PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Cognitive bridge between geometric and numerical learning in monkeys

Jessica F. Cantlon[a,1], Logan R. Brownell[a], Jialin Li[a], and Caroline M. DeLong[b]

Edited by Nora Newcombe, Temple University, Philadelphia, PA; received January 27, 2025; accepted July 16, 2025

Educational research highlights strong developmental links between numerical and spatial cognition in humans, often shaped by cultural tools like the number line. However, emerging evidence suggests that these number–space connections may reflect our evolutionary history and emerge even in the absence of cultural tools. We examined how task-naïve monkeys learn spatial and numerical concepts over time. Monkeys' learning trajectories revealed successful learning of both numerical and geometric concepts. Discriminability effects showed that abstract quantitative features were stronger predictors of accuracy on both tasks than low-level visual similarity, reflecting abstract processing of both tasks. Cross-lagged analyses revealed that learning in the geometry task significantly predicted future gains in numerosity performance, even after controlling session-level trends, within-task stability, and general performance factors. Together, these findings provide evidence that spatial and numerical learning in primates rely on mutual cognitive processes. Such representations could form the evolutionary foundation for the well-documented "cognitive bridge" between numerical and spatial reasoning in human development.

cognitive development | primate cognition | numerical cognition | spatial cognition | mathematical ability

Several theories propose shared evolutionary and developmental mechanisms underlying spatial and mathematical cognition in humans (1–4). Many suggest a fundamental relation between numerical and spatial processes that relates abstract representations of number to geometric features like length and size. However, there are few demonstrations of an evolutionary connection between numerical and spatial cognition in primates. Some evidence from bird species such as newborn chickens is suggestive of a primitive origin for the relation between numerical and spatial processes (5), but the homology between that behavior and human cognition is debated (6). Drucker and Brannon (6) showed that monkeys can track the numerical position of an object in a linear spatial array and that like many humans, they preferentially use a left-to-right orientation of the array rather than a right-to-left orientation. This shows that monkeys, like humans, associate numerical positions with spatial positions. But, while these prior studies investigating the relation between numerical and spatial cognition have focused largely on whether numerical values are mapped to spatial positions on a line, a broader base of cognitive and neural interactions between numerical and spatial representations might be the more critical component of our evolutionary history.

Broad links between numerical and spatial cognition are supported by work showing that numerical and spatial reasoning share representations, reasoning mechanisms, and neural resources. Classic work by Gibbon, Meck, and Church proposed that numerical and temporal magnitudes are processed through a shared representational system (7, 8) and subsequent work on numerosity systems (9) and other magnitudes (10) laid the groundwork for the idea that magnitudes across domains are encoded in a common analog format. In parallel, Holyoak and colleagues demonstrated widespread analog representations of magnitude in semantic memory (11, 12). Together, prior research demonstrates that a critical component of human cognition is the analog magnitude system, which encodes all kinds of quantities as continuous intensities along a normalized, one-dimensional scale. This system allows magnitudes from different formats and dimensions—such as number, size, or length—to be represented and compared within a common mental framework (13). The analog magnitude system is hypothesized to facilitate crosstalk between numerical and spatial domains (13–18). This shared representational system of numerical and spatial stimuli supports the possibility of cross-domain learning and transfer, where training in one domain (e.g., spatial reasoning) facilitates performance in another (e.g., numerical reasoning)

Significance

Developmental research has identified a "cognitive bridge" between spatial and numerical reasoning in early childhood, where spatial skills reliably predict later mathematical achievement. Yet the origins of this bridge remain unclear: Is it constructed through cultural tools, or does it emerge from more primitive quantitative intuitions? Formal concepts in geometry and mathematics, such as number, length, and area, may be grounded in primitive intuitive systems that support both spatial and numerical thinking. Here, we show that monkeys are sensitive to a shared set of abstract numerical and geometric features and, like human children, exhibit cross-domain transfer during spatial and numerical learning. These findings suggest that the bridge that supports spatial and numerical learning in humans is rooted in primate cognition.

Author affiliations: [a]Psychology Department, Carnegie Mellon University, Pittsburgh, PA 15213; and [b]Psychology Department, Rochester Institute of Technology, Rochester, NY 14623

Author contributions: J.F.C., L.R.B., and C.M.D. designed research; J.F.C., L.R.B., and C.M.D. performed research; J.F.C., L.R.B., and J.L. analyzed data; and J.F.C., L.R.B., and C.M.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jcantlon@andrew.cmu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2502101122/-/DCSupplemental.

Published August 18, 2025.

and vice versa (19)—but there are few demonstrations of this with any nonhuman animal.

Research with young children highlights a developmental scaffolding in which spatial reasoning provides a foundational framework for numerical development during early childhood (19). Evidence supporting this view shows that young learners often rely on spatial cues, such as surface area, perimeter, and spatial patterns, to infer numerical relationships (20–23). Similarly, Gunderson, Ramirez, Beilock, and Levine (24) demonstrated that children's spatial skill (mental rotation) at age 5 y predicted their performance on a symbolic approximate number task at age 8, suggesting that spatial reasoning lays a foundation for numerical understanding. Newcombe, Levine, and Mix (3) propose that spatial and numerical cognition share developmental pathways, with spatial reasoning acting as a "cognitive bridge" that facilitates abstract numerical representations. Lourenco and Longo (25) and de Hevia and Spelke (26) found evidence for early links between spatial magnitudes (e.g., size) and numerical magnitudes in infancy, supporting the idea that these domains are connected early in development. Generally, this developmental scaffolding hypothesis emphasizes a unidirectional influence, suggesting that transfer effects are predominantly from spatial to numerical domains, with little reciprocal effect.

Despite advancements in theories about the relations between number and space, most evidence on the origins of the relations has focused on associative links between numerical and spatial values, such as number and size or number and length (e.g., refs. 6, 25–28) and SNARC or Stroop effects (e.g., refs. 27–30). These are associations among *specific values* between domains. For example, the widely studied SNARC (spatial-numeric association in response code) effect highlights implicit associations between smaller numbers and spatial positions on the left and larger numbers on the right (27). Similarly, studies in infants demonstrate that increasing numbers of dots are associated with increasing lengths (26). These findings suggest that spatial and numerical systems are linked at a correlational or associative level, but there is limited insight into how numerical and spatial representations came into contact in the first place, and whether their interactions extend to shared internal properties, computations, or learning transfer. In contrast to prior studies, which examine associations between numerical and spatial stimuli, the current study tests whether similar computations arise across distinct judgments of numerosity and geometry. This question is especially informative in nonhuman primates, who receive no language input or training to link number and space—making any observed interactions evidence of evolutionarily conserved learning processes rather than human-constructed ones.

Interactions between numerical and geometric cognition in nonhuman primates are interesting for two additional reasons. First, some theories propose that unlike humans, nonhuman primates do not decompose shapes into abstract geometric components, such as "number of sides" or "angle," but instead process shapes holistically based on overall visual similarity (e.g., Sable-Meyer et al., (31)). On this view, geometric forms are human-constructed abstractions, and animals are not especially sensitive to their quantitative structure. Second, some theories argue that nonhuman animals solve problems within domain-specific systems but struggle to relate content across domains (32, 33). In this view, number and geometry are distinct core knowledge domains with separate representational currencies. This view is consistent with a long history of evidence that animals typically generalize based on concrete perceptual similarity, not on deeper structural or conceptual commonalities across domains (34–37). Apart from the prior number–space study described above (6), there is little experimental evidence that monkeys transfer information across numerical and geometric domains (e.g., ref. 34).

