factor ANOVA with category and stimulus protocol as main factors (criterion $P < 0.01$) to determine across-category selectivity. Neurons were selected that showed a significant main effect for category but no effect for main factor protocol or interaction between main factors. Second, we additionally calculated Kruskal-Wallis tests to explore differential activity to within-category line length. A category selective neuron was supposed to show no response differences across stimuli within each category (criterion $P \geq 0.05$). Neurons recorded in the two-category task were tested with two Kruskal-Wallis tests (for category ''short'': neuronal responses to sample stimuli S1 vs. S2 vs. S3, and for category ''long'': sample stimuli L1 vs. L2 vs. L3). Neurons recorded in the three-category task were tested with three Kruskal-Wallis testes (''short'': neuronal responses to sample stimuli S1 vs. S2, ''medium'': sample stimuli M1 vs. M2, ''long'': sample stimuli L1 vs. L2). If a neuron fulfilled all of these criteria, i.e., was selective to category in the ANOVA but unselective in the Kruskal-Wallis tests over at least 11 consecutive windows (i.e. 300 ms in total), it was termed 'category-selective'.

A category selective interval was assigned to the sample period if it started no later than 100 ms after sample offset. Later occurring selective intervals were assigned to the delay period. If a neuron had more than one selective interval in the sample or delay period, respectively, only the one with the smallest *P*-value for factor category according to the ANOVA was used for later analyses. The preferred length category within a selective interval was defined as the category which contained the stimulus eliciting the highest mean firing rate. To calculate the average neuronal activity of the category-selective neurons within the selective intervals (separately for the sample and delay phase), each neuron's mean firing rates to the six different line lengths were normalized by setting the highest firing rate to 100% and the lowest to 0%. These were then arranged according to their distance from the category boundary and averaged across all neurons. Due to this definition of category selectivity, the preferred category is expected to be encoded by normalized firing rates that are larger than 50% and maximally 100% (and vice versa for the non-preferred category). However, the normalized firing rates to individual line length stimuli within a category are not part of the definition, which is why this measure and the derived values are suitable to explore neuronal encoding similarities to individual stimuli within categories and coding differences to stimuli between categories, particularly at the category boundary.

We calculated a category index from the average firing rates of the category-selective neurons to the six different sample line lengths analogous to Freedman et al.[37,53] The ''between category difference'' (BCD) was defined as the absolute difference between the average firing rate of a neuron to the sample stimuli adjacent to the category boundary (i.e. S3 vs. L1). For the ''within category difference'' (WCD), we calculated the firing rate differences between all neighboring sample stimuli (to keep the distance between the compared stimuli constant) that belong to the same category (i.e. S1 vs. S2, S2 vs. S3, L1 vs. L2 and L2 vs. L3) and then took the mean of these. From these firing rate differences, we calculated the category index by subtracting the WCD from BCD and dividing it by their sum:

$$\text{Catergory index} = (\text{BCD} - \text{WCD})/(\text{BCD} + \text{WCD}).$$

It resulted in values between -1 and 1 with positive values indicating a higher difference between the firing rates to two sample stimuli of different categories than between stimuli within the same category. Shifts of the category index distributions relative to value 0 were tested with a one-sample *t*-test.

We used principle component analysis (PCA), to investigate how the activity of the category-selective neurons evolved during the course of a trial. For the later purpose of clustering, we included all category-selective neurons with at least 30 correct trials for each line length (two-category task: 54 out of 65 sample-selective neurons and 80 of 86 delay-selective neurons; three-category task: 41 out of 47 sample-selective neurons and 88 out of 93 delay-selective neurons). The neuronal activity in response to a certain stimulus at a certain time point is represented as a *n*-dimensional vector in *n*-dimensional space, with each dimension corresponding to one single neuron.

We used PCA to reduce the dimensionality of the population activity while capturing most of the information. For that, the neuronal data for each trial was smoothed by a 200 ms Gauss kernel with a step size of 1 ms, and the mean firing rate to each line length was calculated in bins of 100 ms (advanced in steps of 10 ms) and then neuron-wise *z*-scored. From this data, we created a population of pseudo-simultaneously recorded neurons. We calculated the PC scores using the implemented *pca* function of MATLAB. To

illustrate the trajectories of the change in neuronal activity, we used the first three principle components which formed a three-dimensional subspace. These three principle components explained 47.2% of the neuronal covariance in the sample period, and 40.6% of the covariance in the delay period of the two-category task. In the three-category task, the first three principle components explained 58.0% of the neuronal covariance of the neurons which were category-selective in the sample phase and 54.6% for the neurons which were selective during the delay.

In a next step, we analyzed the sample and delay periods separately. As before, the neuronal activity was first smoothed by a 200 ms Gauss kernel across the entire trial. Then, we calculated the mean firing rate in each trial across a 600 ms time window starting at sample onset and reaching 100 ms into the delay for the neurons which were category-selective in the sample phase. For the neurons that were selective during the delay, we averaged the firing rate across the last 900 ms of the delay. Then, we randomly drew the firing rates of 30 trials for each sample line length within the given analysis interval, z-scored these neuron-wise and calculated the PC scores.

