RESEARCH ARTICLE

COMPARATIVE BEHAVIOR

# Crows "count" the number of self-generated vocalizations

**Diana A. Liao[1]*, Katharina F. Brecht[1], Lena Veit[2], Andreas Nieder[1]***

**Producing a specific number of vocalizations with purpose requires a sophisticated combination of numerical abilities and vocal control. Whether this capacity exists in animals other than humans is yet unknown. We show that crows can flexibly produce variable numbers of one to four vocalizations in response to arbitrary cues associated with numerical values. The acoustic features of the first vocalization of a sequence were predictive of the total number of vocalizations, indicating a planning process. Moreover, the acoustic features of vocal units predicted their order in the sequence and could be used to read out counting errors during vocal production.**

The process of vocal counting, exemplified by reciting "one, two, three," entails uttering a sequence of number words that represent increasing numerical quantities. The biological origins of this symbolic counting, which require both numerical competency and volitional vocal control, remain unknown.

In human ontogeny, toddlers learning to verbally recite the counting list initially use number words not to represent cardinalities but as simple verbal tallies (*1*). When asked "how many?" they produce as many vocalizations as there are objects (e.g., "one, one, one" or "one, two, three" for three) (*2*). Similarly, some animal species convey information essential for survival through differing numbers of vocalizations. Chickadees, for example, scale the number of "dee" notes in their alarm calls with the size of the predator, thereby conveying the magnitude of perceived threat (*3*, *4*). The usage of repeated sounds predates symbolic counting and serves as a nonsymbolic way of keeping track of quantities. In nonhuman animals, the production of more vocal elements is commonly attributed to increased affective arousal of the signaler (*5*–*7*) and is not deliberately controlled. However, a species could significantly enhance goal-directed communication if it could integrate numerical proficiency with voluntary control over vocalizations. We hypothesized that carrion crows, one of the few bird species (*8*, *9*) that possess not only numerical competency (*10*–*12*) but also volitional vocal control (*13*, *14*), can deliberately control the number of produced vocalizations.

We trained three carrion crows (*Corvus corone*) to flexibly produce a variable number of one to four vocalizations in response to visual and auditory cues (Fig. 1A). After the crows initiated a trial, a vocalization cue (visual: colored Arabic numerals; auditory: distinct 500-ms sounds) instructed the productic a specific number of vocalizations (Fig. 1B). The crows had to produce a target number of vocalizations and indicate the end of the vocal sequence by pecking at a confirmation stimulus ("enter key"). Correct trials in which the produced number of vocalizations matched the cued number were followed by a reward; more or fewer vocalizations than the target number were counted as errors and remained unrewarded.

All crows successfully produced the target number of vocalizations (Fig. 1, C and D). Performance for each cue modality (visual/auditory), and to each number (1 to 4) across sessions was significantly above chance (14.29%, two-sided binomial tests, all $P < 0.0001$). The crows exhibited performance effects characteristic for nonsymbolic number estimation (*15*). When incorrect, they usually erred by producing one more or fewer vocalizations, resulting in bell-shaped performance functions centered around the target number illustrating the "numerical distance effect" (Fig. 1, C and D7). Moreover, the crows tended to make more and larger errors if the target