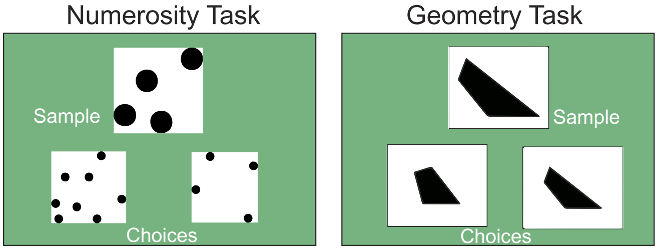

Fig. 1. Task figure. During the match-to-sample task, a sample stimulus appears, followed by two choice options. When subjects touch the sample stimulus, they receive auditory (bing) and visual (border flash) feedback. Correct choices result in positive auditory and visual feedback (bing and border flash) along with a cereal pellet reward, and incorrect responses trigger negative auditory feedback (bonk) and a 4-s timeout with a black screen. A 2-s intertrial interval separates consecutive trials.

In this study, task-naïve primates were trained on separate numerical and geometric tasks, with no explicit requirement to relate the two. Although they are learning to represent numerical and geometric information during the task, any cross-task influence must arise spontaneously from shared processing or representations. We test whether the numerical and geometric tasks rely on common low-level perceptual features or higher-order quantitative abstractions, whether performance in one domain predicts performance in the other, and whether longitudinal learning transfers across domains. This test of the continuity versus construction hypotheses has implications for understanding the origins of human mathematical cognition because if numerical and spatial reasoning interact in nonhuman primates, then the bridge between domains likely precedes and scaffolds human learning rather than being the product of it.

Methods

Participants. Four task-naïve olive baboons (Kalamata, Olivella, Pico, and Sabina; *Papio anubis*; 2 female, 2 male; adults ages 12 to 28 y) at the Seneca Park Zoo in Rochester, New York, participated in this study. They are socially housed with a troop of ten baboons in an indoor-outdoor environment. They participate in computerized touchscreen tasks on an exhibit called the Primate Portal, which is a public STEM outreach exhibit. The 1,920 × 1,080 resolution computerized touchscreen is permanently installed in the outdoor section of their enclosure. The animals participate in touchscreen tasks voluntarily, ad libitum, and are under no dietary restrictions. Subjects complete a variable number of trials each day and work at their own pace. All subjects had prior touchscreen experience with identity matching tasks and object matching tasks, as well as nontouchscreen tasks with real objects (38–41). However, they had no prior experience with explicit numerosity or geometry touchscreen tasks.

Task. Task-naïve baboons were shaped to use a touch-sensitive monitor during the summer of 2020. Shaping involved learning to touch a fluttering butterfly to receive a cereal pellet from an automated MedLine feeder. Once most animals could successfully touch the butterfly to receive a reward, the troop was transitioned from the shaping task to a match-to-sample task with photographs of objects and scenes. Once animals reached an accuracy level above 75% on photograph matching, they began the current match-to-sample task.

During the match-to-sample task (Fig. 1), subjects touch a start-stimulus, a white square in the middle of a green screen, to initiate each trial, then a sample stimulus appears. When subjects touch the sample stimulus, they receive auditory (bing) and visual (border flash) feedback. After a 500 ms delay, two choice options appear on the screen and subjects received positive auditory and visual feedback (bing and border flash) plus a cereal pellet for correct responses, and they receive negative feedback (bonk) and a 4 s timeout showing a black screen for incorrect responses. Each trial was followed by a 2-s intertrial interval.

Stimuli. Images were 500 × 500 pixels and displayed monochrome numerosity dot arrays (Numerosity Task) and 2D geometric shapes (Geometry Task). All stimuli are shown in *SI Appendix*, Fig. S1.

Numerosity stimuli ranged in value from 1 to 18 items, 131 to 5,652 pix^2 surface area, 0.004-0.2 × 10^{-3} density, 6-55 pixel dot radius–and were presented in numerical ratios of 0.25, 0.5, and 0.75 in terms of the smaller/larger ratio. The Sample stimuli were not identical matches to either Choice stimulus. Instead, the parameters of the Sample and Choice stimuli varied such that the numerical match was not always the best match in surface area, density, or perimeter.

Geometry stimuli ranged from triangles (3 sides) to decagons (10 sides)–half were regular and half irregular shapes. The average surface area was 939 pix^2 (range: 412 to 2,234 pix^2) and the average perimeter was 252 pixels (range: 126 to 438 pixels). The area of the Sample was 1.25 times larger or smaller than the Choice stimuli, such that neither choice image was identical to the Sample. This isolated the geometry of the shape from the pixelwise composition of the images.

A control task was included to measure general matching ability, independent of quantitative reasoning. In this task, subjects performed identity matching using a fixed stimulus pair: a red-and-yellow frog and a gray-and-green turtle. Because the same images appeared on every trial, success required sustained attention and trial-by-trial encoding of the sample stimulus in order to select the correct match.

Procedure. Animals were tested generally 4 d per week from 10am to 4 pm, weather permitting. The Numerosity Task, Geometry Task, and Control Task were presented in blocks of approximately 200 trials in a rotation that cycled continuously throughout the day. Animals worked on the task freely and at their own pace, taking turns, and thus completed variable numbers of trials. Each session represents the total trials an animal completed in a day. Only sessions with at least 10 trials were included in analyses. The average number of sessions completed was 262 per animal (Kalamata = 316, Olivella = 257, Pico = 243, Sabina = 232), and the average number of trials completed was 72,079 (Kalamata = 176,032, Olivella = 58,293, Pico = 35,796, Sabina = 18,195). Animals were tested from 8-9-2022 to 11-13-2024.

Analyses. All data processing, analyses, and visualizations were performed in R using libraries **dplyr**, **tidyr**, **ggplot2**, and **lme4**.

Longitudinal learning model. To characterize how monkeys learned each task over time, we used generalized additive mixed models (GAMMs; 41) to model session-level accuracy as a function of task and session number. GAMMs provide a flexible framework for modeling nonlinear change while accounting for repeated measures and subject-level variability. Unlike traditional linear regressions, GAMMs use smooth functions to capture complex, nonmonotonic patterns of change, such as plateaus, inflection points, or periods of rapid learning–without requiring strong assumptions about the shape of the trajectory. This approach is well suited for learning data, where changes in performance often follow nonlinear paths.

We implemented a GAMM using the **mgcv** package in R, predicting trial accuracy from a smooth of Session Number separately for each task (42). The model included random intercepts for subjects to account for individual differences in baseline performance. The model formula was

$$\text{Accuracy} \sim \text{TaskType} + s(\text{SessionNumber},\ \text{by} = \text{TaskType},\ k = 10) + s(\text{Sub},\ \text{bs} = ``\text{re}'').$$

This model includes 1) a parametric term for TaskType to model baseline differences across tasks, 2) task-specific smooths over session number [s(SessionNumber, by = TaskType)] with a basis dimension of 10 to allow moderate flexibility in the shape of each trajectory, and 3) a random intercept for each subject [s(Sub, bs = "re")] to account for between-subject variability. The model was fit using restricted maximum likelihood (REML), which provides stable estimation of smoothness parameters. Accuracy values were averaged within subject, task, and session, and analyses were restricted to sessions where monkeys completed all three tasks (Numerosity, Geometry, and Control) with at least 10 valid trials in each. After listwise deletion of incomplete sessions, the final dataset for this analysis included 309 sessions (M = 78 per monkey, SD = 50). To visualize learning over time, we plotted smoothed accuracy curves for each task.

