Cognition 273 (2026) 106527

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full Length Article

Reaction-time signatures reveal divergent cognitive strategies underlying numerical decisions in monkeys and crows

Tobias Machts [1], Maximilian E. Kirschhock [1,2], Andreas Nieder [*]

Animal Physiology, Institute of Neurobiology, University of Tuebingen, Auf der Morgenstelle 28, 72076 Tuebingen, Germany

ARTICLE INFO

Keywords:
Primate
Corvid
Numerosity
Decision making
Reaction-time dynamics

ABSTRACT

Non-symbolic numerical competence, supported by the approximate number system (ANS), is widespread across animals, yet it remains unclear whether similar numerical performance reflects shared cognitive mechanisms. We compared two rhesus monkeys and two carrion crows on an extended delayed match-to-numerosity task. Animals viewed a sample numerosity (one to four), maintained it across a delay, and searched sequentially through up to three test displays for a match. Behavioral accuracy was comparable across species, with both showing classical ANS signatures such as numerical distance and numerical size effects. However, reaction-time analyses revealed striking species differences: monkeys' responses slowed with increasing numerosity and later test phases, whereas crows responded faster over later test phases and showed numerosity-invariant reaction times. Classifier decoding further confirmed that numerosity and test phase were strongly reflected in monkeys' reaction times but only weakly in crows'. These findings demonstrate that equivalent numerical performance can arise from fundamentally different cognitive strategies, as revealed by distinct reaction-time dynamics, highlighting how evolutionary constraints shape species-specific solutions to the same computational problem.

One-sentence summary: Monkeys solve an abstract-category delayed match-to-sample task through slow, capacity-limited, sequential checking, whereas crows rely on fast, parallel, anticipatory processing with reduced working-memory load.

1. Introduction

Non-symbolic numerical competence is widespread across animals and supports adaptive decision-making (Brannon & Terrace, 1998; Howard, Avarguès-Weber, Garcia, Greentree, & Dyer, 2018; Nieder, 2020; Potrich, Zanon, & Vallortigara, 2022; Rugani, Fontanari, Simoni, Regolin, & Vallortigara, 2009; Scarf, Hayne, & Colombo, 2011; Stancher, Rugani, Regolin, & Vallortigara, 2015). Among vertebrates, both monkeys (primates) and crows (corvids) can discriminate, remember, and act upon numerical quantities (Cantlon & Brannon, 2006; Ditz & Nieder, 2016; Merten & Nieder, 2009). These behaviors are commonly attributed to the approximate number system (ANS), a domain-general mechanism that allows numerical estimation when explicit counting is not possible (Nieder, 2025, 2026). As a signature of the ANS, monkeys and corvids follow similar psychophysical principles — including the numerical distance effect (performance improves as the numerical difference between quantities increases), the numerical size effect (performance declines as overall magnitude increases) (Ditz & Nieder, 2016; Merten & Nieder, 2009), and systematic cognitive judgment biases (Jannasch, Grüb, & Nieder, 2025).

In addition to the ANS, numerical cognition research has proposed a second mechanism, the object tracking system (OTS), which is thought to support precise representation of small numerosities by assigning markers (or pointers) to small numbers of individual elements (typically up to ~3–4 items) (Pylyshyn, 2001). Whereas the ANS is characterized by ratio-dependent performance consistent with Weber's law, OTS processing is typically associated with rapid and accurate individuation of small sets without ratio dependence. Both systems are thought to coexist and may contribute differentially depending on task demands and numerical range (Feigenson, Dehaene, & Spelke, 2004).

So far, it remains unclear how closely the behavioral signatures of numerosity discriminations in monkeys and crows align under identical task conditions; whether the two species share not only the same performance outcomes but also the same cognitive strategies. Primates and

* Corresponding author.
E-mail address: andreas.nieder@uni-tuebingen.de (A. Nieder).
[1] Shared first authors.
[2] Present address: Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel.

https://doi.org/10.1016/j.cognition.2026.106527
Received 19 January 2026; Received in revised form 9 March 2026; Accepted 11 March 2026
Available online 19 March 2026

songbirds differ in telencephalic brain neuroanatomy (layered cortex vs. nuclearly organized pallium, respectively) and sensorimotor systems (eye–hand vs. eye–beak). These differences might indicate that even when performing the same task, the species may rely on distinct cognitive strategies.

To address this question, we trained two rhesus monkeys and two carrion crows on an extended delayed match-to-numerosity task. Animals viewed a briefly presented sample numerosity, maintained it across a delay, and then searched for the matching numerosity in up to three sequential test displays. By varying target numerosity (one to four) and the number of nonmatching items, the task exhibited different demands on numerical discrimination, working memory, and recognition.

These manipulations are known to affect cognitive processes in humans, suggesting that similar cognitive constraints could influence nonhuman animals. For example, human studies show that multiple nonmatch distractors can induce sequential search, slowing reaction times and reducing accuracy (Pomè, Anobile, Cicchini, Scabia, & Burr, 2019; Railo, Koivisto, Revonsuo, & Hannula, 2008; Vetter, Butterworth, & Bahrami, 2008). Likewise, maintaining a visuo-spatial working-memory load while judging multi-item arrays reduces numerosity precision (Castaldi, Piazza, & Eger, 2021; Ester, Ho, Brown, & Serences, 2014). Together, these findings imply that task complexity interacts with memory and attention to shape numerosity judgments. This motivates a cross-species comparison to determine whether monkeys and crows are subject to similar cognitive constraints. Specifically, comparing their respective performances under identical task conditions can reveal whether similar accuracy reflects shared mechanisms or different processing styles, and how neural and memory constraints shape reaction-time patterns.

2. Results

2.1. Behavioral accuracy

Two monkeys and two crows were trained on a modified delayed match-to-numerosity task (Machts, Grüb, & Nieder, 2025; Machts & Nieder, 2026). They viewed sample numerosities from one to four, followed by a delay. Each trial then included up to three sequential test phases presented in equal proportion (see Materials and Methods, Fig. 1A). In each phase, the animals searched for a numerosity that matched the sample. They were rewarded for selecting a matching numerosity in any of the test displays, which encouraged them to inspect each stimulus carefully and continue searching until they found the match.