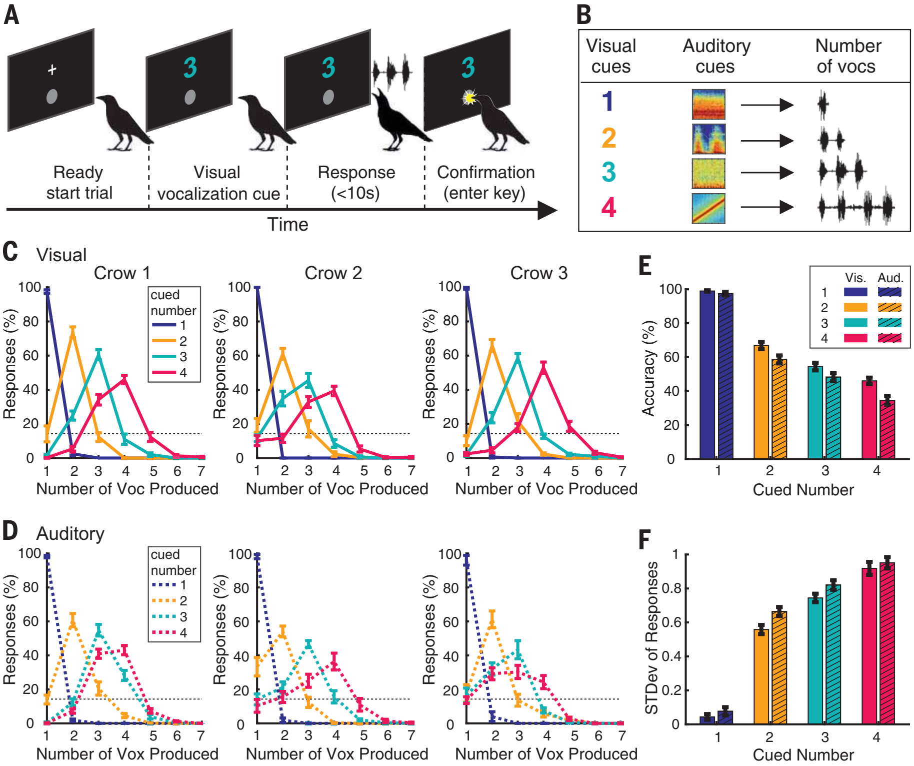

**Fig. 1. Experimental design and performance.** (**A**) Protocol of the vocal production task. In this example, the crow was cued to produce three calls. (**B**) Visual (colored Arabic numerals) and auditory (distinct sounds indicated by spectrograms) cues instructed the crows to produce a certain number of vocalizations. (**C**) Behavioral performance of the three crows to visual cues. The dotted horizontal line indicates chance level (1/7). (**D**) Behavioral performance in regards to auditory cues; layout same as (C). (**E**) Average performance per cued target number for visual and auditory cues. (**F**) Widths (standard deviation) of performance functions displayed in (C) and (D). Error bars represent standard error of the mean (SEM).

---

[1]Animal Physiology, Institute of Neurobiology, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany. [2]Neurobiology of Vocal Communication, Institute of Neurobiology, University of Tübingen Auf der Morgenstelle 28, 72076 Tübingen, Germany.
*Corresponding author. Email: andreas.nieder@uni-tuebingen.de (A.N.); diana.a.liao@gmail.com (D.A.L.)

number increased [mixed-effects generalized linear model (GLM), main effect “cue number”: $P < 0.0001$; Fig. 1E]. This “numerical size effect” resulted in broader performance functions with increasing target numbers (mixed-effects GLM, main effect “cue number”: $P < 0.0001$; Fig. 1F). The modality of the vocalization cues had no significant effect on the crows’ behavior (mixed-effects GLM, main effect “cue modality”: accuracy: $P = 0.627$; standard deviation: $P = 0.347$; Fig. 1, E and F), indicating the formation of an abstract numerical concept to guide vocal production.

How do crows control the production of a specific number of vocalizations? They could assess the number of vocalizations on-the-fly during the process of production, in which case the vocal reaction time (time interval between the cue and the first vocalization) is expected to be equal for all numbers (Fig. 2A). Alternatively, they could plan the future number of vocalizations in advance and show longer reaction times for more numerous vocalizations as the preparation of more vocalizations requires greater cognitive demand and premotor coordination of the vocal apparatus (*16*, *17*). Not only were the crows’ reaction times much longer than any subsequent intervocalization intervals (mixed-effects GLM, main effect “trial interval”: $P < 0.0001$; Fig. 2B),