We also quantified the amount of learning using the area under the curve (AUC) of the first derivative of each smooth term. This measure represents the local rate of accuracy change per session, and integrating it across the full training period yields a cumulative learning index. Derivatives were computed using the derivatives() function from the **gratia** package (Simpson, (43)) for each task-specific smooth term (e.g., s(SessionNumber):TaskTypeNUM). The resulting curves were then grouped by task, and AUCs were calculated using the trapezoidal rule, summing the product of session intervals and the average of adjacent derivative values. To test whether tasks differed in their overall amount of learning, we performed a permutation test (10,000 iterations) comparing pairwise AUC differences between tasks. On each iteration, we randomly shuffled task labels across sessions, recalculated the AUCs, and computed all pairwise differences. Empirical *P*-values were calculated as the proportion of permuted differences greater than or equal to the observed difference in absolute value. This approach controls for any distributional assumptions and provides a robust test of whether task learning trajectories differed in cumulative learning over time.

Distance effects. To quantify stimulus similarity, we computed four distance metrics: Pixelwise Distance, Fourier Distance, Numerosity Distance, and Surface Area Distance. Each metric captures a distinct aspect of the stimuli.

Pixelwise Distance: Pixelwise distance was calculated as the mean squared error (MSE) between the grayscale pixel intensity matrices of two images. This metric captures low-level visual differences across the entire stimulus.

Fourier Distance: Fourier distance measured shape similarity based on stimulus contours, reflecting mid-level perceptual features. Contours were extracted from binarized stimuli, converted into complex coordinates, and analyzed using the Fourier transform. Descriptors were normalized by the DC component and truncated to the first 20 harmonics, preserving global shape while discarding fine noise. The final distance was the Euclidean distance between descriptor vectors.

Numerosity Distance: Numerosity distance was defined as the absolute difference in quantity between stimulus pairs: the number of dots (in the Numerosity Task) or the number of sides (in the Geometry Task). This metric isolates abstract numerical structure independent of spatial layout or low-level appearance.

Surface Area Distance: Surface area distance quantified the difference in total stimulus area (in pixels) across dot arrays or geometric shapes. For dot arrays, this reflects cumulative area of all dots; for geometric shapes, it reflects the filled area within the shape contour. This metric captures size-related cues that are distinct from numerosity.

To investigate how stimulus features influenced task performance, for each trial, distance for each feature (pixelwise similarity, Fourier similarity, numerosity, and surface area) was calculated for each choice relative to the sample using the formula:

$$\text{Distance} = |\text{Incorrect} - \text{Sample}| - |\text{Correct} - \text{Sample}|.$$

More negative distances indicated "harder" trials, where the incorrect choice was closer to the sample than the correct choice, increasing task difficulty.

To investigate how different visual and quantitative features influenced trial-by-trial accuracy (binary outcome: correct or incorrect), we used a mixed-effects logistic regression model. Predictors included low-level visual similarity (pixelwise similarity), mid-level visual similarity (Fourier similarity), and abstract quantitative features (numerical and surface area differences). Analyses were conducted using the lme4 package in R. The model was specified as follows:

$$\text{Accuracy} \sim \text{Pixel} + \text{Area} + \text{Fourier} + \text{Number} + \text{SessionNumber} + (1 \mid \text{Sub}).$$

All predictors were z-scored within task to allow direct comparison of effect sizes across predictors and to account for differences in scale. Pixelwise, Fourier, Numerosity, and Surface Area represent the distance between incorrect and correct choices relative to the sample. SessionNumber accounts for longitudinal trends in performance over time. (1|Sub) represents random intercepts for subjects to account for individual variability.

Model significance was assessed using Wald z-tests on fixed effects, and coefficients (β) are interpreted in terms of log-odds. A positive β indicates that greater feature distance increases the likelihood of a correct response.

Cross-lag longitudinal models. To evaluate whether learning on one task predicted future performance in another, we implemented a cross-lagged extension of the GAMM we described earlier to characterize monkeys' longitudinal learning curves (42). We incorporated lagged predictors to assess the temporal directionality of cross-task interaction while adjusting for within-task stability and general learning trends.

For each session and subject, we computed mean accuracy separately for the Numerosity, Geometry, and Control tasks. We then calculated lagged accuracy scores (t_0) to reflect prior-session performance and used these to predict current-session accuracy (t_1). Two GAMMs were fit using the **mgcv** package in R. The first model predicted Numerosity performance at t_1 as a function of lagged performance on Numerosity, Geometry, and Control tasks at t_0; the second model reversed this structure to predict Geometry performance. Both models included smooth terms for Session Number (to capture nonlinear learning trajectories) and random intercepts for each monkey to account for individual differences:

$$\begin{aligned}\text{Accuracy\_NUM} \sim\ & s(\text{SessionNumber},\ k = 10) + s(\text{Sub},\ bs = ''\text{re}'') \\ & + \text{Lag\_NUM} + \text{Lag\_GEO} + \text{Lag\_CTRL}.\end{aligned}$$

And reciprocally:

$$\begin{aligned}\text{Accuracy\_Geometry} \sim\ & s(\text{SessionNumber},\ k = 10) + s(\text{Sub},\ bs = ''\text{re}'') \\ & + \text{Lag\_Geometry} + \text{Lag\_Numerosity} + \text{Lag\_Control}.\end{aligned}$$

This modeling approach offers several advantages over traditional linear cross-lagged panel models (42). GAMMs accommodate nonstationary and nonlinear learning trends, address repeated-measures dependencies through smooth and random effects, and allow simultaneous inclusion of multiple lagged predictors to isolate domain-specific transfer effects from domain-general fluctuations or within-task consistency.

Coefficient estimates for each lagged predictor reflect their unique contribution to predicting future performance, above and beyond general learning and task stability. Positive coefficients indicate that higher accuracy in the predictor task at t_0 is associated with higher accuracy in the outcome task at t_0. Statistical significance was assessed via Wald tests on the fixed-effect coefficients (e.g., Lag_Geometry → Numerosity). To visualize and interpret the effects of each lagged predictor, we plotted all session-level data points along with partial effect plots, allowing us to observe the cross-task relationships.

We evaluated the adequacy of our GAMMs using standard diagnostics provided by the **mgcv** package. Residual plots, Q–Q plots, and histograms of residuals revealed no severe violations of model assumptions. The low lag-1 autocorrelation (ACF $\approx$ 0.13 to 0.15) suggested minimal temporal dependence after accounting for predictors. We tested for multicollinearity using Variance Inflation Factors (VIFs), which were all well below conventional thresholds (VIFs < 2.6). We also verified that the chosen basis dimension ($k = 10$ to 25) was sufficient by examining the k-index and associated *P*-values. Model convergence was achieved in all cases, and the effective degrees of freedom plateaued at edf $\approx$ 7.5. Overall, the diagnostics indicate that our GAMMs were well specified and robust.

To further assess the robustness of the cross-domain transfer effects observed in the GAMMs, we conducted two follow-up analyses. First, we fit an alternative GAMM that incorporated an autoregressive [AR(1)] correlation structure at the subject level using the corAR1 function from the **nlme** package in R. This model explicitly accounted for potential serial autocorrelation in trial-level residuals across sessions, which could bias estimates of lagged effects. The model structure was otherwise identical to the original GAMM, including smooth terms for session number, random intercepts for subjects, and lagged accuracy predictors for Numerosity, Geometry, and Control tasks. Second, we implemented a linear mixed-effects model using the **lme4** package to provide a more conventional baseline for comparison. This model used the same cross-lagged predictor structure, with lagged accuracy from all three tasks at time t_0 predicting current accuracy at time t_1, and included Session Number as a fixed covariate and random intercepts for each subject. Model significance was assessed using Wald tests.