Both the monkeys and the crows mastered the task (Fig. 2). They reached similar mean performance levels across sessions, all well above

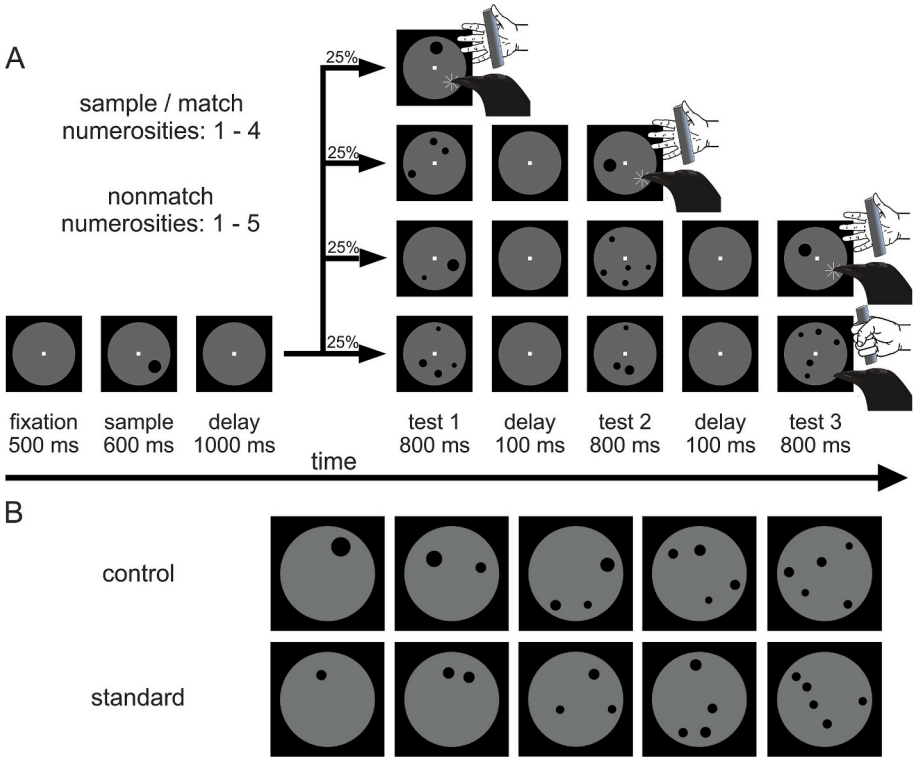

Fig. 1. Task protocol. (A) Both monkeys and crows performed a delayed match-to-sample task involving dot numerosities. The monkeys initiated a trial by grabbing the response bar and fixating on the fixation target, whereas the crows had to position their heads centrally in front of the monitor to close an infrared light barrier. These actions had to be maintained throughout the task, until the monkeys released the bar or the crows pecked at the touchscreen. During a sample phase, a sample stimulus with numerosities ranging from 1 to 4 was presented, followed by a delay phase. The monkeys and crows had to respond whenever a numerosity matching the sample was presented in one of three subsequent test phases. The monkeys had to respond by releasing the bar and the crows had to respond by moving the head out of the IR-light barrier. They had to refrain from responding whenever nonmatching numerosities (ranging from 1 to 5) were presented in the test phases. Up to three sequential test phases (Test 1 – Test 3) were presented pseudo-randomly in different trial conditions; in each test phase, a matching or nonmatching numerosity could appear. If the final Test 3 phase showed a nonmatch, the monkey had to maintain the bar and the crow had to withhold the pecking response until after the display offset. This resulted in a total of four different trial conditions. Sample numerosities and trial conditions were presented in a random and balanced way. (B) The spatial arrangement and diameter of the dots varied depending on the stimulus protocol. The standard stimuli featured randomly positioned and sized dots. In 50% of the trials, the standard stimuli were used, whereby an increasing number of dots resulted in an increasing occupied area. Therefore, control stimuli were presented in the remaining 50% of trials. In these, the dot density and total occupied area were equated across numerosities. Dot density was defined as the average distance between the centers of all dots on a numerosity display. Total occupied area was defined as the cumulative surface area of all the dots on a numerosity display. Thus, as the numerosity increased, the diameter of the dots decreased in the control stimuli. Both standard and control stimuli were presented pseudo-randomly and with equal frequency.