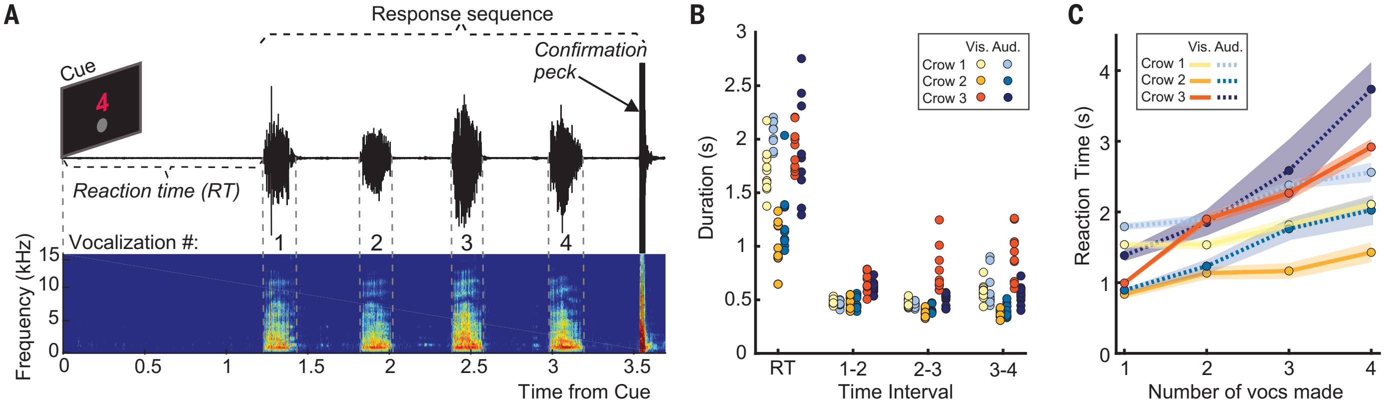

**Fig. 2. Relationship between reaction time and number of impending vocalizations.** (**A**) Time intervals for the vocal production sequence. (Top) Timeline of an example trial for four vocalizations. Reaction time (RT) extends from the monitor display (left) to vocalization 1 (oscillogram 1). Vocalizations 2 to 4 are followed by the noise of the “enter” peck. (Bottom) Temporally aligned spectrogram of the crow’s produced 4 vocalizations. (**B**) Session durations of pertinent time intervals [as shown in (A)]. (**C**) Reaction times to produce the first vocalization after instruction cues for the four target numbers of vocalizations.