Results

To investigate the evolutionary foundations of quantitative reasoning, we examined how monkeys learn and generalize geometric and numerical concepts over time. Spatial reasoning is strongly linked to early mathematical development in humans, but the origins of this connection remain unclear. We asked whether nonhuman primates exhibit similar transfer between spatial and numerical cognition, and whether their learning depends on abstract features such as number and area rather than lower-level visual similarities. To answer these questions, we modeled monkeys' learning trajectories across three tasks: Geometry, Numerosity, and a Control task. We then tested whether performance in one domain predicted future improvement in the other, controlling for session-level learning trends and individual variation. Finally, we used feature-level regression analyses to identify which stimulus dimensions best explained monkeys' trial-by-trial accuracy. Together, these analyses reveal that monkeys not only acquire abstract geometric and numerical representations but also transfer learning across domains.

Longitudinal Performance Across Tasks. Monkeys successfully learned all three tasks over time (Fig. 2). To model learning trajectories, we used a GAMM to predict session-level accuracy from task and session number, allowing for nonlinear, task-specific learning curves and random intercepts for individual monkeys. This approach provides a more realistic characterization of learning than a linear model because it does not assume that learning progresses at a constant rate over time. Instead, it captures potentially nonlinear improvements, plateaus, or periods of rapid change, which are common in cognitive learning trajectories.

The GAMM revealed significant learning-related changes in accuracy for all three tasks (Fig. 3*A*; all smooth terms: $P < 0.001$). The geometry task showed a steep and consistent improvement over time (edf = 4.53, F = 63.04), the numerosity task showed a slightly more complex upward trend (edf = 5.85, F = 12.04), and the control task exhibited a rapid initial increase followed by a plateau with wave-like fluctuations across sessions (edf = 8.13, F = 11.13). Despite their differences, all three tasks showed evidence of learning, as monkeys began around chance and improved with experience. Although all tasks required learning, the relative stability of the Control task provides a baseline to control for general fluctuations in motivation and attention independently of learning.

To measure the overall amount of learning for each task (Fig. 3*B*), we calculated the AUC of the first derivative of each learning trajectory estimated by the GAMM. This metric reflects the cumulative change in accuracy over time, providing a single measure of how much learning occurred. Because the GAMMs nonlinear trajectories, the derivative captures local rates of improvement at each session, and the AUC integrates these values across the entire training period. The results revealed the greatest total learning for the geometry task (AUC = 0.27), followed by the control task (AUC = 0.22) and the numerosity task (AUC = 0.19). Although the geometry task had the highest learning amount, permutation tests comparing pairwise AUC differences across tasks showed no statistically significant differences in total learning (all ps > 0.29). Thus, monkeys exhibited robust learning across all three tasks and comparable cumulative improvement.

Shared Computations Across Tasks. To investigate which specific stimulus features underlie performance in the Numerosity and Geometry tasks, we first conducted logistic regression analyses to test the influence of visual and quantitative features on trial-level accuracy. We examined animals' sensitivity at three levels: low-level vision, mid-level vision, and abstract quantitative processing. Low-level vision was indexed by sensitivity to pixelwise similarity, which captures fine-grained overlap in luminance patterns across images. Mid-level vision was indexed by Fourier similarity, which reflects spatial frequency content and contour organization, properties

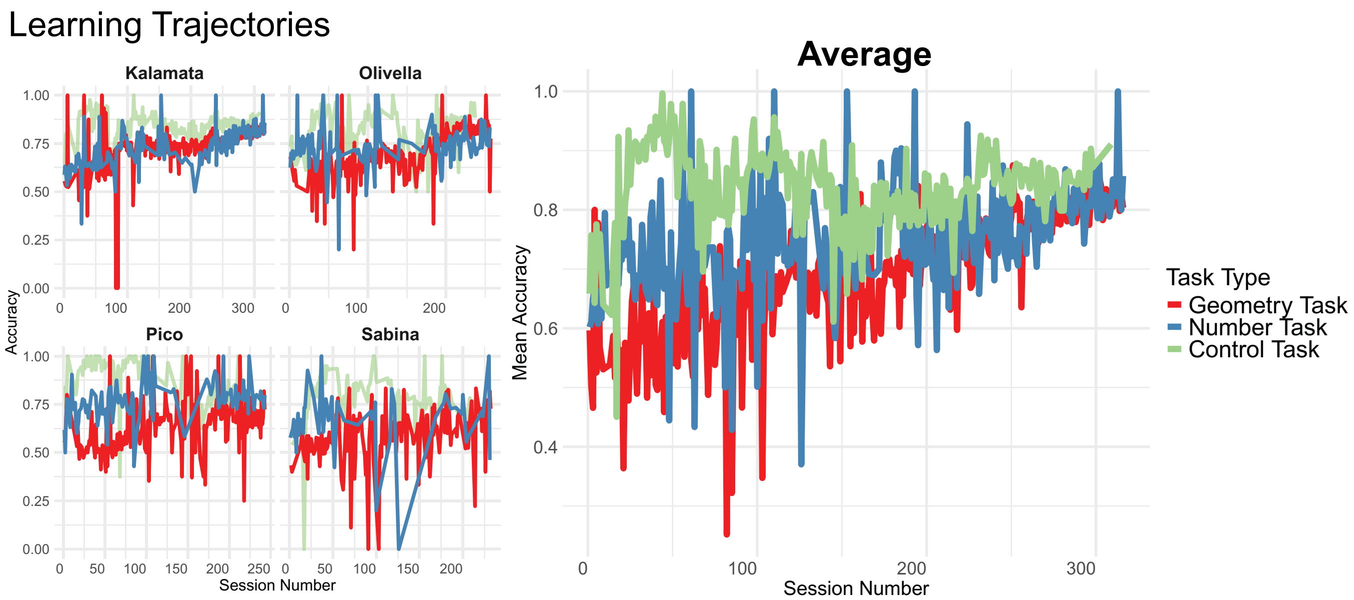

Fig. 2. Accuracy over time for the geometry task, numerosity task, and control task. Accuracy for each session, task, and subject (*Left* panel: Kalamata, Olivella, Pico, Sabina) as well as the aggregate across subjects (*Right* panel).

associated with global shape perception in intermediate visual areas. Abstract quantitative processing was indexed by sensitivity to numerical similarity (number of sides for Geometry and number of elements for Numerosity) and cumulative surface area (between shapes for Geometry and between arrays for Numerosity), reflecting higher-level representations of quantities that are abstracted and compared across objects. They are categorized as *abstract* features because they cannot be computed by image similarity processes.

A mixed-effects logistic regression revealed significant effects of all predictors on trial-by-trial accuracy. Abstract quantitative features Area similarity ($\beta = 0.22$, $P < 0.001$) and Numerical similarity ($\beta = 0.16$, $P < 0.001$) had the largest effects (Fig. 4), followed by lower-level visual features: Fourier similarity ($\beta = 0.07$, $P < 0.001$) and Pixel similarity ($\beta = 0.07$, $P < 0.001$). Task type had a significant effect ($\beta = 0.09$, $P < 0.001$), indicating modest differences in overall accuracy between the numerosity and geometry tasks. Performance also improved over time (Session number: $\beta = 0.16$, $P < 0.001$), consistent with learning. A small random effect for Subject (variance = 0.026) indicates low between-subject variability in performance across sessions. Overall, Area and Number were significantly greater predictors of monkeys' accuracy than Fourier and Pixel level predictors.