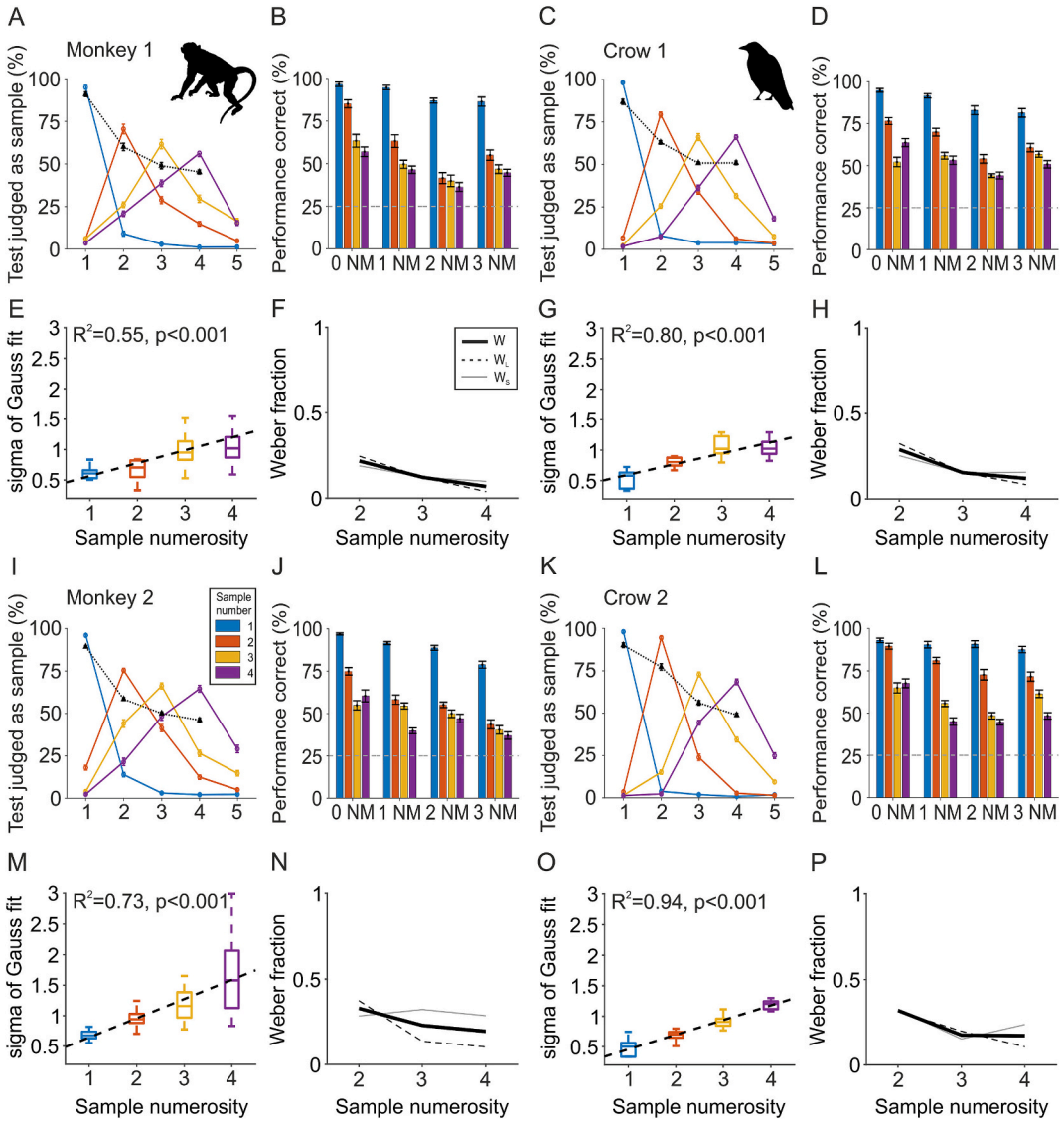

Fig. 2. Behavioral performance. (A, C, I, K) The behavioral performance functions of two monkeys and two crows are shown for sample numerosities of 1–4 (Monkey 1: 27 sessions; Monkey 2: 33 sessions; Crow 1: 20 sessions; Crow 2: 20 sessions). These functions reflect the probability that a monkey or a crow will judge a display in the test period as containing the same number of items as the sample numerosity (indicated by the solid line in four colors). The peak data points of each colored function indicate correct sample matching. Data points in the flanks reflect incorrect responses to non-matching numerosities. The black dotted line indicates the average performance for each sample numerosity. The error bars indicate the standard error of the mean (SEM) across sessions. (B, D, J, L) Average correct performance of the two monkeys and two crows to the four sample numerosities and the four test conditions (NM = nonmatch). Grey dashed line indicates chance level at 25%. Error bars indicate the SEM across sessions. (E, G, M, O) Distributions of widths (σ of fitted Gauss curves) for each sample numerosity 1 to 4 on linear numerical scaling are plotted as box plots (horizontal line indicate the median values). Dashed black line show the linear fit of Gauss width as a function of sample numerosity. (F, H, N, P).Mean Weber fractions (W, solid black line) for smaller (W_S, solid grey line) and larger ($_WL$, black dashed line) sample numerosity discrimination. Animal silhouettes available from phylopic.org under creative commons license.

the 25% chance level (Monkey 1: 27 sessions, 61.32 ± 1.36%, $p < 0.001$; Monkey 2: 33 sessions, 60.95 ± 0.58%, $p < 0.001$; Crow 1: 20 sessions, 59.90 ± 0.64%, $p < 0.001$; Crow 2: 20 sessions, 64.30 ± 1.14%, $p < 0.001$; Binomial test). Furthermore, to assess the animals' ability to identify the correct matching numerosity while ignoring nonmatching numerosities, we computed behavioral performance functions for each session. These functions quantify the probability that monkeys and crows judged a test numerosity as equal to the sample. Across all sample numerosities, both species selected the matching numerosity more often than any nonmatching numerosity (Fig. 2A, C, I, K).

Consistently, performance was highest for numerosity 1 (Monkey 1: 95.02 ± 1.12%; Monkey 2: 96.01 ± 0.52%; Crow 1: 98.12 ± 0.47%; Crow 2: 98.02 ± 0.48%). Accuracy declined as numerosity increased. For numerosity 4, peak performance dropped to 56.31 ± 1.96% (Monkey 1), 64.57 ± 1.99% (Monkey 2), 66.08 ± 1.37% (Crow 1), and 68.45 ± 1.63% (Crow 2). As predicted by the numerical distance effect — which posits that neighboring numerosities are the hardest to discriminate (Ditz & Nieder, 2016; Kirschhock, Ditz, & Nieder, 2021; Merten & Nieder, 2009; Nieder & Miller, 2003) — both monkeys and crows made the most errors when nonmatching numerosities were adjacent to the target. Their accuracy improved noticeably as the nonmatch numerosities became more distant from the target. Additionally, performance functions systematically widened as target numerosities increased, reflecting the numerical size effect, which indicates that discriminating larger numerosities is more difficult at a given numerical distance (Nieder, 2020). These patterns show that both species exhibit the hallmarks of the approximate number system (ANS) — a non-symbolic estimation system also observed in humans when counting is

prevented and they must rely on estimating numerosity (Merten & Nieder, 2009; Nieder, 2026).

Next, we calculated the average performance for each trial type—no nonmatch, one nonmatch, two nonmatches, and three nonmatches across all four sample numerosities (Fig. 2B, D, J, L). As expected, the number of nonmatch numerosities affected both species' ability to recall and identify the matching numerosity (three-way ANOVA with factors 'number of nonmatches', 'sample numerosity' and 'stimulus protocol'; Monkey 1: $F(3,841) = 79.74$, $p < 0.001$; Monkey 2: $F(3,1033) = 73.45$, $p < 0.001$; Crow 1: $F(3,617) = 70.63$, $p < 0.001$; Crow 2: $F(3,617) = 61.78$, $p < 0.001$). Compared to trials with no nonmatch numerosities, both monkeys' and crows' overall performance declined progressively as the number of nonmatches increased. This drop in performance likely reflects the growing cognitive demands on recognition memory (Machts et al., 2025). Monkey 1, Crow 1, and Crow 2—but not Monkey 2—showed a slight, idiosyncratic performance increase for numerosities following three nonmatches.

Furthermore, behavioral performance in both species depended on the sample numerosity within each trial type, shown by a significant interaction between trial type and sample numerosity (Monkey 1: $F(9,841) = 6.85$, $p < 0.001$; Monkey 2: $F(9,1033) = 4.71$, $p < 0.001$; Crow 1: $F(9,617) = 8.00$, $p < 0.001$; Crow 2: $F(9,617) = 10.63$, $p < 0.001$). Reflecting the numerical size effect, trials with numerosity 1 were the easiest for both monkeys and crows, while trials with numerosity 4 were the most challenging for the monkeys (Fig. 2B, D, J, L). Despite this, average performance across all trial types and sample numerosities was significantly above chance ($p < 0.05$, Binomial test), showing that both species were able to identify the correct numerosity most of the time.

Additionally, we used the three-way ANOVA to evaluate whether non-numerical visual cues influenced performance. The control protocol equated dot density and total occupied area across numerosities to minimize potential reliance on continuous visual variables. In crows, the analysis revealed no significant main effect of stimulus protocol and no significant interactions involving stimulus protocol (all $p > 0.05$), indicating that performance was indistinguishable between standard and control stimuli.

In monkeys, we observed statistically significant main effect of stimulus protocol (Monkey 1: $F(1,1033) = 35.51$, $p = 0.0353$; Monkey 2: $F(1,841) = 4,44$, $p < 0.001$). However, the magnitude of this effect was minute: performance differences between protocol types were only approximately 3–4% (Monkey 1: standard 61.3%, control 58.2%; Monkey 2: standard 63.3%, control 59.2%). Consistent with this, the corresponding effect sizes were small (Monkey 1: partial $\eta^2 < 0.01$; Monkey 2: partial $\eta^2 \approx 0.03$), indicating that the influence of stimulus protocol accounted for less than 1% in Monkey 1 and around 3% in Monkey 2 of the variance in behavioral performance.

In addition, weak interactions involving stimulus protocol were observed in individual animals. In one monkey, stimulus protocol interacted with the number of nonmatches (Monkey 1: $F(3,1033) = 5.21$, $p < 0.01$, partial $\eta^2 \approx 0.01$), whereas in the other monkey an interaction between stimulus protocol and sample numerosity was detected (Monkey 2: $F(3,841) = 7.18$, $p < 0.001$, partial $\eta^2 \approx 0.02$). However, these effects were small in magnitude, with effect sizes indicating that the interactions accounted for only a minor proportion of the variance. Overall, these results suggest that continuous visual variables such as dot density or total occupied area had only a limited influence on behavioral performance and cannot account for the main numerical effects observed in the task.

To formally quantify the numerical scaling properties of the behavioral performance functions, we fitted Gaussian functions to the tuning curves for each sample numerosity and estimated the width parameter (σ) of the fits. The width of the tuning curves increased systematically with increasing numerosity, indicating broader discrimination functions for larger numerosities (Fig. 2E,G,M,O). To test this relationship statistically, we correlated σ with sample numerosity. For all animals, σ increased significantly with numerosity (Pearson correlation: Monkey 1: $r = 0.55$, $p < 0.001$; Monkey 2: $r = 0.73$, $p < 0.001$; Crow 1: $r = 0.80$, $p < 0.001$; Crow 2: $r = 0.94$, $p < 0.001$).