To evaluate the optimal number of clusters, we applied the unsupervised *k-means* clustering algorithm using the first two principle components. We used two different criteria, the gap value and the Calinski-Harabasz index.[54,55] The maximum possible number of clusters was set to six. The Calinski-Harabasz index (also called variance ratio criterion) is a measure of how dense the objects within a cluster are and how well different clusters are separated. The optimal number of clusters is the one which yields the highest value. The gap statistic compares the within cluster variation to its variation expected under the assumption of a reference null distribution.[55] A high gap value for a certain number of clusters indicates a large difference from the uniform distribution. These two measures indicated the optimal number of clusters based on the drawn trials. This was repeated 1000 times with newly drawn trials. After that, we calculated the frequency of how often the different cluster numbers were assigned among across the repetitions.

Population analyses

Population analyses were performed on the entire population of recorded neurons. All neurons with an average firing rate of at least 1 Hz and at least 30 correct trials for each sample line length entered the analysis without any pre-selection for category-selectivity (two-category task: $n = 348$, three-category task: $n = 278$). We used the firing rates within a 600 ms fixed window at the end of the delay (starting 400 ms after sample offset) to capture delay-activity which carries the category information needed to be available for the subsequent test phase and at the same time exclude late sample-related responses.

A correlation matrix was created to visualize the firing rate differences between pairs of stimuli and to detect coding patterns. The strength of the relationship between the firing rates to two stimuli was measured by the correlation coefficient r (deviation from the regression line). For that, the firing rates of each neuron were normalized by subtracting the average baseline firing rate (measured within 300 ms before sample onset across all correct trials) and dividing by the respective standard deviation. The coefficients of each correlation were represented as a tile in the correlation matrix. The tiles along the diagonal from lower left to upper right represent the correlation of the stimuli with themselves ($r = 1.0$). The matrix is symmetric to the diagonal.

For quantification, we calculated the mean of the correlation coefficients for the relationships of stimuli which belong to the same category and for stimuli of different categories. Regarding the data from the two-category task, correlations within the same category were: S1 vs. S2, S1 vs. S3, S2 vs. S3, L1 vs. L2, L2 vs. L3 and L1 vs. L3 and correlations between different categories: S1 vs. L1, S1 vs. L2, S1 vs. L3, S2 vs. L1, S2 vs. L2, S2 vs. L3, S3 vs. L1, S3 vs. L2 and S3 vs. L3. In the three-category task, correlations within the same category were: S1 vs. S2, M1 vs. M2 and L1 vs. L2 and between different categories: S1 vs. M1, S1 vs. M2, S1 vs. L1, S1 vs. L2, S2 vs. M1, S2 vs. M2, S2 vs. L1, S2 vs. L2, M1 vs. L1, M1 vs. L2, M2 vs. L1 and M2 vs. L2. The mean difference between these values was statistically verified by a two-sample t-test.

To test whether the activity of the neuronal population is behaviorally relevant, we calculated the correlation coefficients also in error trials (i.e. when the crow responded to the "nonmatch" stimulus or to neither of the two test stimuli). For that, we used all neurons of the analyzed population which had additionally at least 3 error trials for each sample line length (two-category task: $n = 172$, three-category task: $n = 120$). Baseline activity and respective standard deviation for firing rate normalization was measured in error trials within 300 ms before sample onset. Further analysis was done equally as for correct trials.

Additionally, we tested how well a multi-class support vector machine (SVM) classifier categorizes the firing rates of the recorded neurons. We used the LIBSVM toolbox for MATLAB (version 3.24)[57] with default parameters (multi-class classification, radial basis function as kernel). We performed a 5-fold cross-validation with the firing rates of the neuronal population to the six different sample line lengths within the last 600 ms of the delay. For that we randomly drew the firing rates of 30 trials of each neuron for each sample line length and assigned the true category labels to them. Then the drawn set of firing rates was normalized by z-scoring and then split into five equal groups. The classifier was trained with the firing rates of four fifth of the trials (144 trials of each neuron, sample stimuli were balanced, i.e. 24 trials for each sample stimulus) and then tested with the remaining one fifth (36 trials, i.e. 6 trials for each sample stimulus). This procedure was repeated five times so that each split of firing rates was used once as the test set, resulting in one accuracy value per sample line length (percentage of correctly classified firing rates across the five repetitions). Furthermore, we repeated the 5-fold cross-validation with 30 newly drawn trials for each sample line length 1000 times. The resultant confusion matrix shows the averaged classification probability across the trial samplings for the firing rates of each sample line length.

We finally used the SVM classifier to test whether the activity of the neurons to a single stimulus of each category of the two-category task can be used to predict the correct category of firing rates to the remaining stimuli which were not used for training. To that aim, we used three different sets of training and testing stimuli. First, we trained the classifier using the firing rates to the stimuli S1 and

L3 (most distant from the category boundary) and predicted the category of the stimuli S2, S3, L1 and L2. Next, we used the stimuli S2 and L2 for classifier training and S1, S3, L1 and L3 for the prediction and finally, S3 and L1 (adjacent to the category boundary) were used for training and S1, S2, L2 and L3 as prediction stimuli. For the classifier training and testing, we randomly drew the firing rates of 30 trials per line length and normalized these by z-scoring. Then we calculated the percentage of how often the classifier predicted that the test firing rates belong to the ''short'' category. This was repeated 1000 times with newly drawn firing rates. The results were then averaged across the repetitions.