These predictors had consistent effects within each task, as shown in Fig. 5. Notably, for both tasks, Number and Area similarity demonstrated steep and positive trends, consistent with their strong regression coefficients. In contrast, Pixel similarity and Fourier similarity exhibited smaller but significant effects, reflecting the lesser contribution of low- and mid-level visual processing. These common signatures across tasks support the conclusion that number and geometry interact through shared, abstract representations of quantity, not just superficial visual features.

Predictive Relation Between Numerosity and Geometry. To test whether learning in one domain (numerical or spatial) predicted subsequent improvements in the other, we extended the GAMM approach used in our earlier analysis of learning trajectories. In that analysis, we used GAMMs to model how accuracy changed nonlinearly over time within each task. Here, we apply GAMMs in a cross-lagged framework to examine whether prior performance in one domain predicted future performance in another. Specifically, we modeled each task's performance at time t_1 as a function of lagged performance in all three tasks at t_0, while accounting for individual differences (random intercepts) and nonlinear learning over sessions (smooth terms). GAMMs

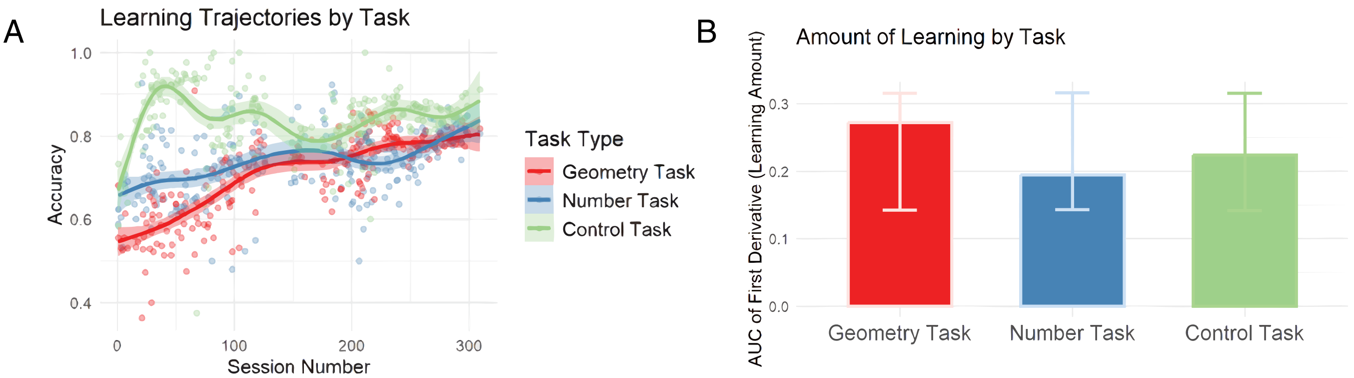

Fig. 3. (*A*) Learning trends for each task. Generalized additive mixed models (GAMMs) revealed improvements in session-level accuracy for all three tasks: Geometry, Numerosity, and Control. Learning trajectories differed across tasks in shape and rate, showing distinct plateaus, accelerations, and inflection points over time. (*B*) Amount of learning per task. Total learning was quantified as the area under the curve (AUC) of the first derivative of each learning trajectory. Error bars represent 95% CI from permutation tests. Differences in AUC between tasks were not statistically significant, indicating comparable overall learning across domains.

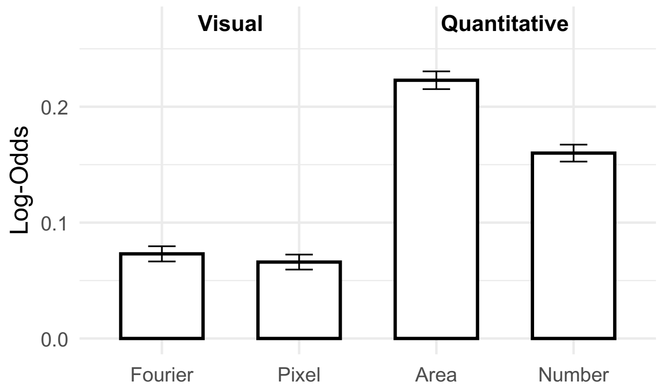

Fig. 4. Predictive strength of visual and quantitative features on accuracy. Standardized log-odds estimates from a mixed-effects logistic regression predicting trial-by-trial accuracy. Predictors include low- and mid-level visual features (Fourier and Pixel similarity) and abstract quantitative features (Area and Number similarity). The y-axis indicates the magnitude of each predictor's effect on accuracy, expressed in log-odds units; error bars reflect SE. Quantitative features (Area and Number) showed the strongest effects, suggesting that monkeys relied more on abstract representations of quantity than on low-level visual similarity when making their choices.

are particularly well-suited for this type of longitudinal analysis because they flexibly capture nonstationary learning dynamics that are common in learning curves.

Crucially, for a cross-domain effect to emerge as significant in this framework, it must explain unique variance in future performance above and beyond several layers of control: 1) general learning trends over time, modeled with smooth functions of session number; 2) within-task performance (e.g., predicting numerosity from earlier numerosity); and 3) domain-general factors captured by performance on the independent control task. This means that any observed cross-domain predictive effect reflects domain-specific transfer that cannot be explained by general learning, session structure, or generic task engagement.

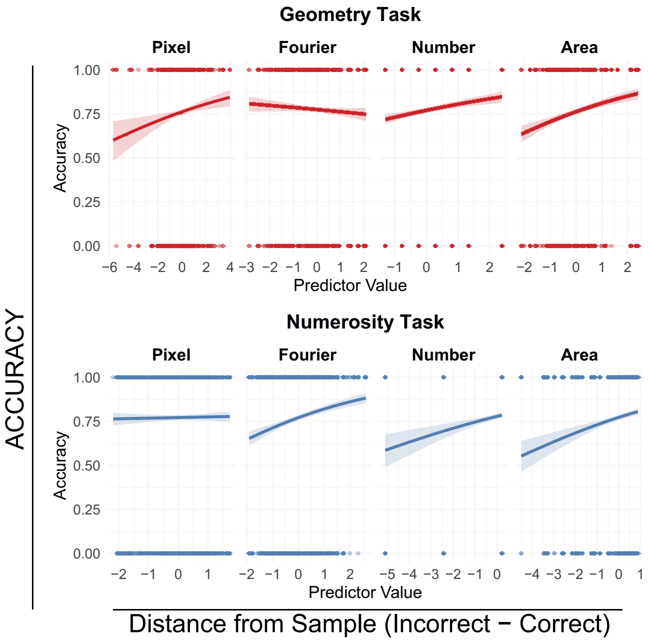

Fig. 5. Distance effects for features on accuracy by task. Logistic regression curves showing how accuracy varies with different feature predictors for each task. Each panel displays accuracy (y-axis) as a function of a single standardized predictor value (x-axis): Pixel similarity, Fourier similarity, Number, or Area. Results are shown separately for the Numerosity task (blue) and Geometry task (red). Curves reflect logistic fits with 95% confidence bands. In both tasks, accuracy increases most steeply with differences in Number and Area, indicating that abstract quantitative features were stronger predictors of performance than holistic visual similarity (Pixel, Fourier).