In addition, we calculated Weber fractions derived from the behavioral tuning functions (Fig. 2F, H, N, P). These remained roughly constant across numerosities (~0.25), consistent with Weber-like scaling. Together, these results demonstrate that behavioral performance follows classical signatures of the approximate number system (ANS).

2.2. Reaction times

Although both species showed similar behavioral performance, reaction times (RTs) differed clearly between monkeys and crows. RT was measured for each matching numerosity and defined as the time from display onset to the animal's response — bar release for monkeys or touchscreen peck for crows (display duration: 800 ms). To account for visual processing and motor preparation, only responses of at least 100 ms after display offset were allowed (see Materials and Methods).

Both monkeys showed a positive relationship between RT and test phase for all four matching numerosities: they responded more slowly as the test phase increased (Fig. 3A, C). This effect was strongest for numerosities 2–4 and weaker or absent for numerosity 1 (Pearson

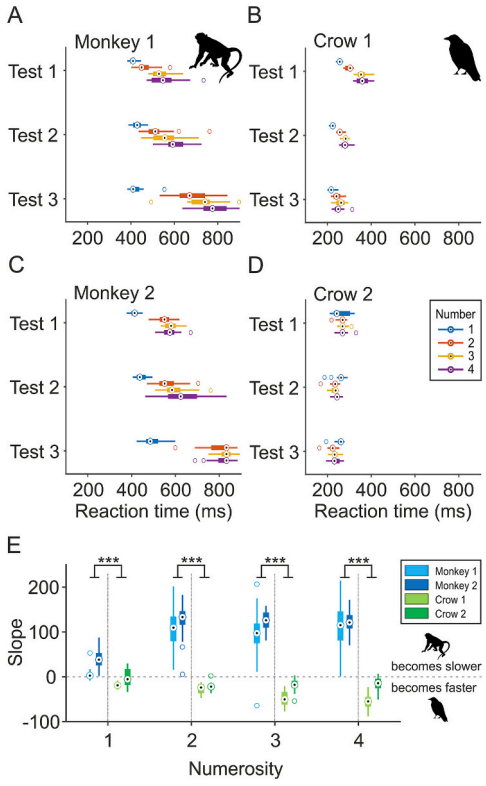

Fig. 3. Reaction times (RT) for correct responses to the matching test stimulus, categorized by the three test phases. (A, B, C, D) The colored boxplots indicate the RT for the respective matching numerosities (1–4) that the monkeys and crows should match to. The edges of the boxes represent the 25th and 75th percentiles, and the black dot inside the boxes demark the respective medians. Outliers are indicated by colored circles. Note that the monkeys become slower as the number of tests increases and the crows become faster (Monkey 1: 27 sessions; Monkey 2: 33 sessions; Crow 1: 20 sessions; Crow 2: 20 sessions). (E) The RTs for the three test phases were fitted using a linear function (linear regression). The resulting slopes were plotted as boxplots for the respective matching numerosities (1–4) and compared between monkeys and crows using a Wilcoxon rank sum test. A positive slope indicates that the animals are getting slower with increasing test phase, while a negative slope indicates that they are getting faster. See (A, B, C, D) for boxplot description. ***: $p < 0.001$. Animal silhouettes available from phylopic.org under creative commons license.

correlation; Monkey 1: numerosity 1: $r = 0.17$, $p = 0.136$; numerosity 2: $r = 0.79$, $p < 0.001$; numerosity 3: $r = 0.73$, $p < 0.001$; numerosity 4: r = 0.79, $p < 0.001$; Monkey 2: numerosity 1: r = 0.73, p < 0.001; numerosity 2: $r = 0.81$, p < 0.001; numerosity 3: $r = 0.84$, p < 0.001; numerosity 4: $r = 0.82$, p < 0.001).

Interestingly, the pattern was reversed in crows (Fig. 3B, D). Both crows showed a negative relationship between RT and test phase: they responded faster as the test phase increased. Pearson correlations were as follows: Crow 1: numerosity 1: $r = -0.78$, p < 0.001; numerosity 2: $r = -0.80$, p < 0.001; numerosity 3: $r = -0.86$, p < 0.001; numerosity 4: r = −0.86, p < 0.001; Crow 2: numerosity 1: $r = -0.07$, $p = 0.608$; numerosity 2: $r = -0.68$, p < 0.001; numerosity 3: $r = -0.63$, p < 0.001; numerosity 4: $r = -0.51$, $p < 0.001$.

This pattern was confirmed by fitting a linear function to the RTs for each matching numerosity across the three test phases and calculating the slopes (Fig. 3E). Within each species, slope steepness was qualitatively similar across numerosities. Monkeys showed positive slopes, indicating slower responses with later test phases, whereas crows showed negative slopes, indicating faster responses. The slopes differed highly and significantly between monkeys and crows (numerosity 1: Z = −6.564, p < 0.001; numerosity 2: Z = −8.411, p < 0.001; numerosity 3: Z = −8.154, p < 0.001; numerosity 4: Z = −8.396, p < 0.001; Wilcoxon rank sum test, two-sided). Notably, slope steepness was much weaker for numerosity 1 compared to the other numerosities in both species — especially in monkeys (Monkey 1: 3.17; Monkey 2: 38.69; Crow 1: −19.00; Crow 2: −5.50) — indicating that responses to numerosity 1 were less influenced by increasing test phases.

We also observed differences in how the animals responded to different numerosities within a test phase. Monkeys were fastest for numerosity 1 and progressively slower for higher numerosities (Fig. 3A, C). Monkey 1's median RTs for numerosity 1 were 409.63 ms, 428.70 ms, and 409.83 ms across the three test phases; Monkey 2's were 413.88 ms, 437.51 ms, and 485.32 ms. In contrast, both monkeys responded slowest to numerosity 4. Monkey 1's RTs were 547.23 ms, 592.78 ms, and 775.68 ms; Monkey 2's RTs were 574.59 ms, 624.23 ms, and 833.19 ms.

Notably, this gradation was absent in both crows, which responded to all four numerosities with similar RTs (Fig. 3B, D). For Crow 1, median RTs for numerosity 1 were 256.75 ms, 224.75 ms, and 217.25 ms across the three test phases, and for numerosity 4 they were 242.75 ms, 262.25 ms, and 261.75 ms. For Crow 2, median RTs for numerosity 1 were 359.50 ms, 280.75 ms, and 249.50 ms, and for numerosity 4 they were 269.00 ms, 244.00 ms, and 232.00 ms.

