PNAS RESEARCH ARTICLE | NEUROSCIENCE

Shared cognitive biases influence numerical judgments in macaques and crows

Lena L. Jannasch[a], Julia Grüb[a], and Andreas Nieder[a,1]

Edited by Ranulfo Romo, El Colegio Nacional, Mexico City, D.F., Mexico; received May 14, 2025; accepted September 25, 2025

Cognitive biases, which come from the shortcuts our brains use to make decisions, often lead people to misjudge sizes, amounts, or other magnitudes. However, it is still unclear how these biases evolved. Nonhuman animals like primates and corvids share with humans a nonsymbolic number sense. Whether biases in magnitude estimation are evolutionarily widespread in animals remains an open question. We investigated numerical estimation in macaque monkeys and carrion crows—two distantly related but numerically proficient vertebrates—using a delayed match-to-numerosity task with dot arrays. Both species exhibited key biases commonly observed in humans during numerical estimation tasks: scalar variability (increasing response variability with numerosity), regression to the mean (overestimation of small and underestimation of large numerosities), and a sequential effect (estimates biased toward previous trial numerosities). Traditionally, these biases have been attributed either to static influences of overall task history or to dynamic effects from recent trials. Here, we demonstrate that a Bayesian model incorporating dynamic priors based on multiple previously encountered numerosities can account for all observed behavioral patterns. Our findings suggest that numerical judgments in both species are shaped by uncertainty and the integration of recent experience, revealing shared Bayesian-like mechanisms for cognitive biases across distantly related animals.

cognitive bias | number | rhesus macaque | carrion crow | evolution

Significance

Cognitive biases, often seen as uniquely human flaws, may instead reflect fundamental decision-making strategies shared across species. By demonstrating that macaque monkeys and carrion crows exhibit human-like biases in numerical estimation—including regression to the mean and sequential effects—our study reveals that these distortions arise from deep-rooted cognitive mechanisms, not symbolic reasoning. These patterns reflect a Bayesian estimation process that reduces sensory uncertainty by integrating prior experience. Our results show that combining multiple past stimuli best explains the cognitive biases observed in monkeys and crows. Our findings offer crucial insights into the evolutionary origins of mental shortcuts.

Judgments made by both humans and animals often deviate from physical reality due to cognitive biases, which arise from the brain's use of mental shortcuts, or heuristics. While these shortcuts enhance efficiency, they can lead to predictable distortions in perception, memory, and reasoning (1–3). Studying cognitive biases is crucial for understanding the limits of rational decision-making and the mechanisms by which the brain simplifies complex information (4). Researching these biases in animals not only reveals how they perceive the world but also offers insights into the evolutionary origins and phylogenetic roots of decision-making strategies in humans.

Humans exhibit cognitive biases when judging numbers. For example, both anchoring and repulsion effects have been observed, where distant numerical values attract judgments, while nearby values tend to repel them (5). In addition, extreme numerical judgments shift closer to the mean of the tested range (6) and infrequent numbers are encoded with greater variability compared to commonly encountered ones (7). Whether these biases in numerical judgment are uniquely tied to humans and their symbolic understanding of numbers, or whether they reflect more widespread and evolutionarily inherited features of cognition, remains elusive.

The origins of abstract human numerical abilities lie in an evolutionarily shared, intuitive sense of nonsymbolic number found across the animal kingdom (8). This "number sense" is supported by the approximate number system, which represents numerosity along a logarithmically compressed mental number line, consistent with the Weber–Fechner law (9–11). Many species from diverse phylogenetic backgrounds can assess the number of items in a set (8). Primates and corvid songbirds exhibit similarly sophisticated, abstract numerical abilities (12, 13), despite the independent evolution of endbrain structures in mammals and birds since their last common ancestor over 320 Mya (14).

We hypothesized that monkeys—close nonhuman primate relatives of humans—and perhaps even distantly related crows might exhibit cognitive biases in numerical estimation. To test this, we assessed macaques and crows using a visual delayed match-to-numerosity task, in which they judged absolute sample numerosities relative to both smaller and larger deviants. This approach allowed us to derive detailed performance functions for a quantitative evaluation of bias effects. We aimed to identify heuristics that are deeply rooted in evolutionary history suggesting that they shape human numerical decision-making.

Author affiliations: [a]Animal Physiology Unit, Institute of Neurobiology, Department of Biology, University of Tübingen, Tübingen 72076, Germany

Author contributions: L.L.J., J.G., and A.N. designed research; performed research; analyzed data; and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: andreas.nieder@uni-tuebingen.de.

Published October 27, 2025.

One of the influential perspectives on magnitude estimation, the Bayesian perspective, posits that individuals combine uncertain sensory information with prior knowledge to optimize their judgments (15). This integration process can lead to systematic biases. The Bayesian framework predicts three key effects that we set out to test in our animals: 1. scalar variability, 2. regression to the mean, and 3. sequential effects (Fig. 1). Scalar variability refers to the proportional increase in the standard deviation (SD) of estimates with increasing magnitude, consistent with the Weber–Fechner law (9, 12, 13). Regression to the mean describes a bias in which estimates are pulled toward the center of the stimulus range, resulting in a compressed perceived range compared to the actual tested values. This bias is more pronounced at the extremes of the tested range and for larger tested ranges—a phenomenon often referred to as the range effect. Finally, sequential effects reflect a dynamic bias whereby current estimates are systematically shifted toward the magnitude of previous trials, leading to a smoothing or carryover effect across trials.

Results

Numerical Discrimination Performance. We tested the hypothesis that magnitude estimation in two phylogenetically distant vertebrate species is shaped by Bayesian integration of sensory information and prior expectations, leading to systematic biases under uncertainty (Fig. 1) (15). We trained two macaque monkeys (*Macaca mulatta*) and two carrion crows (*Corvus corone*) on a computer-controlled numerical discrimination task. This delayed match-to-numerosity task required the animals to flexibly judge whether two visual numerosities that were temporally separated by a delay were the same (Fig. 2*A*). After trial initiation, a dot pattern with a variable number of dots from 1 to 10 (the "sample") was shown to the subject. After a brief delay period, a matching (same numerosity) or nonmatching test display (deviant numerosity) was shown each in 50% of the cases. Within a trial, the subject was required to respond only when the test numerosity matched the sample numerosity. To prevent subjects from discriminating nonnumerical low-level visual features, dot position varied randomly within and between trials (standard condition), and control stimuli with equal cumulative dot area and density but varying dot size across numerosities were shown in half the trials (control conditions; Fig. 2*B*).

We collected the behavioral responses from 75 sessions for monkey #1 with an average of 530 ± 17 trials per session (mean ± SEM), and 57 sessions for monkey #2 with an average of 965 ± 24 trials per session. For crow #1 and crow #2, we collected data from 18 and 17 sessions, respectively, with an average of 555 ± 18 and 579 ± 29 trials per session. All subjects performed above chance in all sessions and for both stimulus conditions [all *P*-values < 0.001; binomial test]. For monkey #1, monkey #2, crow #1, and crow #2, the average performance under standard conditions was 72.9 ± 0.6%, 78.5 ± 0.4%, 78.6 ± 0.9% and 70.1 ± 0.9%, respectively, while under control conditions it was 70.0 ± 0.5%, 75.2 ± 0.5%, 81.0 ± 1.1% and 74.3 ± 1.2% (mean ± SEM).

For each session and individual, we analyzed the detailed response functions (behavioral representations) per sample numerosity 1 to 10 based on the response frequencies to different test numerosities. Fig. 3 illustrates the average response functions for both monkeys and both crows on a logarithmic scale. They show the expected bell shape with a peak at the corresponding sample numerosity (correct response rate in match trials) and monotonically decreasing flanks to both sides (erroneous response rate in nonmatch trials) (12, 13). The increase in peak accuracy with higher sample numerosities—particularly in monkeys—likely reflects both perceptual and methodological factors (*Methods*). This trend is less pronounced in crows, possibly due to fewer training sessions. For further analyses, Gaussian functions were fitted to the response functions to quantify their center, width, and symmetry. Sample numerosity 1 was excluded from the further analysis, as subjects were only required to distinguish it from larger numerosities.

Scalar Variability. To test whether the SD of response functions increased linearly with increasing sample numerosities, we assessed the width of the Gaussian fits to each of the subjects' response functions. For all four subjects, the width of the fits was positively correlated with increasing sample numerosities [$r(646) = 0.71$ for monkey #1, $r(502) = 0.81$ for monkey #2, $r(160) = 0.89$ for crow #1, $r(149) = 0.82$ for crow #2, all $P < 0.001$]. Fig. 4 shows the average width of the Gaussian fits per individual across sample numerosities. Linear fits all yielded positive slopes with mean slopes of 0.32, 0.28, 0.45, and 0.44 for monkey #1 and #2 and crow #1 and #2, respectively, that were significantly different from 0 [$t(75) = 25.2$, $t(56) = 40.2$, $t(17) = 18.8$ and $t(16) = 16.4$ respectively, all $P < 0.001$]. The width of the performance functions systematically increased with increasing numerical values for all subjects.