We fit two models (Fig. 6): one predicting numerosity performance at time t_1 (Fig. 6*A*) and one predicting geometry performance at time t_1 (Fig. 6*B*). Each included all three task performances at time t_0 as predictors, along with smooth terms for session number and random effects of subject. The model predicting numerosity performance showed significant within-task stability (Lag_Numerosity → Numerosity: $\beta = 0.26$, $P < 0.001$) and a significant cross-domain transfer from prior geometry performance (Lag_Geometry → Numerosity: $\beta = 0.15$, $P < 0.05$). Cross-domain performance on the geometry task accounted for over half the predictive strength of prior numerosity performance, highlighting a meaningful transfer of learning from spatial to numerical reasoning. The control task did not significantly predict numerosity outcomes (Lag_Control → Numerosity: $P = 0.78$). Thus, geometry performance at one time point uniquely predicted subsequent improvements in numerical reasoning.

The reciprocal model, predicting geometry performance, showed strong within-task consistency (Lag_Geometry → Geometry: $\beta = 0.24$, $P < 0.001$), but no significant cross-domain effect from numerosity (Lag_Numerosity → Geometry: $\beta = 0.07$, $P = 0.14$) or control task (Lag_Control → Geometry: $P = 0.97$). This asymmetry suggests that while spatial performance contributes to numerical improvement, the reverse effect, from numerical to spatial learning, is not as robust.

Both models explained a substantial proportion of variance in trial-level accuracy. The numerosity model explained 47% of deviance (adjusted $R^2 = 0.44$), while the geometry model explained 76% (adjusted $R^2 = 0.75$). The smooth terms for Session Number were highly significant ($ps < 0.001$) and captured the learning trends over time. These high levels of explanatory power indicate that the models captured the majority of systematic variance in learning across sessions, including general learning trends over time, stable within-task performance, and the predictive relationship between geometry and numerosity.

A null model simulation (*SI Appendix*, Fig. S2) showed that the lagged effect of geometry on future numerosity performance was not a spurious effect. Moreover, the cross-domain effect from geometry to numerosity remained significant even when explicitly modeling autocorrelation in trial-level performance. In a GAMM that included an AR(1) correlation structure to account for temporal dependencies within each subject, geometry performance at time t_0 continued to significantly predict numerosity accuracy at time t_1 ($\beta = 0.15$, $P < 0.05$). This further confirms that the observed transfer effect from geometry to numerosity is robust.

Finally, to benchmark these effects against a more conventional framework, we also fit linear mixed-effects models using the same cross-lagged structure. These models included lagged accuracy from all three tasks at time t_0, Session Number as a covariate, and random intercepts for each subject. The mixed-effects model predicting numerosity performance showed significant within-task stability ($\beta = 0.32$, $P < 0.001$) and a similar cross-domain effect from geometry ($\beta = 0.12$, $P < 0.05$) on numerosity performance, consistent with the GAMM results.

Together, these results provide robust evidence for cross-domain learning in the Numerosity and Geometry tasks: improvements in one domain predict future gains in the other, even when controlling for prior performance, general temporal trends, an independent task, and domain-general factors of motivation or task engagement. Combined with our feature-level analyses—which showed that abstract quantitative features (e.g., number and area)

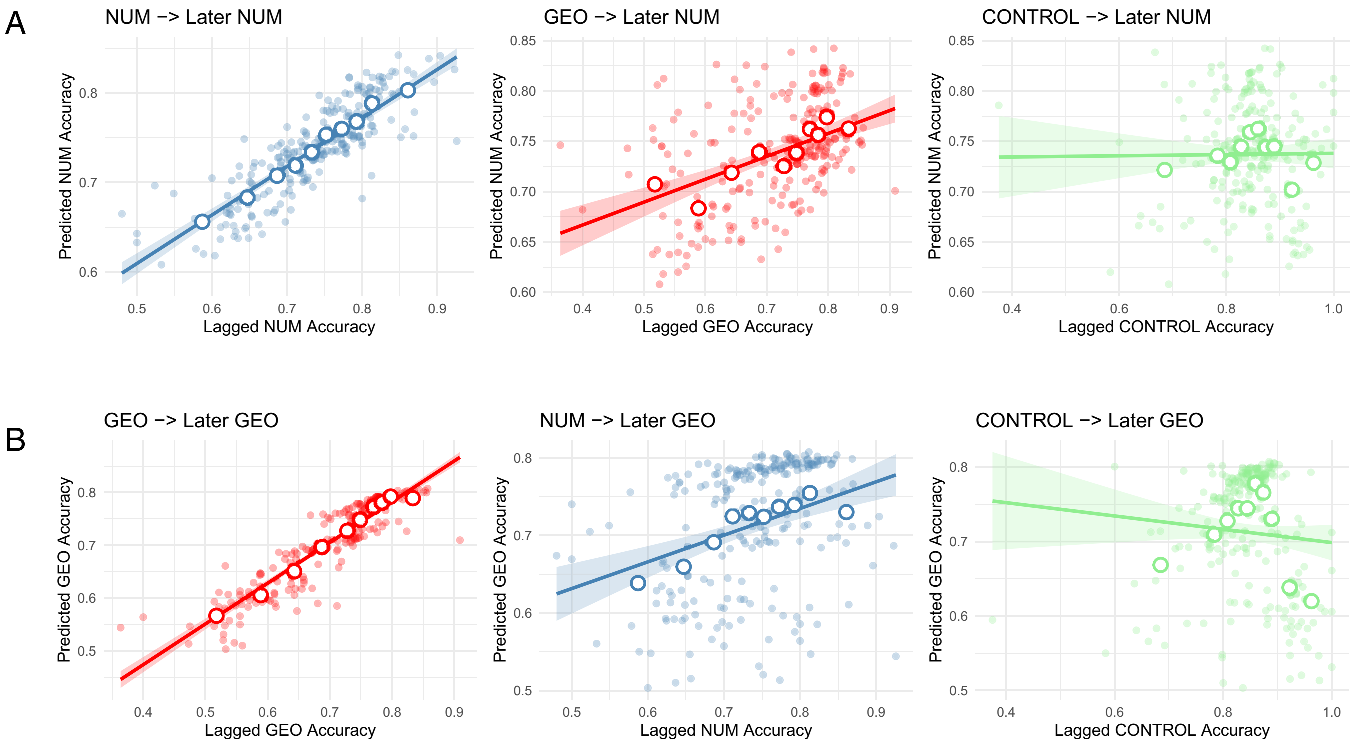

Fig. 6. Cross-lagged predictors of numerical performance. Each panel shows the relationship between lagged accuracy in a specific task at time t_0 (x-axis) and predicted accuracy at time t_1 (y-axis), based on a GAMM. Dots represent individual sessions, and trend lines show linear fits. White-centered circles show mean values. Data are controlled for general learning trends across sessions (modeled with smooth terms for Session Number) and individual differences (random effects for subjects). (*A*) Predictors of Numerosity performance at t_1, based on: (*Left*) prior Numerosity accuracy (Lag_NUM), (*Center*) prior Geometry accuracy (Lag_GEO), and (*Right*) prior Control task accuracy (Lag_CTRL). (*B*) Predictors of Geometry performance at t_1, based on: (*Left*) prior Geometry accuracy, (*Center*) prior Numerosity accuracy, and (*Right*) prior Control task accuracy.

were stronger predictors of performance than holistic image features (e.g., pixel and Fourier similarity)—these findings suggest that monkeys' learning and transfer were driven by shared structural representations of abstract quantity, and not just general cognitive abilities.