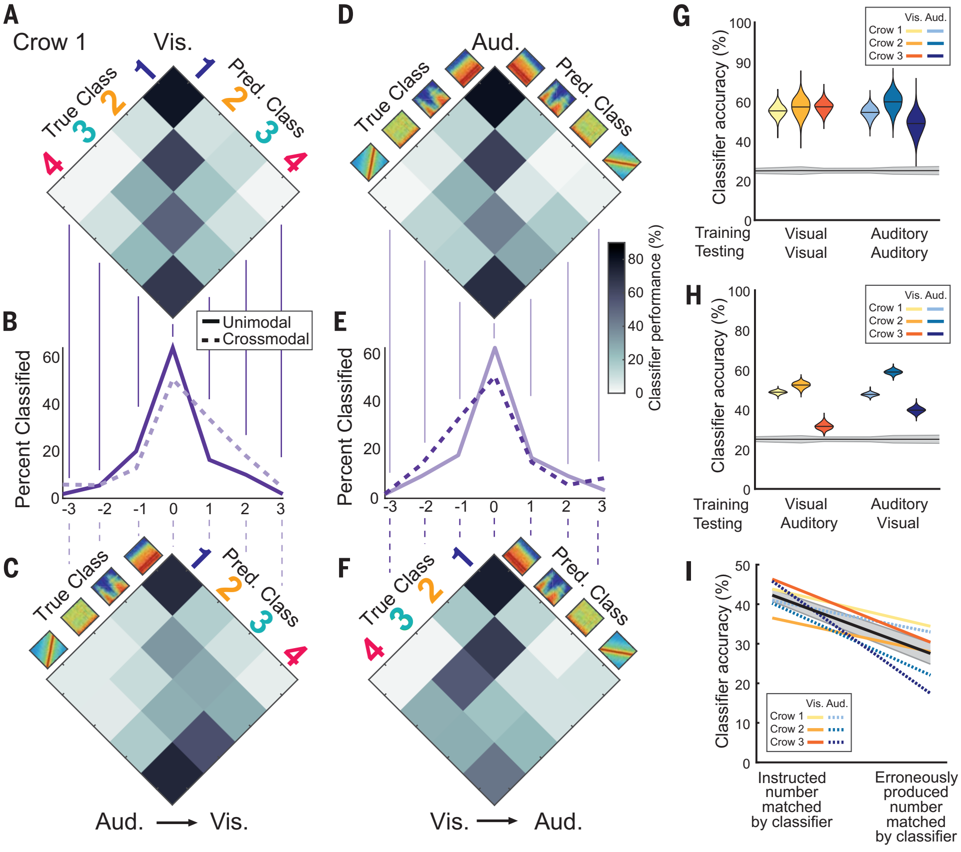

**Fig. 3. Acoustic features of the first vocalization contain information about the number of impending vocalizations.** (**A**) Confusion matrix displaying the accuracy of a support vector machine classifier when predicting the number of 1 to 4 vocalizations of crow 1 performing the vocal task with visual cues. (**B**) The classifier’s performance as a function of absolute numerical distance from the target number. Solid function represents training and testing within visual trials [from matrix in (A); dotted line represents cross-modal transfer, from matrix in (C)]. (**C**) Confusion matrix displaying classifier accuracy when trained with visual cue trials and tested on auditory cue trials. (**D**) Confusion matrix of a classifier for auditory trials. (**E**) Classifier’s performance functions for training and testing with auditory cue trials [solid lines are from the matrix in (D); the dotted line represents cross-modal transfer (from matrix in (F)]. (**F**) Same as (C) except trained with auditory cue trials and tested on visual cue trials. (**G**) Overall accuracy of the unimodal classifiers. (**H**) Overall accuracy of cross-modal classifiers. (**I**) Proportion of first vocalizations in error trials (from 1 to 4) classified as matching the instructed target number or the actual erroneously produced number.