Regression to the Mean. To test whether numerosity estimates were biased toward the center of the tested numerosity range, we quantified the centers of the derived Gaussian fits to the performance curves. If the subject was more likely to confuse the sample with smaller rather than larger numerosities (i.e., underestimated the test numerosity), the center of the Gaussian fit would be biased toward smaller numerosities than the sample numerosity (Fig. 5*A*, green curves). Conversely, if the subject overestimated the sample numerosity, the center of the Gaussian fit

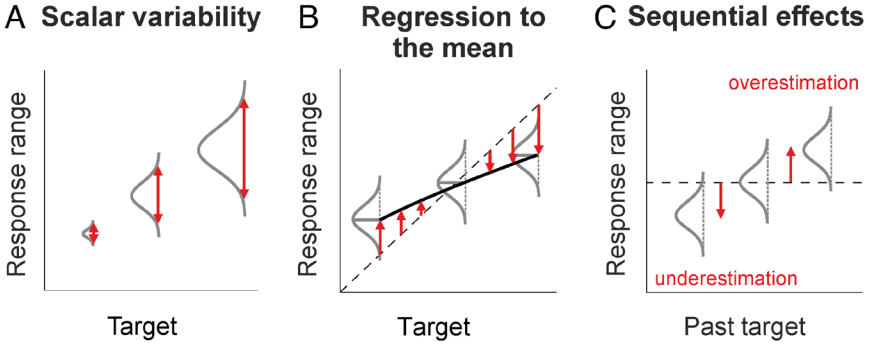

Fig. 1. Behavioral effects of magnitude estimation predicted by a Bayesian framework (15). (*A*) Scalar variability: The SD of behavioral representations (bell-shaped response functions) increases linearly with increasing magnitudes. (*B*) Regression to the mean: The magnitude estimates become biased toward the center of the tested magnitude range. (*C*) Sequential effects: Estimates become dynamically biased toward previously seen magnitudes.

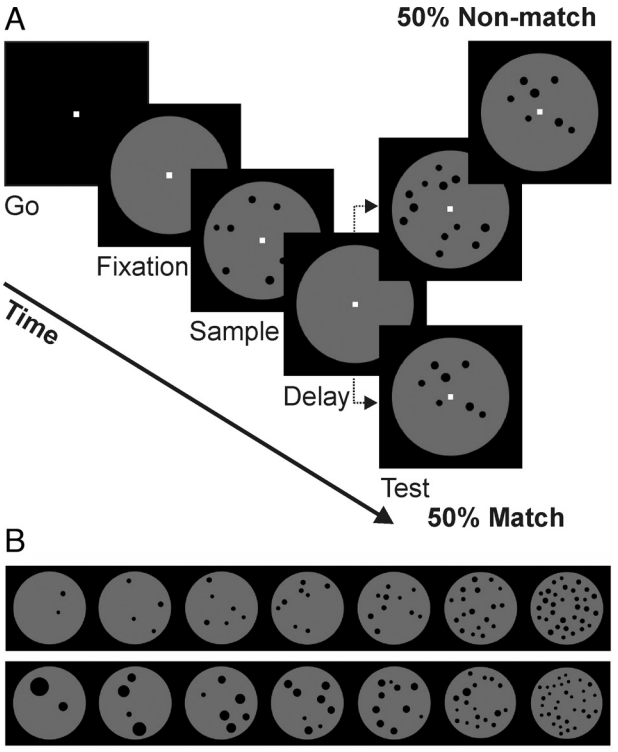

Fig. 2. DMN task. (*A*) Behavioral protocol: A trial was initiated when the animals oriented toward the displays and fixated (300 ms). Subjects were presented with a sample numerosity (500 ms for monkeys, 600 ms for crows) that they had to memorize for 1,000 ms during the delay period. They were required to respond once the same numerosity was shown again in the test period (800 ms for monkeys, 900 ms for crows). In 50% of the trials (match trials), the test numerosity corresponded to the sample numerosity. If the animals responded to the match, they received a reward. The other 50% of the trials were nonmatch trials in which the test numerosity was smaller or larger than the sample numerosity. Here, the animals were not allowed to respond and had to wait until the matching numerosity was presented after the first test period. (*B*) Nonmatch numerosities ranged up to 32. All stimuli were presented as standard and control conditions, in which cumulative area and density were equalized.

would be shifted toward values larger than the sample numerosity (Fig. 5*A*, blue curves).

Fig. 5 *B–E* shows the performance functions' mean center values indicative of the judged numerical value, averaged across sessions per sample numerosity. Small sample numerosities were systematically overestimated, while large sample numerosities were underestimated, resulting in a regression to the mean effect. For all individuals, the shifted centers were well described by a linear fit on a log–log scale with goodness of fit values r^2 = 0.86, 0.93, 0.91, and 0.87 for monkey #1, monkey #2, crow #1, and crow #2, respectively. The slopes of the linear fits were significantly smaller than 1, with mean slopes of 0.78, 0.82, 0.80, and 0.77, respectively [t(75) = −12.8, t(56) = −16.4, t(17) = −10.5, t(16) = −8.7, all $P < 0.001$]. Gaussian fits with centers fixed at the sample numerosity were significantly better (measured by the goodness of fit r^2) on a logarithmic scale compared to a linear scale for all subjects [Bonferroni corrected pairwise comparison, Z = −7.22 for monkey #1, Z = −6.57 for monkey #2, Z = −3.68 for crow #1, Z = −3.62 for crow #2, all $P < 0.001$]. Mean goodness of fit values on a logarithmic scale further increased significantly when the fixed centers were shifted according to a regression to the mean of the tested range with a slope of 0.79 (e.g., 3 would be fixed at 3.41 and 8 would be fixed at 7.39). For all subjects, this led to a significantly better fit [Z = −6.88 for monkey #1, Z = −5.67 for monkey #2, Z = −3.72 for crow #1, Z = −3.43 for crow #2, all $P < 0.001$]. This indicates that performance functions were most symmetric when plotted on a logarithmic scale, with centers shifted from the sample numerosity toward the center of the tested range. The observed regression to the mean provides evidence for the influence of the tested range, the complete task and/or training history, on single trial outcomes.

Sequential Effects. To test for a dynamic influence of task history on numerosity judgments, we examined the effect of the immediate past from one trial to the next. The presence of a sequential effect would predict that, if the sample numerosity in the previous trial was larger (or smaller) than the sample numerosity of the current trial, the animal would respond more often to a nonmatch numerosity that was also larger (or smaller) than the current sample numerosity. Therefore, we examined the proportion of erroneous responses to the nonmatching test-display in nonmatch trials as a function of the sample numerosity of previous (match and nonmatch) trials (Fig. 6*A*).

The trials were split into two groups: In one group (congruent group), the sample numerosity of the previous trial and the nonmatching test numerosity of the current trial were both smaller or larger relative to the sample numerosity of the current trial. In the other group (incongruent group), the sample numerosity of the previous trial was smaller (or larger) than the sample numerosity of the current trial, when the nonmatching test numerosity of the current trial was larger (or smaller) than the sample numerosity of the current trial. Only nonmatch trials of central sample numerosities 3 to 7 were included in the analysis of the sequential effect, as these allowed us to test directly whether subjects were more likely to judge the sample numerosity as smaller or larger than its actual value. Fig. 6*B* depicts the effect of the previous sample (=match) numerosity S/M*1* on error rates of the current trial. We found that for all subjects, the congruent group had significantly higher error rates than the incongruent group [chi-square test for independence $X^2(1)$ = 86.5, n = 7,047 for monkey #1, $X^2(1)$ = 86.6, n = 10,901 for monkey #2, $X^2(1)$ = 43.5, n = 1,844 for crow #1, $X^2(1)$ = 19.6, n = 1,678 for crow #2, all $P < 0.001$]. To control for possible interaction effects with the regression to the mean, we separately analyzed trials of sample numerosity 5 only—combined per species to account for lower trial numbers. Again, the error rates of the congruent and incongruent group differed significantly [$X^2(1)$ = 19.4, n = 3,686, $P < 0.001$ for monkeys and $X^2(1)$ = 14.2, n = 723, $P < 0.001$ for crows]. Our results indicate that monkeys and crows biased their numerical judgments toward the sample (=match) numerosity of the previous trial.