2.3. SVM classifier predicts numerosity and test phase based on the reaction times

Collectively, our analyses indicate that, unlike carrion crows, rhesus monkeys respond distinctively to both numerosity and test phase in our task. To confirm this, we used an SVM classifier to predict numerosity and test phase based on each animal's reaction times, training the classifier separately for each subject with 7-fold cross-validation and 150 resampling iterations. Because decoding accuracy depends on sample size, we created pseudo-populations to balance the number of sessions between monkeys and crows. The SVM classifier performed well above the 8.33% chance level for both monkeys and crows (Fig. 4A). However, monkeys achieved significantly higher decoding accuracy than crows (Z = −19.0613, $p < 0.001$, Wilcoxon rank sum test, two-sided). Decoding success was similar between the two monkeys, with no significant difference (Monkey 1: 43.94%, Monkey 2: 44.84%; Z = −0.5394, $p = 0.590$). In crows, decoding success was higher in Crow 1 (32.68%) than in Crow 2 (27.10%; Z = 8.3495, p < 0.001, Wilcoxon rank sum test, two-sided).

In addition to overall decoding accuracy, we generated confusion matrices for each animal from the classifier's predictions. In monkeys, the highest accuracy clearly aligned along the main diagonal (Fig. 4B, D), indicating reliable classification. This pattern was absent in crows,

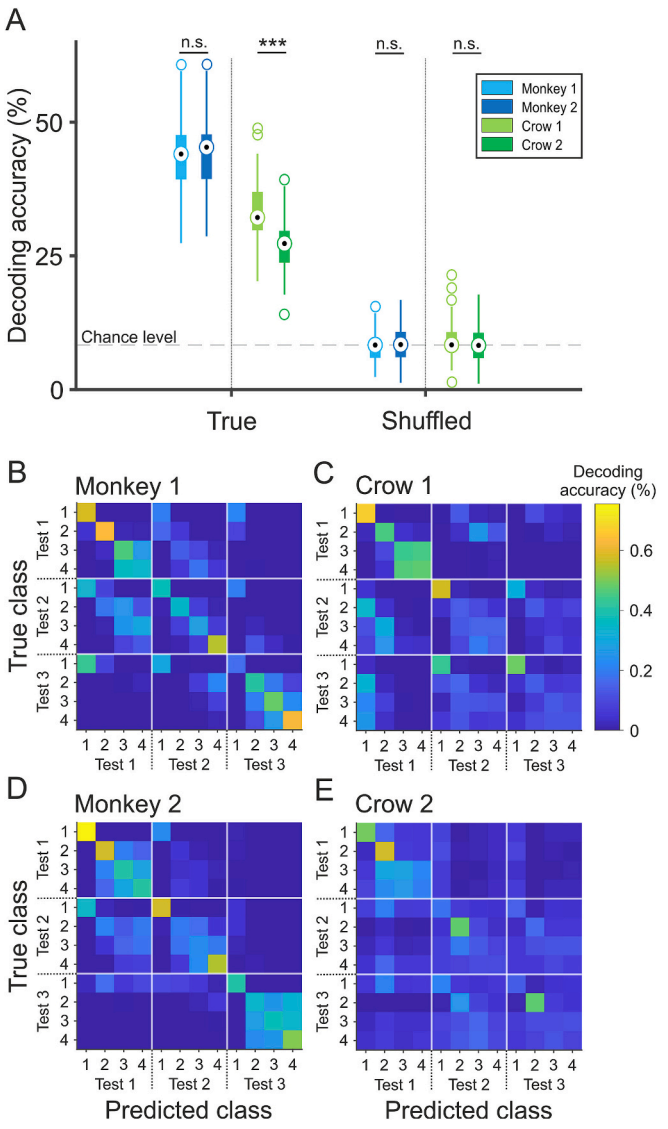

Fig. 4. Decoding numerosity and test phase from RT with a SVM classifier. (A) Mean decoding accuracy of the classifier for true and shuffled labels. The edges of the boxes represent the 25th and 75th percentiles, and the black dot inside the boxes the respective medians. Outliers are indicated by colored circles. Grey dashed line indicates chance level at 8.33%. Decoding accuracy was tested between the animals using a Wilcoxon rank sum test. ***: p < 0.001; n.s.: not significant. (B-E) Confusion matrices for sessions with at least 7 RTs for each correct match numerosity and test phase (Monkey 1: 18 sessions; Monkey 2: 29 sessions; Crow 1: 20 sessions; Crow 2: 20 sessions). Rows within the confusion matrices correspond to classification performance curves per numerosity and test phase, which was evaluated by a seven-fold cross-validation within the sessions. The sessions count was balanced across the animals by creating pseudo populations. The scaling of color maps is the same across the four confusion matrices. Color values in the confusion matrices show the average over cross-validation runs and 150 resamples. Note that for both monkeys, the classifier can predict the correct match numerosity of the respective test phase based on the RT, whereas this is not possible for neither crows.

where decoding of individual classes was indistinguishable (Fig. 4C, E). These results show that an ideal observer could predict numerosity and test phase from RTs far more accurately in rhesus monkeys than in carrion crows.

2.4. Accuracy in relation to reaction times

Finally, we examined the relationship between accuracy and

reaction time using Pearson correlations for each numerosity and test phase. Both monkeys tended to respond faster when performing well, although this pattern was not consistent across all numerosities and test phases and was more pronounced in Monkey 1 (Fig. 5A, C). No significant cases were found where monkeys were slower when performing well. In contrast, crows tended to be slower when performing well, with this effect being more noticeable in Crow 2 (Fig. 5B, D). As with monkeys, this opposite effect — faster responses with better performance — was not statistically significant in crows. Importantly, faster RTs in crows were not associated with reduced accuracy; if anything, higher performance tended to co-occur with shorter reaction times, arguing against a simple speed–accuracy tradeoff.

3. Discussion

We found that monkeys' reaction times (RTs) increased with later test phases and larger numerosities, whereas crows' RTs decreased with test phase and remained largely invariant across numerosities. SVM decoding revealed strong structure in monkeys' RTs but only weak structure in crows', while overall accuracy was comparable across species. These results indicate that, despite similar performance levels, monkeys and crows solve the same numerical task using fundamentally different cognitive strategies. Both species show signatures of the approximate number system (ANS), yet their opposing temporal profiles suggest that monkeys' behavior is best explained by slow, memory-dependent, sequential evaluation, whereas crows' behavior is consistent with fast, parallel processing strategies.

Because the present study included numerosities up to five, it is conceivable that object tracking system (OTS)-related mechanisms contributed to performance. However, the behavioral signatures observed here even for small numerosities—most notably the ratio-dependent discrimination performance consistent with Weber's law—align more closely with predictions of ANS-based processing. This is consistent with previous studies reporting ratio-dependent discrimination across small and large numerosities in these trained species (Cantlon & Brannon, 2006; Ditz & Nieder, 2015, 2016; Jannasch et al., 2025; Jordan & Brannon, 2006; Liao, Brecht, Veit, & Nieder, 2024; Merten & Nieder, 2009; Nieder & Miller, 2003). Given this shared Weber-like scaling across numerosities, we would therefore expect the species-specific reaction time dynamics observed here to persist even for larger sample numerosities.