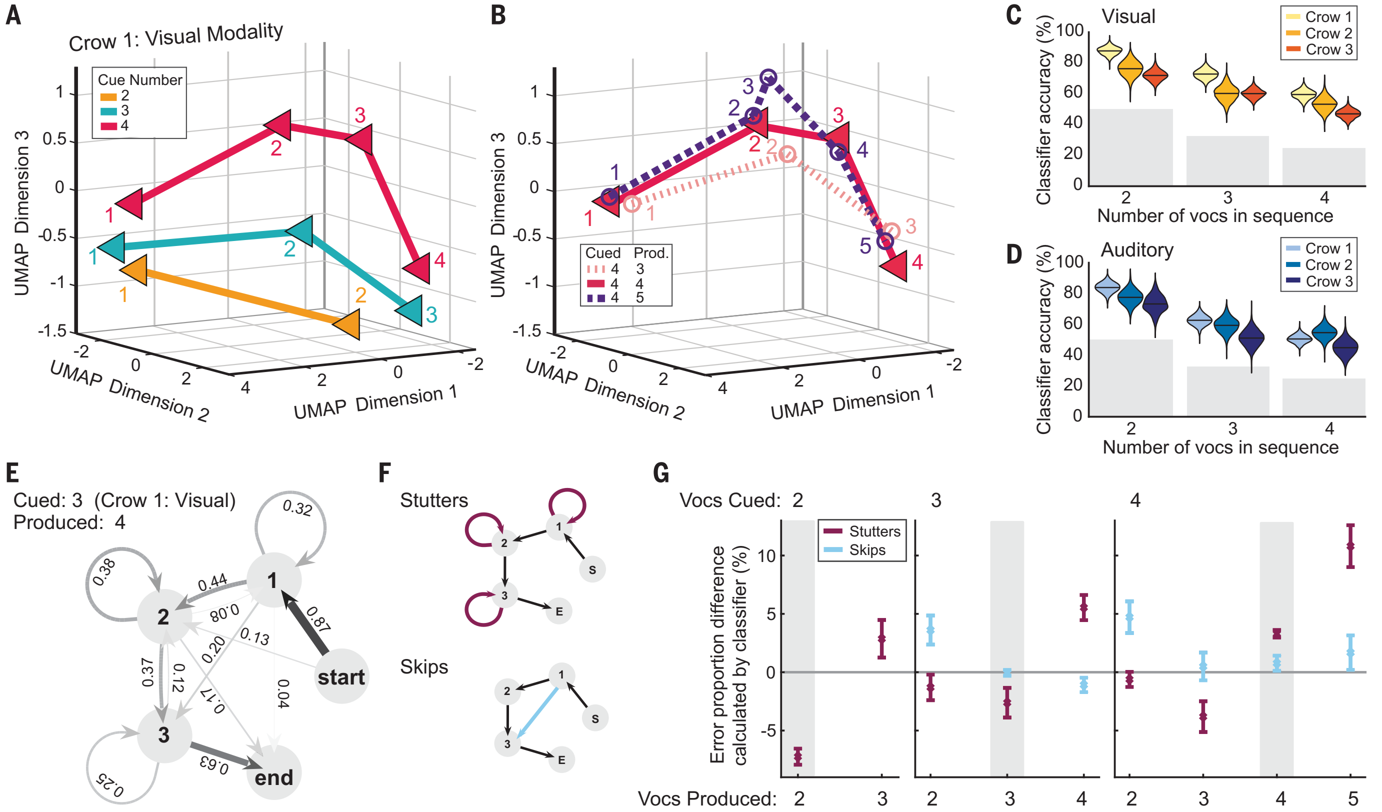

**Fig. 4. Trajectories of correct and erroneous numbers of produced vocalizations.** (**A**) Low-dimensional acoustic space trajectories for sequences of 2, 3, and 4 correctly produced vocalizations. Numbers adjacent to data symbols of respective color-coded paths represent the ordinal position of vocalizations in the sequence. (**B**) Example trajectories of error trials relative to correct trials with cue 4 for the same crow as in (A). (**C**) Overall classification accuracies for ordinal position in the vocal sequence during visual cues. Gray columns indicate chance levels of the classifiers' given number of calls in the sequence. (**D**) Same as (C) except for auditory cues. (**E**) Example transition diagram for decoding of erroneous cue 3 trials where the crow produced four vocalizations. The initiation of the trial is set as the "start" state and the peck to conclude a trial is set as the "end" state. Numbers, line widths, and shading indicate the probability of a transition from each node to another. (**F**) Depiction of two major types of errors. "Stutters" are indicated by outward circles in which a vocalization is repeated whereas "skips" are indicated by internal lines in which a vocalization jumps over the next position in the sequence. (**G**) Error proportion difference (relative to shuffles) of "stutters" and "skips" when 2, 3, and 4 vocalizations were cued. Gray bars mark the correct trials for each cue number; error bars represent the SEM.

but reaction times were also systematically longer the more vocalizations were impending. This was true for both visual and auditory cues and for data combined across all three crows (Spearman correlation: $\text{rho}_{\text{visual}} = 0.6262$, $P = 0.029$; $\text{rho}_{\text{auditory}} = 0.7773$, $P = 0.003$; $n = 12$ per correlation), as well as for data within individual crows (all six correlations for three crows and two cues: $P < 0.002$; Fig. 2C). This increase in reaction times with increasing vocalization numbers suggests that crows plan the entire number of impending vocalizations before motor production.