We analyzed this effect in more detail and found that the sequential effect is characterized by a numerical distance dependency. For all subjects, the average difference in error rates increases with numerical distance between the current and previous sample numerosity S/M*1* [$r(415)$ = 0.31 for monkey #1, $r(338)$ = 0.37 for monkey #2, $r(97)$ = 0.36 for crow #1, $r(95)$ = 0.49 for crow #2, all $P < 0.001$]. Linear regressions resulted in all positive slopes with mean values of 4.1, 3.3, 4.2, and 6.0 for monkey #1, monkey #2, crow #1, and crow #2 as depicted in Fig. 6 *C–F* and significantly different from 0 [t(75) = 5.1, $P < 0.001$, t(56) = 6.8, $P < 0.001$, t(17) = 2.6, $P = 0.02$ and t(16) = 6.7, $P < 0.001$ respectively]. We did not find a clear correlation between the sequential effect and the elapsed time between the previous and current sample onset [$r(150)$ = −0.05, $P = 0.57$ for monkey #1, $r(112)$ = −0.12,

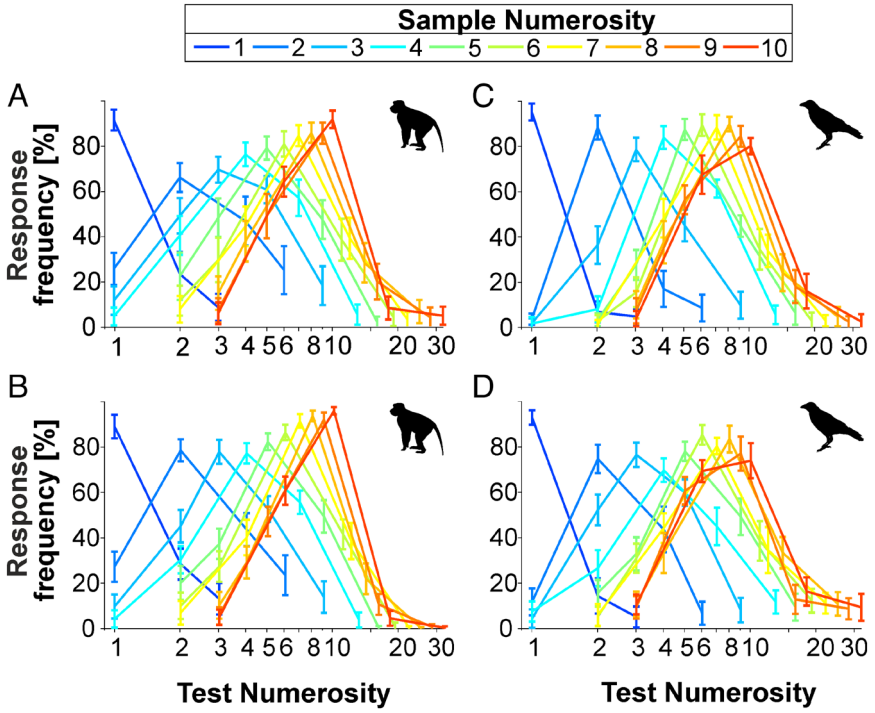

Fig. 3. Response functions per subject. Performance functions for monkey #1 (*A*), monkey #2 (*B*), crow #1 (*C*), and crow #2 (*D*). Response functions resembled bell-shaped curves with a peak at the sample numerosity (related to correct response rate in match trials) and monotonically decreasing flanks to both sides (related to erroneous response rate in nonmatch trials). Depicted are mean ± SEM across sessions.

P = 0.20 for monkey #2, $r(34)$ = 0.15, P = 0.39 for crow #1, $r(32)$ = −0.26, P = 0.14 for crow #2].

Samples further in the past, of the last to previous trial also influenced the numerical judgment. The error rates of the congruent group of the last to previous trial (with respect to its sample and match numerosity S/M*2*) are again higher than those of the incongruent group with an average difference across individuals of 4.7 ± 1.0%, compared to 11.0 ± 1.2% for the previous trial S/M*1*. This sequential effect with respect to the second last sample S/M*2* reaches significance for three out of four subjects [$X^2(1)$ = 7.3, n = 5,025, P = 0.007 for monkey #1, $X^2(1)$ = 8.0, n = 8,317, P = 0.004 for monkey #2, $X^2(1)$ = 9.6, n = 1,388, P = 0.002 for crow #1, $X^2(1)$ = 2.4, n = 1,174, P = 0.121 for crow #2]. Fig. 6 *G–J* shows the average error rate differences between congruent and incongruent trials with respect to the sample and match numerosities of the past four trials (S/M*4* being the sample of the 4th last trial). A positive correlation between the sequential effect and task history is significant for all subjects except for crow #2 [$r(271)$ = 0.16, P = 0.009 for monkey #1, $r(226)$ = 0.24, P < 0.001 for monkey #2, $r(64)$ = 0.42, P < 0.001 for crow #1, $r(57)$ = 0.10, P = 0.45 for crow #2]. A linear regression for the significant correlations performed per session resulted in slopes significantly different from 0 [t(73) = 2.84, P = 0.006 for monkey #1, t(56) = 3.62, P < 0.001 for monkey #2, t(16) = 5.19, P < 0.001 for crow #1]. The number of trials decreased with greater task history depth, increasing measurement uncertainty—especially in crows, where fewer trials were recorded overall.

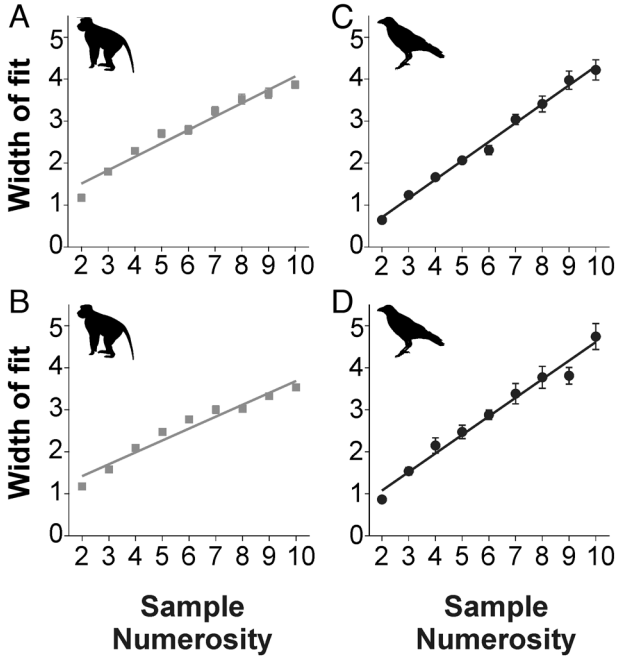

Fig. 4. Scalar variability effect. The width of the response functions increased linearly with increasing numerical value. A linear regression fitted to the mean values resulted in slopes greater than 0 for monkey #1 (*A*), monkey #2 (*B*), crow #1 (*C*), and crow #2 (*D*). Error bars indicate ± SEM across sessions.

Bayesian Observer Model. Last, we simulated trial outcomes using a Bayesian-like integration of prior information to examine how different priors influence the observed biases. The model operates on a logarithmically compressed mental number line, consistent with the Weber–Fechner law of numerical perception across species (9, 16). The observer combines noisy sensory input (likelihood) with internal beliefs shaped by prior experience (prior) to form a posterior estimate, which drives the behavioral response. Regression to the mean is typically modeled by a static prior centered within the tested range, while sequential effects arise from a dynamic prior anchored to the previous sample numerosity (6, 17–21). Therefore, we implemented a fixed model structure (Fig. 7*A*) with priors at different locations influencing the internal estimation process to compare their predicted outcomes.

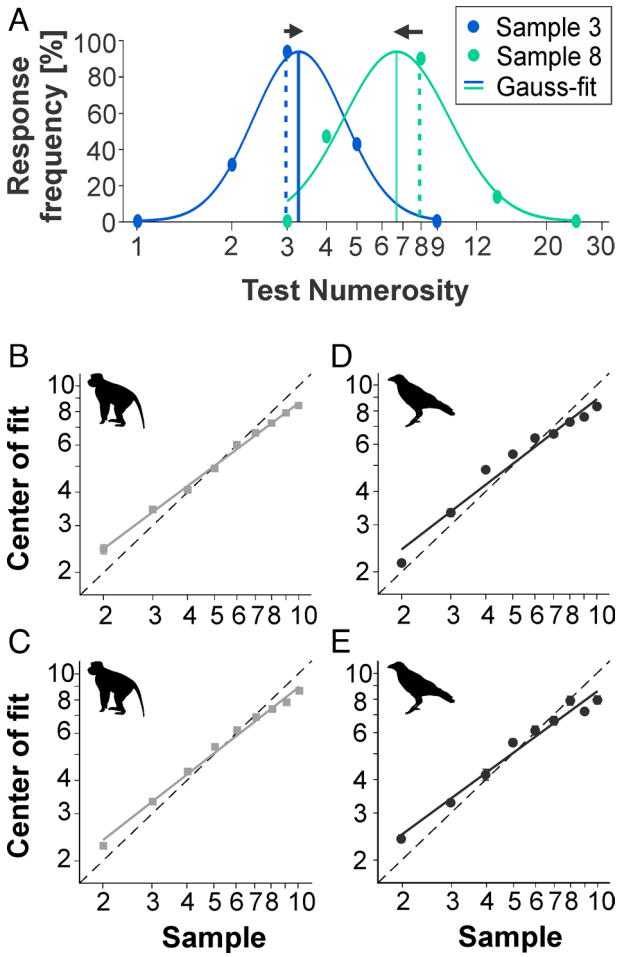

Fig. 5. Regression to the mean effect. (*A*) Example Gaussian fits to behavioral data of sample numerosities 3 (blue, overestimated) and 8 (green, underestimated) on a logarithmic scale. (*B–E*) show the average centers of the fits per sample numerosity across sessions. A linear regression fitted to the mean values resulted in slopes smaller than 1 for monkey #1 (*B*), monkey #2 (*C*), crow #1 (*D*), and crow #2 (*E*). SEMs are smaller than 0.05 and lie within the marker edges.