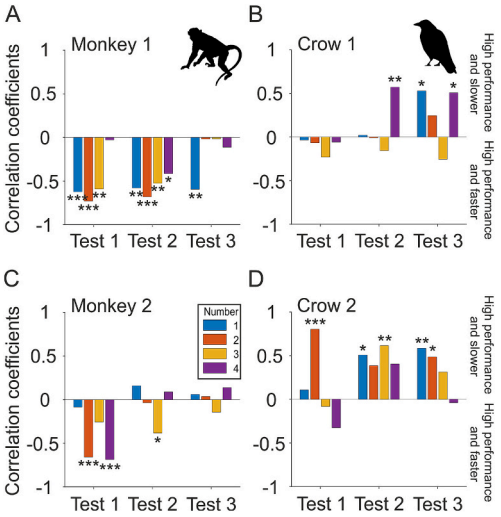

Fig. 5. Pearson correlation coefficients showing the relationship between performance and RT. (A, B, C, D) A Pearson correlation was used to investigate the relationship between performance and RT of monkeys and crows. The resulting correlation coefficients (Pearson's r) were plotted for each correct match numerosity and test phase. Both monkeys tend to have a negative correlation, i.e. they have a better performance when they are faster. In contrast, crows tend to have a positive correlation, i.e. they have a better performance, when they are slower. Stars above or below the bars indicate statistical significance for the respective correlation coefficient. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. Animal silhouettes available from phylopic.org under creative commons license.

3.1. Monkeys use a slow, sequential, memory-limited search strategy

Monkeys responded more slowly when the target appeared later in the sequence and when numerosity increased. The SVM classifier could reliably decode both numerosity and test phase from monkeys' RTs, confirming that their behavior carries strong signatures of serial, memory-dependent evaluation that resembles controlled processing in humans. Their performance slows when more items must be processed or when the target appears later in a sequence, indicating that working-memory load increases over time (Yakovlev, Bernacchia, Orlov, Hochstein, & Amit, 2005). Monkeys also rely on serial, position-based encoding, as shown by ordinal-position errors in sequence recall tasks (Orlov, Yakovlev, Amit, Hochstein, & Zohary, 2002). Even in numerosity judgments — often thought to be fast and parallel — reaction times rise with numerical magnitude, suggesting that decisions become slower as quantity increases (Nieder & Miller, 2004). Monkeys also fail to extract relational or structural patterns from sequences and instead depend on item-specific working-memory representations, which require more effort and are less efficient than human strategies (Zhang et al., 2022). Neurophysiological studies further show that monkeys hold a target numerosity in working memory and check each test display in sequence, using sustained prefrontal and parietal activity to guide recognition (Machts et al., 2025). Together, these findings indicate that monkeys rely on slow, effortful, and capacity-limited processing when they must maintain information and evaluate items one by one. This aligns with evidence that primate numerical cognition depends on fronto-parietal working memory in sequential numerical tasks (Nieder, 2012; Nieder, Diester, & Tudusciuc, 2006).

We acknowledge that the motor responses required from monkeys and crows were not identical: monkeys released a bar, whereas crows moved their heads out of a light barrier. However, in both cases the task required an active, self-initiated movement to report the decision. We therefore consider it unlikely that these relatively small differences in response modality account for the relative differences in reaction times observed between the two species.

3.2. Crows appear to use a parallel, rapid, or anticipatory strategy

Crows, in contrast, responded faster with later test phases and showed near-constant RTs across numerosities, indicating that neither working-memory load nor numerical magnitude imposed measurable costs. Importantly, the speeding observed in crows across later test phases cannot be explained by a simple speed–accuracy tradeoff, as faster responses were not associated with reduced performance; if anything, higher accuracy tended to co-occur with shorter reaction times. This stands in sharp contrast to the monkeys' behavior. Crows also showed shallower SVM decoding accuracy: numerosity and test phase could still be decoded above chance, but with markedly weaker structure, consistent with less systematic variation in RTs.

This pattern matches a body of evidence suggesting that avian numerical processing is rapid, parallel, and feedforward-like. Even numerically naïve crows possess neurons tuned to visual numerosity (Wagener, Loconsole, Ditz, & Nieder, 2018), and young chicks show similar innate number-selective coding (Kobylkov, Mayer, Zanon, & Vallortigara, 2022). Crows also rely on approximate magnitude representations rather than serial counting across sequential and simultaneous inputs (Ditz & Nieder, 2020). Their numerosity-to-action behavior, performed with very short response latencies (≈389–444 ms), likewise suggests rapid stimulus–response mapping without heavy

memory involvement (Kirschhock & Nieder, 2022). These findings support the interpretation that crows rely on fast, low-cost perceptual strategies, with reduced dependence on sequential checking or sustained working memory (Bobrowicz & Osvath, 2019).

3.3. Different cognitive strategies despite similar accuracy

The opposing temporal signatures — load-dependent slowing in monkeys versus anticipatory speeding in crows — demonstrate that similar behavioral performance does not imply shared cognitive mechanisms. Monkeys appear to rely on capacity-limited, controlled processes, whereas crows employ fast, parallel, and heuristic strategies. This divergence reflects a broader comparative pattern: evolution can produce convergent cognitive abilities while relying on distinct neural architectures and computational solutions (Nieder, 2021; Olkowicz et al., 2016). Future work should examine whether these species differences persist under controlled temporal conditions, in the presence of distractors, or with increased working-memory demands.

4. Methods

4.1. Experimental animals

We trained two male adult rhesus monkeys (*Macaca mulatta*, named Monkey 1 and Monkey 2) and two adult carrion crows (*Corvus corone*, named Crow 1 and Crow 2) at the institute's facility and acquired data from them. The monkeys were housed socially under a 12/12-h light–dark cycle. They were provided with environmental enrichment for climbing and foraging, and had access to primate food at all times, with fresh fruit and vegetables provided at regular intervals. During the experiments, the monkeys were kept on a controlled water protocol and earned water as a reward during training.

The crows were housed in small social groups in large indoor aviaries. They were kept on a controlled feeding protocol and earned food as a reward during training. If necessary, water/food was supplemented after the sessions. All experimental procedures had been ethically reviewed and approved and were regulated and supervised by the national authorities, the Regierungspräsidium Tübingen, Germany.

4.2. Experimental apparatus and stimuli

Training sessions were conducted for both animal species in a darkened operant conditioning chamber. The monkeys were seated in primate chairs 57 cm from a 15-in. flat-screen monitor with a resolution of 1024 × 768 pixels and a refresh rate of 60 Hz. An infrared eye-tracking system (ISCAN, Cambridge, MA) was used to control eye fixation while presenting test stimuli. An automated water dispenser was used for reward delivery.

The crows were loosely strapped to a wooden perch with leather jesses for training and positioned in front of a 15-in. touchscreen monitor (3 M Microtouch; 60 Hz refresh rate). An infra-red (IR) light barrier ensured that the crows kept their heads still and maintained a central head position in front of the monitor at a viewing distance of 14 cm throughout the trial. This light barrier consisted of an infrared light emitter/detector fixed to the ceiling of the chamber and a reflector foil attached to the crow's head. The reward (Birdseed pellets and mealworms) was delivered by an automated feeder that was positioned below the monitor. Loudspeakers (Visation WB10) were installed in the chamber to provide auditory feedback, and an infrared camera (Genius iSlim 321R) was installed to enable observational control. The NIMH Cortex software (Bethesda, MD) was used to present the stimuli and acquire behavioral data in both set-ups.