We further explored whether the acoustic features of the first vocalization were predictive of the total number of vocalizations and characterized individual vocalizations by extracting 23 predefined acoustic features (*18*) (supplementary methods). We used these vocal parameters including the reaction time to train classifiers (support vector machines) (*19*) on different numbers of impending vocalizations for each crow. The classifiers were then tested on a subset of correct trials set aside to explore how well they could predict the number of impending vocalizations. The resulting confusion matrices spanning the true numbers of vocalizations against the classifier-predicted vocalization numbers for an example crow are shown in Fig. 3, A and D, for visual and auditory cues, respectively. The high accuracy values along the diagonal of the matrices show that the classifiers were able to reliably predict the forthcoming number of vocalizations from the first vocalization at about 55% accuracy for each crow (chance level ~25%) (Fig. 3G). The classifiers mainly confused numerically adjacent numbers, resulting in a bell-shaped performance function showing a numerical distance effect (Fig. 3, B and E, and fig. S1) that mirrored the crow's behavioral performance (Fig. 1, C and D). Significant performance of the classifiers was observed even when the reaction time was excluded as a feature (fig. S2), highlighting that the acoustic parameters of the first vocalization alone predict the number of impending vocalizations.

Demonstrating that the crows abstracted numerical values across cue modalities, a classifier trained on visual cues was able to generalize to auditory cues and vice versa (Fig. 3, C and F). This resulted in cross-modal classifier performance curves (dotted lines in Fig. 3B for auditory → visual and 3E for visual → auditory). Significant cross-modal generalization in the range of 30 to 55% accuracy was present with each crow (Fig. 3H). Little generalization was seen between crows, indicating that each crow used distinct vocalizations to solve the task (fig. S3). In sum, the acoustic features and reaction time of the first vocalization were informative of the planned number of vocalizations irrespective of cue modality within each crow. A comparison of the classifier performance based on vocal features collected during test sessions with those collected early during training suggests that the crows adjusted their vocal parameters with increasing proficiency in the task (fig. S4).

Our previous analyses examined only correct trials. Next, we explored whether errors—when the crows produced more or fewer vocalizations than instructed—occurred because the motor plan was incorrect from the start or because the crows started out with a correct plan but "lost track" while vocalizing. We applied classifiers trained on acoustic features of the first vocalization plus reaction time in correct trials to predict the number of vocalizations in error trials. This revealed that the first vocalization in error trials better reflected the instructed number compared with the erroneously produced number (Fig. 3I, binomial test of proportions: $P < 0.0001$). This suggests that crows started out correctly but "lost track" during production.

To explore this hypothesis, we analyzed all vocalizations in the response sequence and projected their acoustic features into a low-dimensional acoustic space, resulting in trajectories of the uttered vocalizations [uniform manifold approximation and projection (UMAP)] (*20*). We found that all vocal sequences, irrespective of the produced number of vocalizations, followed similar trajectories with the start and end points converging in similar regions of the acoustic space (Fig. 4A). This suggested that the acoustic features of ordinal vocalizations change in systematic ways to permit a readout of how far along the vocal sequence a particular vocalization has advanced. In addition, the trajectories of error trials provided an indication as to where the crow went "off track." Figure 4B shows an example in which four vocalizations had been instructed but the crow sometimes erroneously produced three or five calls. Both "more vocalizations" and "fewer vocalizations" errors follow a similar overall trajectory as the correct trials. However, there were distance differences between vocalizations. For example, when the crow erroneously made five vocalizations, the second and third vocalizations sat closely in acoustic space, i.e., the crow appeared to repeat a vocalization with similar acoustic features signifying an erroneous repetition along the enumeration process.