Discussion

Geometric and numerical abstraction were learned equally well by task-naïve nonhuman primates. Monkeys used a common set of abstract representations to solve the numerical and geometric tasks. By analyzing longitudinal learning data, we found that monkeys exhibit cross-domain transfer during numerical and geometric learning, specifically from geometric to numerical performance. The emergence of these effects in nonhuman primates, who lack symbolic language and formal instruction, supports the continuity hypothesis that number–space interactions are fundamental, and not uniquely constructed through human culture (1–4).

Monkeys learned the Geometry task and performed it as accurately as the Numerosity task, demonstrating robust learning, and they also showed signatures of representing geometric structure. Recent theories propose that only humans represent geometric shapes in terms of structural features—like number of sides, angles, and symmetries—while nonhuman animals are limited to holistic, contour-based representations (Sable-Meyer et al., 2022). However, our findings suggest a more nuanced view. While monkeys may not represent all of the formal geometric properties used by humans, they appear sensitive to at least a subset of abstract quantitative features that define geometric forms. "Number of sides" in the Geometry task and "number of elements" in the Numerosity task were the strongest predictors of performance across primates, along with cumulative surface area (for arrays) and shape area (for geometry polygons). This pattern suggests that animals were not simply solving the tasks through holistic visual matching strategies. If performance were primarily driven by visual similarity, such as perceptual size-scaling, contour overlap, or spatial frequency, we would expect pixelwise and Fourier similarity to be the dominant predictors of monkey's performance (44). Instead, abstract quantitative features like number and area best explained accuracy, indicating that monkeys were extracting structural properties of the stimuli, such as "how many parts" and "how much stuff," rather than relying on surface-level image features. This provides evidence that nonhuman primates engage similar quantitative dimensions when reasoning about numerical and geometric stimuli and suggests that geometric representations in nonhuman primates may be more structured than previously assumed (Sable-Meyer et al., 2022).

The numerical and geometric tasks also showed further evidence of shared cognitive processes. Cross-lag analyses showed predictive transfer whereby prior accuracy in the Geometry task significantly predicted later performance in the Numerosity task. The lack of a robust effect in the reverse direction suggests asymmetric learning transfer from geometric learning to numerical learning. These findings cannot be explained by domain-general motivation or task engagement because the Control task was a reliable measure of general performance that captured initial learning and fluctuations in engagement. The cross-lag models also controlled for within-task performance over time, filtering out general drifts in attention or motivation. Only cross-task effects that exceeded both within-task correlations and Control task correlations counted as evidence of domain transfer. The overall level of control across analyses strengthens the interpretation that our findings reflect shared processing mechanisms between number and geometry, rather than general cognitive factors.

The stronger predictive path from Geometry to Numerosity echoes developmental research showing that spatial reasoning plays a foundational role in children's early numerical learning. A large body of work has demonstrated that spatial skills—such as spatial scaling and geometric comparison—predict children's ability to learn numerical estimation, symbolic arithmetic, and mathematical reasoning (24, 45–48). These findings have led to the "cognitive bridge" hypothesis: the idea that spatial reasoning serves as a scaffold that supports the construction of numerical concepts during development. The emergence of a similar asymmetry in nonhuman primates suggests that this bridge is not uniquely built through human language, education, or culture. Instead, it may reflect an evolutionarily conserved learning pathway in which spatial structure provides a foundation for extracting and organizing quantitative information, such as the number of elements or the total amount of material, from complex visual input.

Despite their impressive perceptual learning abilities, primates are not known for abstract generalization (35). Studies that test matching and concept learning in nonhuman animals often show limited generalization and transfer across domains (e.g., refs. 36, 37, 49). According to core knowledge theory (32, 33), animals possess distinct, domain-specific systems for representing number and geometry that are fundamentally siloed from one another. Under this view, abstract generalizations and learning transfer between domains would be expected only after cultural scaffolding in humans. However, it is possible that the explicit tasks traditionally used to assess cross-domain transfer in nonhuman animals are not well-suited to detecting shared computations. In research with humans, for example, robust interactions between number and space often emerge not from explicit transfer tasks, but from incidental effects, such as the SNARC effect, where people implicitly link numerical magnitude with spatial position (27). These kinds of cross-domain computations are not the result of explicit generalization and may be unavailable for deliberate use in explicit generalization tasks, particularly by children and animals. Similarly, the learning transfer observed here in monkeys may reflect spontaneous domain interactions that occur implicitly when number and space tap into shared computations.

While our findings reveal interactions between numerical and spatial reasoning in nonhuman primates, it is important to acknowledge that our conclusions are based on correlational models. These approaches, though statistically rigorous and well-controlled, do not allow for definitive causal inference, and the fact that we could test only one kind of control task means we cannot rule out the possibility of unmeasured influences on performance. However, the combination of analyses here offers a robust framework for isolating directionality during learning and testing hypotheses about cognitive interactions in nonverbal subjects. These methods go beyond prior observations of space-number mappings in animals (5, 6) by modeling naïve learning, feature-specific variance, task-specific learning trajectories, and directional transfer over time.

Overall the results suggest that abstract quantitative concepts serve as cognitive bridges between spatial and numerical domains, facilitating generalization and transfer during learning. Primates are capable of decomposing sets and shapes into meaningful features such as number, length, and area. This primitive ability to carve inputs into component dimensions provides a representational framework not only for spatial analysis but also for identifying and comparing quantitative properties across domains, such as the number of elements in a set or the number of sides in a shape. Our study shows that nonhuman primates extract analogous structural information from both numerical and geometric stimuli, which suggests that the bridge between number and space in humans originated in quantitative systems that structure the environment into comparable abstract units. These primitive spatial and numerical abilities, along with computations like arithmetic (50, 51), may form the evolutionary foundation for abstract mathematical reasoning of humans.

Data, Materials, and Software Availability. Behavior data have been deposited in KiltHub (TBD) (52).

ACKNOWLEDGMENTS. We thank the Seneca Park Zoo: Keepers Clare Belden, Jenna Bovee, Linda Velasquez, Maggie Kinsella, Assistant Curator Lindsay Brinda, Superintendent Steve Lacy, and General Curator David Hamilton. We thank the children, parents, and teachers at The Children's School at Carnegie Mellon University. We thank Hugo Angulo for technical innovation and support and Jessica Wegman and Katie Becker for assistance with the Primate Portal exhibit. This work was supported by the NSF DRL2026416 and NSF2148343.