We used random dot arrays as stimuli representing numerosities ranging from one to five dots. The diameter of these dots ranged from 0.5° to 0.9° for the monkeys and from 1.2° to 5.5° for the crows. Each unique dot pattern was presented on a grey background circle subtending 5.4° (for the monkeys) and 26.1° (for the crows) of visual angle. This pattern was used as both the sample stimulus and the test stimulus. To prevent the monkeys and crows from memorizing stimulus sequences, a MATLAB (The MathWorks, Natick, MA) program generated the stimuli anew before each training session. The sample and test stimuli of the same countable numerosity were never identical within a trial or session. To control for low-level, non-numerical visual features, dot diameters were randomly selected in half of the trials (standard protocol, Fig. 1B). In the other half of the trials, the dot density and the total area occupied were equated across stimuli (control protocol, Fig. 1B). Trials of the standard and control protocols were presented in a pseudo-random and unpredictable order.

4.3. Behavioral protocol

We trained two macaque monkeys and two carrion crows on a modified, extended version of the delayed match-to-numerosity task, in which the animals had to match the numerosity of the test stimulus to that of the sample stimulus (Fig. 1A). Nonmatching numerosities presented between the sample and the matching test stimulus had to be ignored. Up to three sequential numerical nonmatches were presented, i.e., the task consisted of up to three test phases, in each of which the correct or incorrect numerosity could be presented.

While the monkeys initiated a trial by grabbing a bar and maintaining eye fixation within a visual angle of 3.5° of a central white dot, the crows had to position their heads centrally in front of the monitor to close an infrared light barrier. Trials were immediately aborted and excluded from further analysis if the monkeys broke fixation or the crows moved out of the predefined position. Initially, a grey circular background was displayed for 500 ms (the pure fixation phase). Subsequently, the sample stimulus containing one to four dots appeared for 600 ms. This sample stimulus had to be memorized over a delay phase of 1000 ms (again with a grey circular background) and then compared in a subsequent sequence of test numerosities. Each test stimulus was presented for 800 ms, with a 100 ms delay between each stimulus.

If the numerosity of the test stimulus was the same as that of the sample stimulus (match situation: 1–4), the monkeys had to respond by releasing the bar and the crows had to respond by moving the head out of the IR-light barrier. If the numerosities were different (nonmatch situation: 1–5), the monkeys continued holding the bar and the crows maintained the predefined position.

A nonmatch numerosity was never shown twice per trial. Sample numerosities and trial types (i.e. whether no, one, two or three nonmatch numerosities were presented) were balanced across conditions and displayed with equal probability (25% each, pseudo-randomly intermixed). Thus, the animals could not predict whether or when the match stimulus would appear. A correct decision per trial was rewarded with a drop of water for the monkeys and a small piece of food for the crows. In 25% of trials involving three nonmatch numerosities in sequence, the monkeys were rewarded for holding the bar and the crows for maintaining the predefined position throughout the trial until the display offset. Since the number of trials that a monkey or crow can perform during an experimental session is limited (on average, Monkey 1 completed 483 trials, Monkey 2 completed 534 trials, Crow 1 completed 640 trials and Crow 2 completed 631 trials), not all possible combinations of sample (1–4) and test numerosities (1–5) could be presented. Therefore, a predefined collection of 64 unique sample-test combinations was presented instead.

Behavioral chance level calculation was as follows: For each test phase, the animals have only two options: to respond or to no respond, corresponding to a probability of $p = 0.5$. Cascading this probability for the subsequent test phases (by multiplication), the following probabilities arise: Test 1 phase: $p = 0.5$; Test 2 phase: $p = 0.25$; Test 3 phase: $p = 0.125$. Since the animals do not know how many test phases to expect (i.e., trial types are unpredictable), the average probability results in $p = 0.25$.

4.4. Data analysis

All analyses were conducted in MATLAB (version R2021a, MathWorks Inc.). All values in the main text and figures refer to the mean ± SEM. SEM was calculated as the standard deviation divided by the square root of the number of samples.

4.5. Behavioral performance

Behavioral performance (% correct) was calculated by counting the number of correctly performed trials in each session and dividing this by the total number of trials. Percent correct performance for each stimulus combination was calculated for each session to construct behavioral performance functions in response to each sample numerosity. Mean response functions were computed as the average of all individual functions per session.

To assess the effects of task demands on behavioral accuracy, we conducted three-way repeated-measures analyses of variance (ANOVA; Interaction Model) separately for each animal. The dependent variable was percent correct performance per session. The within-session factors were 'number of nonmatches' (0–3), 'sample numerosity' (1–4) and 'stimulus protocol' (standard, control). Furthermore, we computed their interaction term. Because we test distributions of means (% accuracy), we can assume normal distributions suitably tested with parametric tests. For all ANOVAs, we report F-values, degrees of freedom, and associated p-values. Effect sizes (partial η^2) were calculated to quantify the magnitude of observed effects.

To quantify the performance functions, we fitted Gaussian bell curves to the performance functions on a linear numerical scale. The center of each fitted Gaussian curve was fixed to the respective target number and the width, height and offset parameters were fitted to minimize the overall residual error.

We calculated Weber fractions to assess the just noticeable difference between two stimuli (Weber, 1851). The following formulas were used to calculate the smaller and larger slopes of the performance functions, respectively:

$W_S = \frac{n - n_S}{n_S}$ for smaller Weber fractions, and

$W_L = \frac{n_L - n}{n}$ for larger Weber fractions.

With n being the target number, n_S is a number smaller than the target number, which the monkeys and crows could discriminate at a rate of 50%. Likewise, n_L is a number greater than the target number and is discriminated in 50% of cases.

4.6. Reaction times (RTs)

RTs were measured for each match numerosity (1–4 in Test 1 – Test 3) and was defined as the time taken by the monkey or crow to react (by releasing the bar in monkeys or by moving the head out of the IR-light barrier in crows) to the matching numerosity after display onset (the display duration was 800 ms). Due to the visual response latency and the time required for motor preparation, the animals were permitted to respond to the stimulus for up to 100 ms after display offset. Consequently, using this approach, responses occurring 100 ms after onset were also counted as errors and excluded from further analysis. Thus, the crows and monkeys could react within a time window of 800 ms, 100 ms after stimulus onset.

We performed Pearson correlation to investigate the relationship between RT and increasing test phase for each match numerosity. Furthermore, we fitted linear functions to the RTs using the following formula:

$$f(x) = mx + n,$$

where m is the slope, n is the $f(x)$-intercept and x is the respective test phase. This was repeated for each session to obtain a distribution of slopes for each animal. These distributions were then compared across species using a Wilcoxon rank sum test (two-sided).

Finally, we conducted Pearson correlations to investigate the relationship between RT and behavioral performance as a percentage for each match numerosity and test phase.

4.7. SVM classifier

We trained a linear multiclass support vector machine (SVM) classifier using one-versus-one classification (Chang & Lin, 2011) on response times evaluated for four matching numerosities and three test phases (i.e. 12 different classes). We only used sessions in which we could determine at least seven RTs for each class. This criterion was met in 18 sessions with Monkey 1, 29 sessions with Monkey 2, and 20 sessions with Crow 1 and Crow 2. Seven-fold cross-validation was performed, resulting in six trials for training and one trial for testing per class. Within each cross-validation loop, a new training and test subset was selected. We repeated the training and testing process 150 times, selecting a new set of trials at random with each iteration. The tested SVM classifier provides predicted class labels for the trials in the test subset. This enabled us to compute the main decoding success, as well as construct confusion matrices in combination with the true class labels of the test subset. Both possess an accuracy measure expressed as a percentage. We evaluated the decoding performance of the SVM classifier in balanced pseudo-populations of sessions, drawing a new set of equal size with each iteration (the decisive factor was the session number of Monkey 1). To create a chance-level performance distribution, we randomly shuffled the association between firing rates and trial class labels (numerosity) with each iteration. We tested the distributions of decoding success between animals using a Wilcoxon rank sum test (two-sided, with alpha correction).