We used the acoustic features of all vocalizations to train classifiers to predict the ordinal position of each vocalization in sequences of 2 to 4 vocalizations (Fig. 4, C and D, and fig. S5). Using the positions predicted by the classifiers we constructed transition diagrams of the vocal sequences (Fig. 4E and fig. S6). In addition to correct progressions through the sequences where vocalizations advance in steps of 1 (i.e., Pos. $1 \rightarrow 2 \rightarrow 3$), different types of errors (start errors, stutters, skips, backward, and end errors) could be separated (fig. S6). This analysis provided a behavioral readout of the ordinal position and relative transition the crow was representing for every single vocalization. We normalized the classifiers' performance (i.e., proportion of different errors) relative to a shuffled distribution (Fig. 4, F and G, and fig. S6) and focused on "skip" errors where intermediate vocalizations were missing (i.e., $1 \rightarrow 3$) and "stutter" errors where the same ordinal position was repeated (i.e., $1 \rightarrow 1$). Across trials (Fig. 4G), the types of errors depended on the number of vocalizations [cue 2: 1-factor analysis of variance (ANOVA), $P < 0.001$; cue 3: 2-factor ANOVA, interaction: $P < 0.0001$; cue 4: 2-factor ANOVA, interaction: $P < 0.0001$]: stutters were more likely in "more" error trials than in "fewer" error or correct trials; by contrast, skips were more likely in "fewer" error trials than in "more" error or correct trials. This confirmed that error trials tended to start out correctly (Fig. 3I) but advanced either with too few or too many vocalizations, as revealed by the acoustic features of vocalizations that signal where the crow went "off track."

Our results demonstrate that crows can flexibly and deliberately produce an instructed number of vocalizations by using the "approximate number system" (*21*), a nonsymbolic number estimation system shared by humans and animals (*15*). This ability is especially impressive given that volitional vocalizations are more difficult to produce and require much longer reaction times (~1 to 2 s) for crows (*13*, *14*) and monkeys (*22*) as compared with instructed pecks or head movements in crows (*12*, *23*) or hand movements in monkeys (*22*, *24*), which show reaction times of only a few hundreds of milliseconds. The crows' cognitive vocal control opens the possibility that some bird species may deliberately use the number of vocalizations to convey ecologically relevant information (*3*, *25*). This competency in crows also mirrors toddlers' enumeration skills before they learn to understand cardinal number words and may therefore constitute an evolutionary precursor of true counting where numbers are part of a combinatorial symbol system (*1*, *2*).

This vocal flexibility in crows is likely enabled by brain nuclei supporting singing and cognition in songbirds (*26*). Neuronal representations of variable numbers of repeat syllables in song have been found in certain song nuclei (*27*) and manipulations of song nuclei can influence the number of repeats produced (*28*). Moreover, neurons selective to the number of objects and self-generated actions (*11*, *12*), and neurons signaling volitional vocal initiation (*14*), are present in the telencephalic nidopallium caudolaterale, a brain area representing executive functions in birds. Notably, better vocal learning in songbirds correlates with enhanced problem solving (*29*). Studying the neuronal underpinnings of cognitively controlled vocal production in crows offers a chance to explore the interaction between numerical and communication systems in the brain.