A perceptual function f transforms the sample numerosity into a noisy sensory representation, modeled as a lognormal distribution with mean μ_n centered on the sample numerosity (likelihood). The prior is also modeled as a lognormal distribution, with its mean μ_P varying according to the specific model. The posterior is computed by integrating the likelihood and prior via Bayes' rule, resulting in a lognormal distribution. In line with the Weber–Fechner law, the variances of the likelihood, prior, and posterior are assumed constant on a logarithmic scale. The posterior mean μ_{post} reflects a weighted shift of the likelihood mean μ_n toward the prior mean μ_p with weight $w \in (0,1)$. This reflects Bayes-optimal weighting, where the prior exerts the greatest influence—on a linear scale—at higher numerosities, where sensory input is least reliable, and the least influence at lower, more precise numerosities. The response function r transforms the posterior distribution back to the linear stimulus scale, producing probabilities of responding to specific test numerosities (i.e., judging them as equal to the sample). To control for effects of stimulus statistics and analysis methods, we simulated single-trial outcomes and compared them to the subjects' responses.

To determine which prior best explains each subject's behavior, we fit models with different prior locations separately to each monkey and crow. The baseline model (M0), without a prior, and models M1–M3 all showed scalar variability, evidenced by a positive correlation between Gaussian fit widths and sample numerosity (Fig. 7*B*). However, when comparing model-predicted scalar variability slopes to behavioral data, models M0–M3 showed significantly steeper slopes than observed in monkeys #1 and #2 [post hoc pairwise comparison, all $P < 0.001$]. Model M0 did not produce regression to the mean, as indicated by slopes near 1 for the centers of Gaussian fits across sample numerosities, nor a sequential effect [yellow markers in Fig. 7 *C–E*; post hoc comparisons: regression to the mean for monkeys #1, #2 and crows #1, #2 versus M0, $P < 0.001$; sequential effect for monkeys #1, #2 and crow #1 versus M0, $P < 0.01$].

In M1, we implemented a static prior centered at the sample range midpoint ($\mu_P = 5.5$), representing the influence of the entire task history on perception. The model's free parameter w was fit to replicate the Gaussian centers across sample numerosities, successfully capturing the regression to the mean observed in the animals (Fig. 7*C*). Statistically, the implemented regression to the mean in M1 also produced a sequential effect with numerical distance dependency (Fig. 7*E*). However, this sequential effect did not diminish with task history (Fig. 7*D*), and the error rate difference relative to the previous trial was smaller than observed, reaching significance in monkey #2 [post hoc comparison versus M2: $P < 0.01$].

A continuous bias toward previous stimuli (sequential effect) may accumulate over trials, producing an overall regression to the mean. To test this, model M2 incorporated a bias toward the perceived sample from the immediately preceding trial.

Since subjects lacked direct access to the physical magnitude of the previous trial's sample, we implemented a recursive bias toward the internal perception of that sample (the previous trial's posterior), following a similar approach to ref. 18 on number-to-space mapping in humans. The weight w of the prior of M2 was fit to reproduce a sequential effect toward the previous sample S/M*1* as observed in the animal subjects. Model M2 predicts a sequential effect that diminishes with task history and shows numerical distance dependency (Fig. 7 *D–E*). However, its predicted regression to the mean is smaller than observed in the animals' behavior [post hoc comparisons for monkeys #1, #2, and crow #2 versus M2: $P < 0.01$].

A direct influence of past stimuli—unlike the purely recursive effect in M2—amplifies their impact on current perception and strengthens the regression to the mean. Therefore, in model M3, the prior was chosen as a weighted sum of the past eight sample numerosities. The free parameter w was fit to replicate the centers of Gaussian fits across sample numerosities. We found that biasing estimates toward multiple past stimuli fully accounted for both the regression to the mean and the sequential effect observed in the animals [green markers in Fig. 7 *C–E*; post hoc comparisons versus M3 for monkeys #1, #2, crow #1, and crow #2: all $P = 1.0$]. Fig. 7 *F–I* show simulated trial outcomes per subject based on model M3, illustrating how biases in numerical judgments in monkeys and crows can be fully explained by Bayesian integration on a logarithmic mental number line that incorporates information from multiple past stimuli.

Discussion

In our study, we tested numerical discrimination in two phylogenetically distant vertebrate species, macaque monkeys and carrion crows, using a delayed match-to-numerosity task. Our findings revealed key patterns of numerical processing similar to those observed in humans, including scalar variability, a regression to the mean, and a sequential effect. Scalar variability was evident, with the width of the

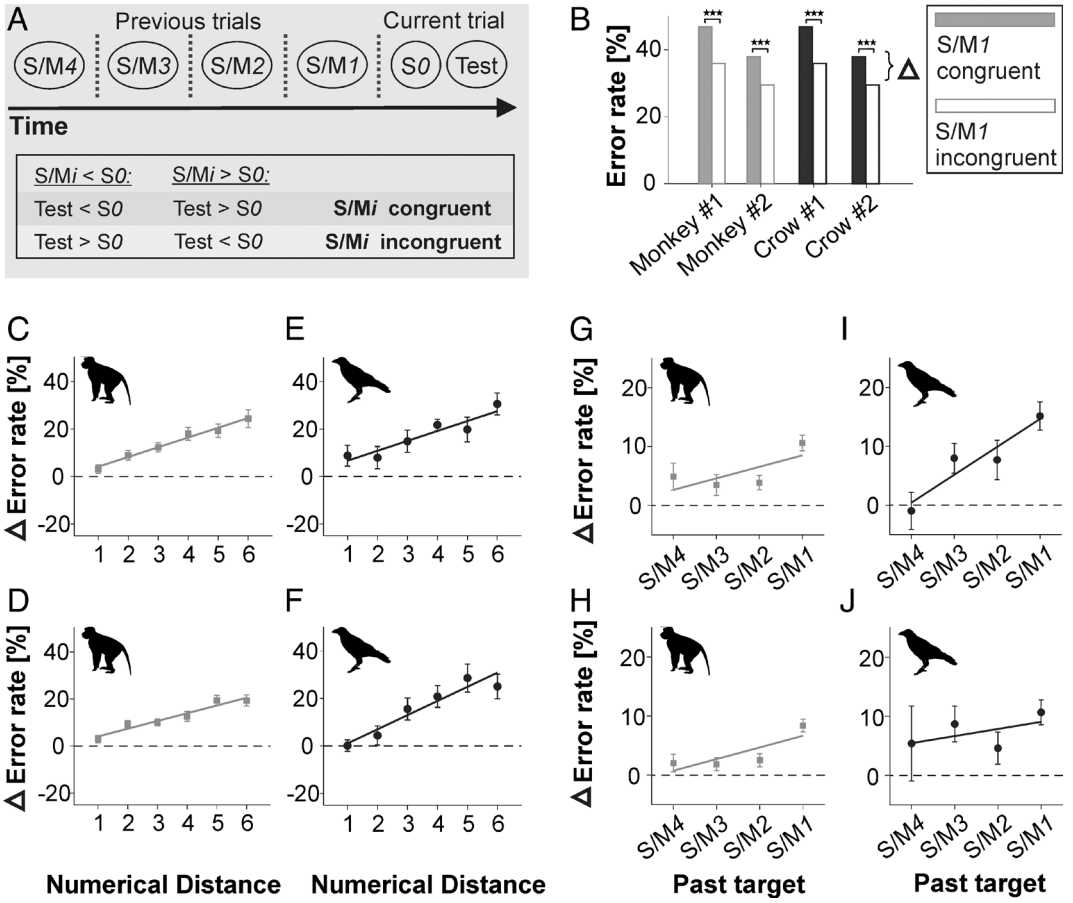

Fig. 6. Sequential effect. (*A*) Error rates were evaluated depending on whether the sample (=match) numerosity of the previous trial (S/M*1*), 2nd previous trial (S/M*2*), 3rd (S/M*3*), and 4th previous trial (S/M*4*) were congruent/incongruent to the nonmatching test numerosity in the current trial with respect to the current sample. (*B*) Subjects were more likely to respond erroneously to a smaller (larger) nonmatching test numerosity if the previous sample was smaller (larger) as well. ***$P < 0.001$. (*C–F*) Show the difference in error rates (Δ with respect to S/M*1*) depending on the numerical distance between the current and previous sample numerosity. The error rate increased progressively with increasing numerical distance. A linear regression fitted to the mean values results in positive slopes for monkey #1 (*C*), monkey #2 (*D*), crow #1 (*E*), and for crow #2 (*F*). (*G–J*) show the sequential effect extending further into the task history. Error rate differences are depicted with respect to the sample and match numerosities of the past four trials. Again, a linear regression resulted in positive slopes for monkey #1 (*G*), monkey #2 (*H*), crow #1 (*I*), and crow #2 (*J*).

response functions increasing linearly with sample numerosities, reflecting greater uncertainty in numerosity judgments at higher values. We observed a regression to the mean effect, where both species systematically overestimated smaller numerosities and underestimated larger ones, suggesting a bias toward the center of the tested range. Lastly, sequential effects were observed, in which responses in the current trial were influenced by the previous trial's numerosity. These biases reflect both a static influence of the overall tested distribution and dynamic effects of recent task history—two factors that have largely been studied separately. A Bayesian observer model that biases current numerical judgments toward multiple previously seen numerosities can account for all observed behavioral biases. This suggests a shared mechanism whereby the task history shapes numerical judgments in both monkeys and crows.