CRediT authorship contribution statement

Tobias Machts: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Maximilian E. Kirschhock:** Writing – review & editing, Investigation, Conceptualization. **Andreas Nieder:** Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

None.

Acknowledgments

This work was supported by a DFG grants NI 618/15-1 and NI 618/17-1 (FOR 5159) to A.N.

Data availability

The data and code underlying all the findings of this manuscript are available on Figshare (DOI: https://doi.org/10.6084/m9.figshare.31045399).

References

Bobrowicz, K., & Osvath, M. (2019). Cognition in the fast lane: Ravens' gazes are half as short as humans' when choosing objects. *Animal Behavior and Cognition, 6*(2), 81–97.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, 282*(5389), 746–749.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science, 17*(5), 401–406.

Castaldi, E., Piazza, M., & Eger, E. (2021). Resources underlying visuo-spatial working memory enable veridical large numerosity perception. *Frontiers in Human Neuroscience, 15*, Article 751098.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*, 27.

Ditz, H. M., & Nieder, A. (2015). Neurons selective to the number of visual items in the corvid songbird endbrain. *Proceedings of the National Academy of Sciences of the United States of America, 112*(25), 7827–7832.
Ditz, H. M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of the Royal Society B: Biological Sciences, 283*(1827), 20160083.
Ditz, H. M., & Nieder, A. (2020). Format-dependent and format-independent representation of sequential and simultaneous numerosity in the crow endbrain. *Nature Communications, 11*(1), 686.
Ester, E. F., Ho, T. C., Brown, S. D., & Serences, J. T. (2014). Variability in visual working memory ability limits the efficiency of perceptual decision making. *Journal of Vision, 14*(4), 2.
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.
Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D., & Dyer, A. G. (2018). Numerical ordering of zero in honey bees. *Science, 360*(6393), 1124–1126.
Jannasch, L. L., Grüb, J., & Nieder, A. (2025). Shared cognitive biases influence numerical judgments in macaques and crows. *Proceedings of the National Academy of Sciences of the United States of America, 122*(44), Article e2512219122.
Jordan, K. E., & Brannon, E. M. (2006). Weber's Law influences numerical representations in rhesus macaques (Macaca mulatta). *Animal Cognition, 9*(3), 159–172.
Kirschhock, M. E., Ditz, H. M., & Nieder, A. (2021). Behavioral and neuronal representation of numerosity zero in the crow. *Journal of Neuroscience, 41*(22), 4889–4896.
Kirschhock, M. E., & Nieder, A. (2022). Number selective sensorimotor neurons in the crow translate perceived numerosity into number of actions. *Nature Communications, 13*(1), 6913.
Kobylkov, D., Mayer, U., Zanon, M., & Vallortigara, G. (2022). Number neurons in the nidopallium of young domestic chicks. *Proceedings of the National Academy of Sciences of the United States of America, 119*(32), Article e2201039119.
Liao, D. A., Brecht, K. F., Veit, L., & Nieder, A. (2024). Crows "count" the number of self-generated vocalizations. *Science, 384*(6698), 874–877.
Machts, T., Grüb, J., & Nieder, A. (2025). Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices. *Progress in Neurobiology, 251*, Article 102794.
Machts, T., & Nieder, A. (2026). Coordinated parieto–frontal neuronal communication is critical for abstract quantity judgments in primates. *Cell Reports*. https://doi.org/10.1016/j.celrep.2026.117147. Advance online publication.
Merten, K., & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience, 21*(2), 333–346.
Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences of the United States of America, 109*(29), 11860–11865.
Nieder, A. (2020). The adaptive value of numerical competence. *Trends in Ecology & Evolution, 35*(7), 605–617.
Nieder, A. (2021). The evolutionary history of brains for numbers. *Trends in Cognitive Sciences, 25*(7), 608–621.
Nieder, A. (2025). The calculating brain. *Physiological Reviews, 105*(1), 267–314.
Nieder, A. (2026). The making of number: From content to representation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2025.12.011. S1364-6613(25)00360–2. Advance online publication.
Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science, 313*(5792), 1431–1435.
Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*(1), 149–157.
Nieder, A., & Miller, E. K. (2004). Analog numerical representations in rhesus monkeys: Evidence for parallel processing. *Journal of Cognitive Neuroscience, 16*(5), 889–901.
Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences of the United States of America, 113*(26), 7255–7260.
Orlov, T., Yakovlev, V., Amit, D., Hochstein, S., & Zohary, E. (2002). Serial memory strategies in macaque monkeys: Behavioral and theoretical aspects. *Cerebral Cortex, 12*(3), 306–317.
Pomè, A., Anobile, G., Cicchini, G. M., Scabia, A., & Burr, D. C. (2019). Higher attentional costs for numerosity estimation at high densities. *Attention, Perception & Psychophysics, 81*(8), 2604–2611.
Potrich, D., Zanon, M., & Vallortigara, G. (2022). Archerfish number discrimination. *eLife, 11*, Article e74057.
Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.
Railo, H., Koivisto, M., Revonsuo, A., & Hannula, M. M. (2008). The role of attention in subitizing. *Cognition, 107*(1), 82–104.
Rugani, R., Fontanari, L., Simoni, E., Regolin, L., & Vallortigara, G. (2009). Arithmetic in newborn chicks. *Proceedings. Biological sciences, 276*(1666), 2451–2460.
Scarf, D., Hayne, H., & Colombo, M. (2011). Pigeons on par with primates in numerical competence. *Science, 334*(6063), 1664.
Stancher, G., Rugani, R., Regolin, L., & Vallortigara, G. (2015). Numerical discrimination by frogs (Bombina orientalis). *Animal Cognition, 18*(1), 219–229.
Vetter, P., Butterworth, B., & Bahrami, B. (2008). Modulating attentional load affects numerosity estimation: Evidence against a pre-attentive subitizing mechanism. *PLoS One, 3*(9), Article e3269.
Wagener, L., Loconsole, M., Ditz, H. M., & Nieder, A. (2018). Neurons in the endbrain of numerically naive crows spontaneously encode visual numerosity. *Current Biology, 28*(7), 1090–1094.e4.
Weber, E. H. (1851). *Die Lehre vom Tastsinne und Gemeingefühle: Auf Versuche gegründet*. Braunschweig, Germany: Friedrich Vieweg und Sohn.
Yakovlev, V., Bernacchia, A., Orlov, T., Hochstein, S., & Amit, D. (2005). Multi-item working memory–a behavioral study. *Cerebral Cortex, 15*(5), 602–615.
Zhang, H., Zhen, Y., Yu, S., Long, T., Zhang, B., Jiang, X., … Wang, L. (2022). Working memory for spatial sequences: Developmental and evolutionary factors in encoding ordinal and relational structures. *Journal of Neuroscience, 42*(5), 850–864.