## REFERENCES AND NOTES

1. S. Carey, *The origin of concepts* (Oxford Univ. Press, 2009).
2. H. Wiese, *Trends Cogn. Sci.* **7**, 385–390 (2003).
3. C. N. Templeton, E. Greene, K. Davis, *Science* **308**, 1934–1937 (2005).
4. C. N. Templeton, E. Greene, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5479–5482 (2007).
5. R. M. Seyfarth, D. L. Cheney, *Annu. Rev. Psychol*. **54**, 145–173 (2003).
6. J. Fischer, T. Price, *Neurosci. Biobehav. Rev.* **82**, 22–31 (2017).
7. G. Szipl, E. Ringler, M. Spreafico, T. Bugnyar, *Front. Zool.* **14**, 57 (2017).
8. D. Scarf, H. Hayne, M. Colombo, *Science* **334**, 1664 (2011).
9. R. Rugani, G. Vallortigara, K. Priftis, L. Regolin, *Science* **347**, 534–536 (2015).
10. A. Smirnova, O. F. Lazareva, Z. A. Zorina, *J. Exp. Anal. Behav.* **73**, 163–176 (2000).
11. H. M. Ditz, A. Nieder, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7827–7832 (2015).
12. M. E. Kirschhock, A. Nieder, *Nat. Commun.* **13**, 6913 (2022).
13. K. F. Brecht, S. R. Hage, N. Gavrilov, A. Nieder, *PLOS Biol.* **17**, e3000375 (2019).
14. K. F. Brecht, S. Westendorff, A. Nieder, *Cell Rep.* **42**, 112113 (2023).
15. A. Nieder, *Trends Ecol. Evol.* **35**, 605–617 (2020).
16. S. Sternberg, S. Monsell, R. L. Knoll, C. E. Wright, *Perception and production of fluent speech*, R. A. Cole, Ed. (Lawrence Erlbaum, 1980), pp. 469–505.
17. D. A. Rosenbaum, R. G. Cohen, S. A. Jax, D. J. Weiss, R. van der Wel, *Hum. Mov. Sci.* **26**, 525–554 (2007).
18. J. E. Elie, F. E. Theunissen, *Anim. Cogn.* **19**, 285–315 (2016).
19. C. Cortes, V. Vapnik, *Mach. Learn.* **20**, 273–297 (1995).
20. T. Sainburg, M. Thielk, T. Q. Gentner, *PLOS Comput. Biol.* **16**, e1008228 (2020).
21. G. Anobile, R. Arrighi, E. Castaldi, D. C. Burr, *Trends Cogn. Sci.* **25**, 24–36 (2021).
22. N. Gavrilov, A. Nieder, *elife* **10**, e62797 (2021).
23. M. Quest, P. Rinnert, L. Hahner, A. Nieder, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205515119 (2022).
24. H. Koda, T. Kunieda, T. Nishimura, *R. Soc. Open Sci.* **5**, 180879 (2018).
25. R. D. Magrath, T. M. Haff, P. M. Fallow, A. N. Radford, *Biol. Rev.* **90**, 560–586 (2015).
26. J. J. Bolhuis, K. Okanoya, C. Scharff, *Nat. Rev. Neurosci.* **11**, 747–759 (2010).
27. H. Fujimoto, T. Hasegawa, D. Watanabe, *J. Neurosci.* **31**, 10023–10033 (2011).
28. Y. S. Zhang, J. D. Wittenbach, D. Z. Jin, A. A. Kozhevnikov, *J. Neurosci.* **37**, 2600–2611 (2017).
29. J. N. Audet, M. Couture, E. D. Jarvis, *Science* **381**, 1170–1175 (2023).
30. D. A. Liao, K. F. Brecht, L. Veit, A. Nieder, Data from: Crows can 'count' the number of self-generated vocalizations, Dryad (2024). doi: 10.5061/dryad.qjq2bvqpz.

## ACKNOWLEDGMENTS

We thank M. Kirschhock and A. Song for useful discussions. **Funding:** This work was supported by grants from the German Research Foundation (DFG NI 618/11-1) to A.N. and (DFG BR 5908/1-1) to K.F.B. and a Research Fellowship from the Alexander von Humboldt Foundation to D.A.L. **Author contributions:** D.A.L. and A.N. designed the experiment. D.A.L and K.F.B. conducted the experiments. D.A.L., L.V., and A.N. analyzed the data. D.A.L. and A.N. wrote the manuscript. D.A.L., K.F.B., L.V., and A.N. edited the manuscript. A.N. supervised the study. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data to reproduce analyses and figures are available at the Dryad Digital Repository (*30*).

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adl0984
Materials and Methods
Figs. S1 to S6
References (*31–34*)

Submitted 28 September 2023; accepted 22 April 2024
10.1126/science.adl0984