Scalar Variability. The finding that the width of the response functions in both monkeys and crows increased linearly with increasing sample numerosities supports the idea that numerosity discrimination in these species follows the Weber–Fechner law. As the sample numerosity increases, the variability in the animals' judgments grows proportionally, indicating a scaling effect in which larger numerosities are represented with less precision. These results are consistent with previous studies in both monkeys (12) and crows (13), further confirming that the Weber–Fechner law applies to numerical cognition across species.

Research on numerical cognition revealed that humans, too, possess an approximate number system independent of counting, that is characterized by scalar variability (12, 22, 23). The approximate number system supports basic arithmetic operations in both animals and preverbal infants (24). Scalar variability is also a key feature of other magnitude estimations, such as interval timing or speed discrimination (25, 26). This suggests that scalar variability is a general property of magnitude estimation, which humans share with numerically competent animals. In an optimal Bayesian integration mechanism, the higher precision of small numerosities implies that sensory information is more reliable. As a result, there is less reliance on prior expectations, and sensory evidence dominates the judgment. In contrast, the influence of the prior increases for larger numerosities, as it helps guide the decision-making process in the face of greater uncertainty (6, 27). This combination of sensory information and prior expectations enables more accurate judgments.

Regression to the Mean. The regression to the mean effect observed in both monkeys and crows indicates a systematic bias in numerical estimation, where sample numerosities are perceived as closer to the center of the tested range. Specifically, small numerosities were overestimated, while large numerosities were underestimated, resulting in a compression of the perceived numerical range. This effect, evident as a shift in response function centers toward the mean of the tested range, was consistent across

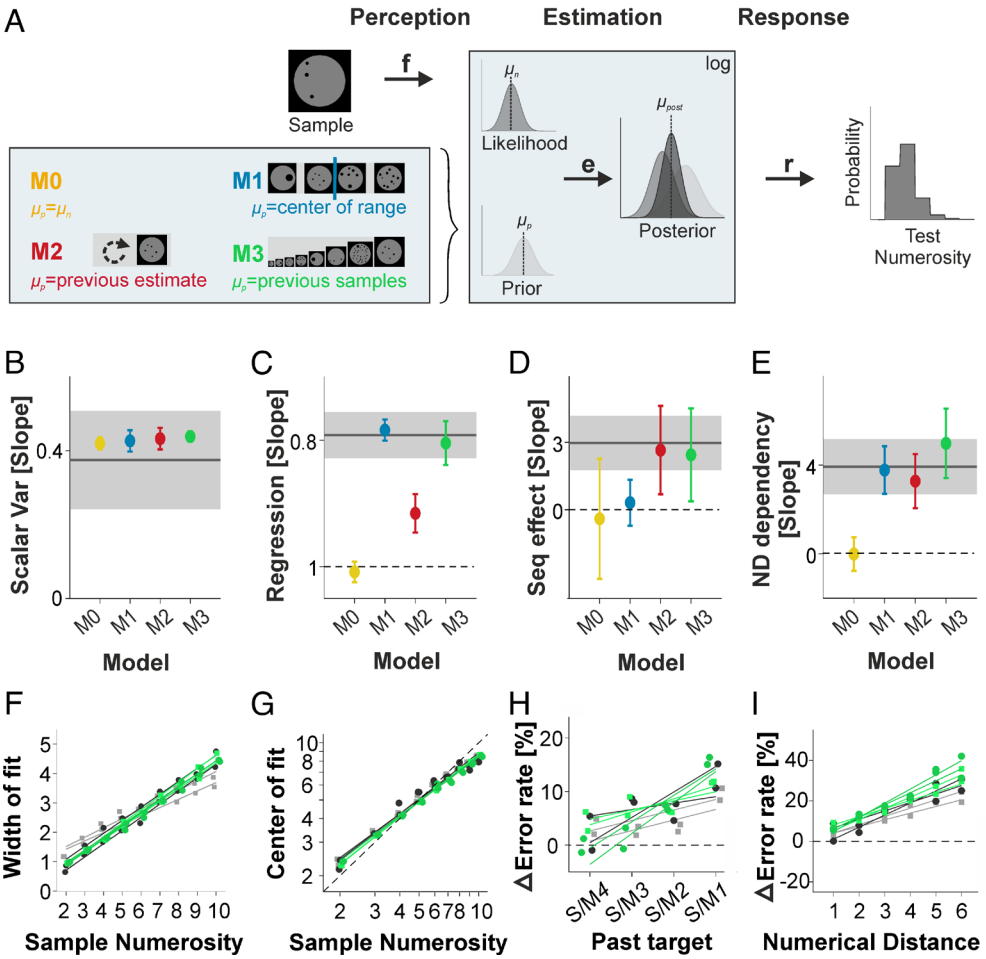

Fig. 7. Biases emerging from a Bayesian model with different priors. (*A*) The basic model structure involves an estimation process on a logarithmic mental number line. A noisy sensory input (likelihood) becomes integrated with prior information according to Bayes' rule to reduce the variability of the estimation process. Responses are simulated for four different priors M0, M1, M2, and M3 per individual. (*B–E*) Illustrate the biases predicted by each model in the form of the slopes of a linear regression fitted to the mean values across all trial outcomes predicted by M0-M3 per individual. Plotted are mean ± 95% CI across individuals. The black line and gray area show the mean ± 95% CI of the subjects' (monkeys and crows) behavioral data. (*F–I*) show the biases resulting from M3 per subject resembling scalar variability (*F*), regression to the mean (*G*) and sequential effects (*H* and *I*) as observed in the behavioral data.

both species and suggests that numerical judgments are influenced by the broader context of the numerical range.

Research on magnitude estimation consistently shows regression to the mean effects in humans (28–31). The phenomenon has been observed in various contexts, including numerical and time judgments, path integration, alphabetic interval estimation, and general probability judgments (17, 29, 31, 32). Interestingly, the regression effect is not limited to adults but is also observed in children's area judgments, where the number sense biases their estimations (33).

Explanations for this bias include Bayesian models that incorporate prior experience in the form of a static prior at the center of the tested range into current estimates (6, 17, 19, 31, 32). Our model demonstrates that regression to the mean could also emerge from a dynamic prior, but only integrating multiple previously seen stimuli can fully explain the strength of the effect observed across all subjects.

Sequential Effect. In our study, we observed a clear sequential effect in both monkeys and crows, where numerical judgments were systematically biased toward the sample numerosity of the preceding trial. It depended on the numerical distance between the current and previous sample and extended to samples that lay further in the past, but with diminishing influence. These findings support the idea of an online, trial-by-trial updating process.

Sequential effects in number discrimination tasks have been observed in humans. They occur in symbolic number comparison tasks, with faster responses when the same number is repeated, independent of the numerical distance effect (34). Sequential processing is a common mechanism in attention orienting and number processing systems, potentially governed by universal mechanisms across perception and cognition (35).

The sequential effect we found is consistent with Bayesian models of cognitive processing, where each trial serves as a new piece of evidence that refines the animal's prior expectations (18, 36, 37). The observed pattern of a diminished effect over longer task histories suggests that the animals are continuously adjusting their internal numerical representations in response to recent experiences. This dynamic updating process might be an adaptive and efficient strategy for maintaining a stable perception of numerosity in the face of inherent sensory noise (37, 38).

Our dataset contains substantially more sessions from monkeys than crows, limiting direct statistical comparisons—especially for sequential effects over multiple trials due to smaller crow sample size. Nevertheless, the crow data are robust enough to confirm that key biases—scalar variability, regression to the mean, and sequential effects—are consistently present in both species.

Interaction of Effects. The three effects we identified have largely been studied separately, and little is known about how they relate to one another. The modeling of different priors shows that scalar variability alone, fails to produce a regression to the mean or a sequential effect. A regression to the mean and a sequential effect coemerge due to inherent statistical dependencies and are hard to disentangle. However, the fact that both effects systematically influenced numerical judgments across all numerosities—including those at the center of the tested range—suggests they may have distinct underlying origins. The modeling results of a prior located at the center of the tested range or at the previous sample numerosity show that one effect can not fully account for the other. This independence of both effects—as well as the statistical dependencies—has also been described for color and line length judgments in humans (39). When studying one phenomenon in magnitude estimation tasks, the other must likewise be accounted for in the analysis.

Alternatively, a combination of a static central prior (as in M1) and a dynamic prior (as in M2) may also explain the observed behavior. Analyses of recurrent spiking neural networks trained on temporal interval discrimination have shown that neuronal firing rates encode information about both the most recent and penultimate stimuli during the current trial (40). The resulting population activity represents a Bayesian estimate of interval duration shaped by the task history. Future behavioral studies could test whether magnitude judgments rely on combined static and short-term dynamic priors or on a dynamic prior spanning longer task histories.

Evolutionary Roots of Cognitive Biases. Our findings in nonhuman primates and corvids shed light on the evolutionary origins of cognitive biases, challenging the idea that these biases are exclusive to primates or humans. While our results suggest that cognitive biases in magnitude estimation are evolutionarily conserved across species, the origin of these similarities—whether from shared ancestry or convergent evolution—remains unclear. One hypothesis proposes that birds and primates inherited these biases from a common ancestor over 320 Mya (14, 41), indicating that such cognitive biases may be a shared trait among all amniotes. Alternatively, convergent evolution may explain the independent emergence of cognitive biases in birds and mammals, as both groups evolved sophisticated cognitive abilities via distinct evolutionary paths (42, 43). This convergence suggests that cognitive biases may also exist in other highly intelligent, distantly related animals. Demonstrating similar biases in fish—an early-diverging vertebrate lineage predating the bird-mammal split—would strongly support the idea that magnitude-related biases originated in a common vertebrate ancestor. However, detailed behavioral data, including full performance functions or evidence of numerical biases, are currently lacking for fish and other early-diverging vertebrates. Further controlled studies are needed to clarify the phylogenetic origins of these cognitive biases and their role in the evolution of cognition across species.

Neurobiological evidence suggests that similar numerical biases in distantly related species like primates and corvids may result from convergent evolution. Numerical cognition involves distinct pallial regions that evolved independently: the intraparietal and prefrontal cortices in macaques (12) and the nidopallium caudolaterale in crows (44). This anatomical divergence supports the idea that similar computational demands drove the independent emergence of analogous cognitive strategies. Thus, convergence remains a plausible explanation, highlighting the adaptive value of Bayesian-like inference in numerical cognition across vertebrates.

Research on cognitive biases in time estimation in macaques reveals striking parallels to humans, both exhibiting scalar variability and regression to the mean (19, 32). In magnitude estimation tasks involving time intervals, spatial distances, and vibrotactile frequencies, neurons in parietal and frontal cortices encode Bayesian estimates biased toward the center of the tested range (19–21, 45, 46). However, macaques struggle to integrate reward magnitude and prior reward likelihood in categorical decisions, tending to overadjust for reward magnitude but not for reward likelihood (47). This indicates that, although macaques share certain cognitive biases with humans in magnitude estimation, species-specific differences exist in how this information is processed and utilized.

Given that both species exhibit cognitive biases in numerical estimation, it is plausible that number neurons in the intraparietal sulcus and prefrontal cortex of monkeys, and in the nidopallium caudolaterale of crows, encode biased numerosity estimates. A key future direction is to investigate the neuronal correlates of these biases along processing pathways in both species.

Methods

Animals. We trained and acquired data from two male macaque monkeys (*M. mulatta*), age 9 and 12 y, and two male carrion crows (*C. corone*), age 6 and 11 y, from the institute's facility. Both monkeys were on a controlled water protocol and were trained to perform the behavioral task to receive fluid rewards. Crows were kept on a controlled feeding protocol during the experiments and earned food (birdseed pellets and mealworms) as a reward during training. For monkeys, dry food was always provided ad libitum, while for crows, water was provided ad libitum. If necessary, water or food was supplemented after the daily sessions.

All individuals were experienced with cognitive tasks from previous studies, for which monkeys had been implanted with titanium head posts allowing for head fixation and crows had been implanted with a small holder for attaching the reflector to enable head tracking (for details, see refs. 13 and 48). All procedures involving the animals were in alignment with German and European law. They have been evaluated by an independent ethics committee and were approved by the national animal experimentation authority, the Regierungspräsidium Tübingen.

Experimental Apparatus. During daily training and experimental sessions lasting 1 to 3 h, monkeys were head-restrained and seated in primate chairs within darkened chambers. They viewed stimuli on a 15-inch monitor with a 75 Hz refresh rate placed 57 cm in front of them. Eye movements were tracked continuously and sampled within trials using an infrared eye camera (ISCAN, Woburn, MA). Responses were registered using a custom-fitted touch-responsive bar in the primate chairs. Crows sat on a wooden perch placed inside a darkened chamber in front of a 15-inch touchscreen on which the stimuli were presented (3 M Microtouch; 60 Hz refresh rate). The crows' head position was tracked in 3D via a reflector foil mounted onto the crows' head by a two-camera system (FLIR CM3-U3-13Y3M) to ensure a central head position in front of the monitor at a viewing distance of ~14 cm. The presentation of stimuli, behavioral monitoring, and reward administering for both species were automated using the CORTEX program (National Institute of Mental Health).

Task Design. All subjects performed a delayed match-to-numerosity (DMN) task in which they had to memorize a sample numerosity and match it to the same numerosity displayed later to gain reward (Fig. 1*A*). The go-stimulus indicated that the subject may initiate a trial by fixating on the screen. After trial initiation no further fixation stimulus was displayed for the crows. For monkeys a fixation stimulus (a small white square) in the center of the screen was presented throughout the whole trial. Additionally, monkeys had to grab and hold the response bar. After a brief waiting time (300 ms), a sample stimulus displaying a specific number of black dot items (sample numerosity) on a gray background circle was presented for 500 ms (monkeys) or 600 ms (crows). The subjects were required to accurately assess the number of displayed items and memorize it during the following delay period (1,000 ms). In the subsequent test period, the subjects had

to respond only when a dot display with the same numerosity was shown. In 50% of the randomized trials, the test stimulus displayed the matching numerosity, prompting the subjects to respond (match trials). In the other 50% of the trials (nonmatch trials), the test stimulus showed a nonmatching numerosity. In this case, the subjects had to refrain from responding and wait 800 ms (monkeys) or 900 ms (crows) until the matching stimulus appeared, which was presented for the same duration. To respond, monkeys released a response bar, while crows moved their head or pecked on the screen. The chance level of correct responses was 50%. After 80 correct answers the subjects had to perform a second task independent from the DMN task analyzed here, in which subjects were rewarded for maintaining fixation only.

Instruction Stimuli. Sample stimuli for the numerosity range 1 to 10 were used. Sample numerosity 1 was tested against larger test stimuli with numerosity 2 and 3 only. Sample numerosities 2 to 9 were each tested against two smaller and two larger nonmatching test numerosities ranging from 1 to 32. Ratios (sample versus test) were fixed equally spaced on a logarithmic scale at 0.316, 0.56, 1.77, and 3.16, and rounded integer test stimuli were chosen accordingly. As a result, most sample numerosities were also presented as nonmatching test numerosities, albeit with varying frequencies of appearance. The cut-off of higher sample numerosities led to an unavoidable overrepresentation of nonmatch trials involving smaller numerosities (e.g., 2 and 3), which may result in greater familiarity and thus enhanced responses (also if wrong) at the low end of the numerosity scale. Because only integer numerosities are used, the actual ratios between match and nonmatch values are not really evenly distributed across the numerosity range. This may result in greater perceptual distances between stimuli at higher numerosities, potentially enhancing discriminability and improving performance (as visible for the monkeys in Fig. 3). Varying numbers of black dots were pseudorandomly placed within a gray background circle. Two types of conditions were implemented per species to control for covarying continuous features. In the standard condition for the monkeys, dot size varied pseudorandomly and for the crows it was held constant. In the control condition of both species, density and the cumulative total area occupied by all black dots were kept constant while dot sizes varied. Example Stimuli are shown in Fig. 1*B*. Each trial consisted of within-condition images, and trials were arranged in a pseudorandom order unknown to the experimenter. Error trials were repeated with a pseudorandom delay automized via CORTEX to prevent animals from avoiding hard trials. To prevent animals from memorizing individual stimuli, stimuli with randomly varying dot positions were newly generated before each session using prewritten MATLAB functions.

Data Analysis. Data analyses and simulations were carried out using MATLAB (Version R2023a, MathWorks Inc). Gaussian fits and linear regressions were done by minimizing the square of the residuals implemented by the Curve-fitting Toolbox. All values were calculated per session and reported are mean ± SEM across sessions per individual–unless stated otherwise. For all effects, we first tested for the existence of a correlation, by calculating Pearson correlations per individual. To quantify correlations, linear regressions were fitted per session and the resulting slopes tested against 0 (or 1 to test for the regression to the mean).

Scalar variability and regression to the mean. To construct response functions, the relative response frequency for different test numerosities was calculated per sample numerosity and session. We fitted Gaussian curves with free parameters amplitude, mean and variance to the response functions for sample numerosities 2-10. Fitting continuous functions, such as Gaussian curves, to discrete behavioral data has been a well-established method in numerical cognition research (9, 49–51), as it allows detailed characterization of response distributions across the entire response profile. The width was measured as the half-width of Gaussian fits on a logarithmic scale. To exclude unreasonable fits, only those fits with an r^2 value above 0.6 were included in the analysis. This amounted to >95% of response curves being included per individual. 648 (monkey #1), 504 (monkey #2), 162 (crow #1), and 151 (crow #2) fits were included (Figs. 4 and 5). To further validate the scalar variability and regression to the mean effect, we compared how well response functions on different scales with different means were described by Gaussian curves. This can be taken as a measure of symmetry of the subjects' response distributions. For each session, response functions were plotted on a linear or logarithmic scale. Gaussian curves with free parameters amplitude and variance were fitted to these. The center was either fixed at the sample numerosity n or for the logarithmic scale at $\log(n)$ or at the shifted sample numerosity $\log(5.5) - 0.79 \cdot (\log(5.5) - \log(n))$ according to a regression to the mean with slope 0.79. The mean goodness-of-fit (r^2) values were compared per individual.

Sequential effect. To test for sequential effects, single trial outcomes of all sessions were analyzed per individual. Only nonmatch trials of central sample numerosities 3–7 were included. To control which numerosity the subjects saw previously, only trials for which the corresponding previous trial was done correctly were included. Therefore, the sample numerosity of the respective past trial was always shown twice as sample and match. 7047 (monkey #1), 10901 (monkey #2), 1844 (crow #1), and 1678 (crow #2) trials were included to test for the effect of the previous sample (=match) numerosity S/M*1* (Fig. 6*B*). To further characterize the sequential effect, trials were split into two groups. The congruent group inhabited all trials for which the previous sample numerosity and the current test numerosity were both smaller or both larger than the current sample numerosity. The incongruent group consisted of all trials for which the previous sample numerosity and current test numerosity were opposing one another (one smaller, one larger) with respect to the current sample numerosity. To further characterize the effect, the change of the difference in error rates (congruent group – incongruent group) with respect to the sample numerosity of the previous trial S/M*1* was evaluated depending on the numerical distance or time between trials. Nonmatch trials of all sample numerosities were grouped according to the absolute numerical distance between the current and previous sample. Datapoints consisting of less than 10 trials were excluded, leading to fewer datapoints for higher numerical distances (Fig. 6 *C*–*F*) amounting to 76, 76, 76, 74, 70, 45 included datapoints for monkey #1, 57, 57, 57, 57, 57, 55 for monkey #2, 18, 18, 18, 17, 17, 11 for crow #1 and 17, 17, 17, 17, 16, 13 included datapoints for crow #2, reported per numerical distance 1-6, respectively. To test for a correlation between the sequential effect and the time passed between the current and previous sample onset, trials were grouped into 50% of shortest intervals (median = 5.4 s, 5.4 s, 4.8 s, 5.0 s for monkey #1, monkey #2, crow #1, crow #2 respectively) and 50% of the longest intervals (median = 6.2 s, 7.0 s, 7.0 s, 7.1 s for monkey #1, monkey #2, crow #1, crow #2 respectively). To test for a sequential effect of the ith (second to fourth) previous trial, only those trials for which the ith previous until previous trial were all done correctly, were included to ensure subjects perceived the respective sample numerosities (Fig. 6 *G*–*J*). This led to less included trials the further back we moved in task history. Again, datapoints consisting of less than 10 trials were excluded, leading to 76 (S/M*1*), 74 (S/M*2*), 70 (S/M*3*), and 53 (S/M*4*) included datapoints for monkey #1, 57, 57, 57, 57 for monkey #2, 18, 17, 17, 14 for crow #1 and 17, 17, 14, 11 for crow #2, respectively. Statistics were done across sessions.

Bayesian Observer Model. The basic model structure consists of a perceptual function f, that transforms a sample numerosity x_n into a conditional probability distribution $p(x_m|x_n)$ describing the probability of different numerical measurements x_m for the specific sample numerosity x_n (likelihood). It is modeled as a normal distribution on a logarithmic scale with its mean located at the corresponding sample numerosity $\mu_n = x_n$. The prior $p(x_n)$ is again modeled as a normal distribution on a logarithmic scale with its mean μ_P varying depending on the model. Justified by the Weber–Fechner law, we assumed that the likelihood and the prior both have constant variance on a logarithmic scale across numerosities. The posterior $p(x_n|x_m)$–the probability of different sample numerosities x_n given a measurement x_m–is formed by integrating the likelihood and the prior via Bayes' rule $p(x_n|x_m) = p(x_m|x_n) \cdot p(x_n)$, again resulting in a lognormal distribution. Its mean μ_{post} is described by a shift of the mean of the likelihood μ_n toward the mean of the prior μ_p:

$$\mu_{post} = (1-w) \cdot \mu_n + w \cdot \mu_p,$$

with constant $w \in (0, 1)$. We implemented four different versions of the prior. Eqs. [**1**] refers to model #0 [M0], [**2**] to model #1 [M1], [**3**] to model #2 [M2], and [**4**] to model #3 [M3]:

$$\mu_p = \mu_n \rightarrow \mu_{post} = \mu_n, \quad [\mathbf{1}]$$

$$\mu_p = 5.5, \quad [\mathbf{2}]$$

$$\mu_p = \mu_{post}^{-1}, \quad [\mathbf{3}]$$

$$\mu_p = 0.41 \cdot \mu_n^{-1} + 0.16 \cdot \mu_n^{-2} + 0.15 \cdot \mu_n^{-3} + 0.09 \cdot \mu_n^{-4} + 0.08 \cdot \mu_n^{-5} + 0.13 \cdot \mu_n^{-6} + 0.11 \cdot \mu_n^{-7} + -0.12 \cdot \mu_n^{-8}, \quad [4]$$

whereby μ^{-i} denotes the respective mean of the likelihood or posterior of the i-th last trial. As an approximation of the coefficients of the sample numerosities of the eight previous trials in μ_p of M3, a calculation of the sequential effect on the combined data of all subjects was used. Measurement uncertainty increases with greater task history depth due to fewer available data points. Fixing these parameters before individual fitting ensures comparability across models, as each retains a single free parameter. This approach enables a clearer conceptual comparison between prior types, even if individual fitting might capture behavior more precisely. Instead of an estimator function that projects the posterior onto a single value, the response function r discretizes the continuous posterior distribution $ln(X) \sim N\left(\mu_{post}, \sigma^2_{post}\right)$, where the probability of responding to a test numerosity t is given by

$$P(t) = \int_{ln(t-0.5)}^{ln(t+0.5)} N\left(ln(x); \mu_{post}, \sigma^2_{post}\right) dx.$$

It is normalized such that the probability of responding to a test numerosity equals 1 when test = prior = sample. Based on this probability distribution–which depends on sample, test, and previous sample numerosities–Boolean trial outcomes were simulated using MATLAB's pseudorandom number generator.

Models were fit per subject to minimize the summed squared residual error to the behavioral data of the relevant measure for the chosen effect. First σ^2_{post} of M0 was fit to the observed scalar variability (width of Gaussian fits of sample numerosities 2 to 10). The argmin of 10 consecutive simulations was chosen as the width of the posterior distribution: $\sigma^2_{post} = 0.2,\ 0.18,\ 0.18,\ 0.19$ for monkey #1, monkey #2, crow #1, and crow #2, respectively. It was kept fixed for M1–M3 to ensure comparability between models within a subject. For M1 as well as M3, the weight $w \in (0, 1)$ was chosen to minimize the summed squared residual error of the centers of Gaussian fits of sample numerosities 2 to 10. The argmin of 10 consecutive simulations resulted in $w = 0.23, 0.23, 0.23, 0.24$ for M1 and $w = 0.2, 0.2, 0.23, 0.24$ for M3, respectively per subject. For M2, the weight $w \in (0, 1)$ was chosen to minimize the summed squared residual error of the sequential effect with respect to S/M*1*. The argmin of 10 consecutive simulations resulted in $w = 0.1, 0.12, 0.1, 0.12$ for monkey #1, monkey #2, crow #1, and crow #2, respectively. Assuming an optimal estimation process, the width of the likelihood and prior can be derived from w and σ^2_{post} (31, 52).

To control for effects due to stimulus statistics or analysis methods, we simulated single trial outcomes of all trials for each subject according to models M0, M1, M2, and M3, calculated the biases as described above per session and tested whether the predicted slope of the scalar variability, regression to the mean and the difference in error rate with respect to S/M*1* differed. Per animal and effect a Friedman test for paired data and Bonferroni corrected post hoc pairwise comparisons between 5 groups–M0, M1, M2, M3, and the behavioral data–were performed. We report all significant results ($P < 0.01$) of post hoc pairwise comparisons between the simulated data and observed data, as well as all results comparing M3 to the behavioral data for the regression to the mean and the sequential effect. The task history and numerical distance dependency of the sequential effect were not calculated per session due to limited data and were only evaluated as fit to the data combined across sessions (Fig. 7 *B*–*E*).

Data, Materials, and Software Availability. All study data are included in the main text.

ACKNOWLEDGMENTS. We thank Maximilian E. Kirschhock for help with initial crow training and Michael Baumgärtner for careful reading of the manuscript. This research was supported by a PhD grant from the Studienstiftung des Deutschen Volkes to L.L.J., and DFG grants NI 618/13-1 (FOR 5159) and NI 618/15-1 to A.N.

1. H. von Helmholtz, "Concerning the perceptions in general, 1867" in *Readings in the History of Psychology* (*Century Psychology Series*, Appleton-Century-Crofts, 1948), pp. 214–230.
2. A. Tversky, D. Kahneman, Judgment under uncertainty: Heuristics and biases. *Science* **185**, 1124–1131 (1974).
3. F. P. De Lange, M. Heilbron, P. Kok, How do expectations shape perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).
4. I. Diester *et al.*, Internal world models in humans, animals, and AI. *Neuron* **112**, 2265–2268 (2024).
5. J. Spicer, J.-Q. Zhu, N. Chater, A. N. Sanborn, Perceptual and cognitive judgments show both anchoring and repulsion. *Psychol. Sci.* **33**, 1395–1407 (2022).
6. G. M. Cicchini, G. Anobile, E. Chelli, R. Arrighi, D. C. Burr, Uncertainty and prior assumptions, rather than innate logarithmic encoding, explain nonlinear number-to-space mapping. *Psychol. Sci.* **33**, 121–134 (2022).
7. A. Prat-Carrabin, M. Woodford, Efficient coding of numbers explains decision bias and noise. *Nat. Hum. Behav.* **6**, 1142–1152 (2022).
8. A. Nieder, The evolutionary history of brains for numbers. *Trends Cogn. Sci.* **25**, 608–621 (2021).
9. A. Nieder, E. K. Miller, Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* **37**, 149–157 (2003).
10. A. Nieder, Absolute numerosity discrimination as a case study in comparative vertebrate intelligence. *Front. Psychol.* **11**, 1843 (2020).
11. G. T. Fechner, *Elemente der psychophysik* (Breitkopf und Härtel, Leipzig, 1860).
12. K. Merten, A. Nieder, Compressed scaling of abstract numerosity representations in adult humans and monkeys. *J. Cogn. Neurosci.* **21**, 333–346 (2009).
13. H. M. Ditz, A. Nieder, Numerosity representations in crows obey the Weber-Fechner law. *Proc. R. Soc. B* **283**, 20160083 (2016).
14. S. Kumar, S. B. Hedges, A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
15. F. H. Petzschner, S. Glasauer, K. E. Stephan, A bayesian perspective on magnitude estimation. *Trends Cogn. Sci.* **19**, 285–293 (2015).
16. H. M. Ditz, A. Nieder, Neurons selective to the number of visual items in the corvid songbird endbrain. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7827–7832 (2015).
17. G. Anobile, G. M. Cicchini, D. C. Burr, Linear mapping of numbers onto space requires attention. *Cognition* **122**, 454–459 (2012).
18. G. M. Cicchini, G. Anobile, D. C. Burr, Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7867–7872 (2014).
19. M. Jazayeri, M. N. Shadlen, A neural mechanism for sensing and reproducing a time interval. *Curr. Biol.* **25**, 2599–2609 (2015).
20. H. Sohn, D. Narain, N. Meirhaeghe, M. Jazayeri, Bayesian computation through cortical latent dynamics. *Neuron* **103**, 934–947.e5 (2019).
21. L. Serrano-Fernández, M. Beirán, R. Romo, N. Parga, Representation of a perceptual bias in the prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2312831121 (2024).
22. J. Whalen, C. R. Gallistel, R. Gelman, Nonverbal counting in humans: The psychophysics of number representation. *Psychol. Sci.* **10**, 130–137 (1999).
23. S. Dehaene, V. Izard, E. Spelke, P. Pica, Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* **320**, 1217–1220 (2008).
24. K. McCrink, S. Dehaene, G. Dehaene-Lambertz, Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Percept. Psychophys.* **69**, 1324–1333 (2007).
25. C. Malapani, S. Fairhurst, Scalar timing in animals and humans. *Learn. Motiv.* **33**, 156–176 (2002).
26. H. Nover, C. H. Anderson, G. C. DeAngelis, A. Logarithmic, Scale-invariant representation of speed in Macaque middle temporal area accounts for speed discrimination performance. *J. Neurosci.* **25**, 10049–10060 (2005).
27. J. Z. Sun, G. I. Wang, V. K. Goyal, L. R. Varshney, A framework for Bayesian optimality of psychophysical laws. *J. Math. Psychol.* **56**, 495–501 (2012).
28. H. L. Hollingworth, The central tendency of judgment (1910). https://philpapers.org/rec/HOLTCT-2. Accessed 8 August 2025.
29. J. Jou, G. E. Leka, D. M. Rogers, Y. E. Matus, Contraction bias in memorial quantifying judgment: Does it come from a stable compressed memory representation or a dynamic adaptation process? *Am. J. Psychol.* **117**, 543–564 (2004).
30. P. Ashourian, Y. Loewenstein, Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One* **6**, e19551 (2011).
31. F. H. Petzschner, S. Glasauer, Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *J. Neurosci.* **31**, 17220–17229 (2011).
32. M. Jazayeri, M. N. Shadlen, Temporal context calibrates interval timing. *Nat. Neurosci.* **13**, 1020–1026 (2010).
33. R. C. Tomlinson, N. K. DeWind, E. M. Brannon, Number sense biases children's area judgments. *Cognition* **204**, 104352 (2020).
34. W. Schwarz, A. Ischebeck, Sequential effects in number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 1606–1621 (2000).
35. D. Gökaydin, P. Brugger, T. Loetscher, Sequential effects in SNARC. *Sci. Rep.* **8**, 10996 (2018).
36. G. M. Cicchini, K. Mikellidou, D. C. Burr, The functional role of serial dependence. *Proc. R. Soc. B* **285**, 20181722 (2018).
37. M. Manassi, Y. Murai, D. Whitney, Serial dependence in visual perception: A meta-analysis and review. *J. Vis.* **23**, 18 (2023).
38. J. Fischer, D. Whitney, Serial dependence in visual perception. *Nat. Neurosci.* **17**, 738–743 (2014).
39. T. P. Saarela, S. M. Niemi, M. Olkkonen, Independent short- and long-term dependencies in perception. *J. Vis.* **23**, 12 (2023).
40. L. Serrano-Fernández, M. Beirán, N. Parga, Emergent perceptual biases from state-space geometry in trained spiking recurrent neural networks. *Cell Rep.* **43**, 114412 (2024).
41. S. B. Hedges, The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**, 838–849 (2002).
42. N. S. Clayton, N. J. Emery, Avian models for human cognitive neuroscience: A proposal. *Neuron* **86**, 1330–1342 (2015).
43. A. Nieder, Evolution of cognitive and neural solutions enabling numerosity judgements: Lessons from primates and corvids. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, 20160514 (2018).
44. H. M. Ditz, A. Nieder, Sensory and working memory representations of small and large numerosities in the crow endbrain. *J. Neurosci.* **36**, 12044–12052 (2016).

45. C. Preuschhof, T. Schubert, A. Villringer, H. R. Heekeren, Prior information biases stimulus representations during vibrotactile decision making. *J. Cogn. Neurosci.* **22**, 875–887 (2010).
46. D. Benozzo, L. Ferrucci, A. Genovesio, Effects of contraction bias on the decision process in the macaque prefrontal cortex. *Cereb. Cortex* **33**, 2958–2968 (2023).
47. T. Teichert, V. P. Ferrera, Suboptimal integration of reward magnitude and prior reward likelihood in categorical decisions by monkeys. *Front. Neurosci.* **4**, 186 (2010).
48. P. Viswanathan, A. Nieder, Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11187–11192 (2013).
49. M. Piazza, V. Izard, P. Pinel, D. Le Bihan, S. Dehaene, Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* **44**, 547–555 (2004).
50. A. J. Kersey, J. F. Cantlon, Neural tuning to numerosity relates to perceptual tuning in 3-6-year-old children. *J. Neurosci.* **37**, 512–522 (2017).
51. E. F. Kutter *et al.*, Distinct neuronal representation of small and large numbers in the human medial temporal lobe. *Nat. Hum. Behav.* **7**, 1998–2007 (2023).
52. M. O. Ernst, M. S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).