1. S. Dehaene, E. Brannon, Eds., *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought* (Academic Press, 2011).
2. M. D. de Hevia, How the human mind grounds numerical quantities on space. *Child Dev. Perspect.* **15**, 44–50 (2021).
3. N. S. Newcombe, S. C. Levine, K. S. Mix, Thinking about quantity: The intertwined development of spatial and numerical cognition. *Wiley Interdiscip. Rev. Cogn. Sci.* **6**, 491–505 (2015).
4. R. Rugani, M. D. de Hevia, Number–space associations without language: Evidence from preverbal human infants and non-human animal species. *Psychon. Bull. Rev.* **24**, 352–369 (2017).
5. R. Rugani, G. Vallortigara, K. Priftis, L. Regolin, Number-space mapping in the newborn chick resembles humans' mental number line. *Science* **347**, 534–536 (2015).
6. C. B. Drucker, E. M. Brannon, Rhesus monkeys (Macaca mulatta) map number onto space. *Cognition* **132**, 57–67 (2014).
7. J. Gibbon, Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* **84**, 279–325 (1977).
8. R. M. Church, W. H. Meck, J. Gibbon, Temporal integration in duration and number discrimination. *J. Exp. Psychol. Anim. Behav. Process.* **11**, 591–597 (1985).
9. C. R. Gallistel, R. Gelman, Non-verbal numerical cognition: From reals to integers. *Trends Cogn. Sci.* **4**, 59–65 (2000).
10. V. Walsh, A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* **7**, 483–488 (2003).
11. K. J. Holyoak, A. L. Glass, Recognition confusions among quantifiers. *J. Verbal Learn. Verbal Behav.* **17**, 249–264 (1978).
12. J. E. Hummel, K. J. Holyoak, A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* **110**, 220–264 (2003).
13. C. D. Bonn, J. F. Cantlon, Spontaneous, modality-general abstraction of a ratio scale. *Cognition* **169**, 36–45 (2017).
14. S. Dehaene, L. Cohen, Towards an anatomical and functional model of number processing. *Math. Cogn.* **1**, 83–120 (1995).
15. S. Dehaene, L. Cohen, Cultural recycling of cortical maps. *Neuron* **56**, 384–398 (2007).
16. J. F. Cantlon, M. L. Platt, E. M. Brannon, Beyond the number domain. *Trends Cogn. Sci.* **13**, 83–91 (2009).
17. T. Gebuis, R. Cohen Kadosh, W. Gevers, Sensory integration in the numerical brain: A theoretical review. *Front. Psychol.* **7**, 1253 (2016).
18. E. M. Hubbard, M. Piazza, P. Pinel, S. Dehaene, Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* **6**, 435–448 (2005).
19. Z. C. Hawes, K. A. Gilligan-Lee, K. S. Mix, Effects of spatial training on mathematics performance: A meta-analysis. *Dev. Psychol.* **58**, 112 (2022).
20. L. Cantrell, L. B. Smith, Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition* **128**, 331–352 (2013).
21. S. F. Lourenco, L. S. Aulet, A theory of perceptual number encoding. *Psychol. Rev.* **130**, 155 (2023).
22. K. S. Mix, J. Huttenlocher, S. C. Levine, *Quantitative Development in Infancy and Early Childhood* (Oxford University Press, 2002).
23. B. N. Verdine, R. M. Golinkoff, K. Hirsh-Pasek, N. S. Newcombe, Spatial skills, their development, and their links to mathematics. *Monogr. Soc. Res. Child Dev.* **82**, 7–30 (2017).
24. E. A. Gunderson, G. Ramirez, S. L. Beilock, S. C. Levine, The relation between spatial skill and early number knowledge: The role of the linear number line. *Dev. Psychol.* **48**, 1229–1241 (2012).
25. S. F. Lourenco, M. R. Longo, General magnitude representation in human infants. *Psychol. Sci.* **21**, 873–881 (2010).
26. M. D. de Hevia, E. S. Spelke, Number-space mapping in human infants. *Psychol. Sci.* **21**, 873–881 (2010).
27. S. Dehaene, S. Bossini, P. Giraux, The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* **122**, 371–396 (1993).
28. R. Rugani, G. Vallortigara, K. Priftis, L. Regolin, Numerical magnitude, rather than individual bias, explains spatial numerical association in newborn chicks. *eLife* **9**, e54662 (2020).
29. A. Viarouge, M. D. Hevia, What makes different number-space mappings interact? *Psychol. Res.* **88**, 2389–2400 (2022).
30. F. Hurewitz, R. Gelman, B. Schnitzer, Sometimes area counts more than number. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19599–19604 (2006).
31. M. Sablé-Meyer *et al.*, Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023123118 (2021).
32. E. S. Spelke, *What Babies Know: Core Knowledge and Composition* (Oxford University Press, 2022), vol. 1.

33. E. S. Spelke, K. D. Kinzler, Core knowledge. *Dev. Sci.* **10**, 89–96 (2007).
34. A. A. Roberts, W. A. Mazmanian, Concept learning at different levels of abstraction by pigeons, monkeys, and people. *J. Exp. Psychol. Anim. Behav. Process.* **14**, 247–260 (1988).
35. W. A. Premack, Human and animal cognition: Continuity and discontinuity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7871–7878 (2018).
36. E. A. Wasserman, T. R. Zentall, Eds., *Comparative Cognition: Experimental Explorations of Animal Intelligence* (Oxford University Press, 2006).
37. A. A. Wright, J. S. Katz, Mechanisms of same/different concept learning in primates and avians. *Behav. Processes* **72**, 234–254 (2006).
38. A. M. Barnard *et al.*, Inherently analog quantity representations in olive baboons (Papio anubis). *Front. Psychol.* **4**, 253 (2013).
39. J. F. Cantlon *et al.*, The origins of counting algorithms. *Psychol. Sci.* **26**, 853–865 (2015).
40. S. Ferrigno, K. D. Hughes, J. F. Cantlon, Precocious quantitative cognition in monkeys. *Psychon. Bull. Rev.* **23**, 141–147 (2016).
41. S. Ferrigno, Y. Huang, J. F. Cantlon, Reasoning through the disjunctive syllogism in monkeys. *Psychol. Sci.* **32**, 292–300 (2021).
42. S. N. Wood, *Generalized Additive Models: An Introduction with R* (Chapman and Hall/CRC, 2017).
43. G. L. Simpson, gratia: An R package for exploring generalized additive models. arXiv [Preprint] (2024). https://doi.org/10.48550/arXiv.2406.19082 (Accessed 31 July 2025).
44. C. T. Zahn, R. Z. Roskies, Fourier descriptors for plane closed curves. *IEEE Trans. Comput.* **100**, 269–281 (2009).
45. A. Henik, Y. Gliksman, A. Kallai, T. Leibovich, Size perception and the foundation of numerical processing. *Curr. Dir. Psychol. Sci.* **26**, 45–51 (2017).
46. K. S. Mix, Y.-L. Cheng, The relation between space and math: Developmental and educational implications. *Adv. Child Dev. Behav.* **42**, 197–243 (2012).
47. B. N. Verdine, R. M. Golinkoff, K. Hirsh-Pasek, N. S. Newcombe, Spatial skills, their development, and their links to mathematics. *Monogr. Soc. Res. Child Dev.* **82**, 7–30 (2017).
48. J. L. Booth, R. S. Siegler, Numerical magnitude representations influence arithmetic learning. *Child Dev.* **79**, 1016–1031 (2008).
49. S. M. Fleming, M. J. Beran, D. A. Washburn, Disconnect in concept learning by rhesus monkeys (*Macaca mulatta*): Judgment of relations and relations-between-relations. *J. Exp. Psychol. Anim. Behav. Process.* **33**, 55 (2007).
50. J. F. Cantlon, How evolution constrains human numerical concepts. *Child Dev. Perspect.* **12**, 65–71 (2018).
51. J. F. Cantlon, D. J. Merritt, E. M. Brannon, Monkeys display classic signatures of human symbolic arithmetic. *Anim. Cogn.* **19**, 405–415 (2016).
52. J. Cantlon, Cognitive bridge between geometric and numerical learning in monkeys. Kilthub. https://kilthub.cmu.edu/articles/dataset/_b_Cognitive_Bridge_between_Geometric_and_Numerical_Learning_in_Monkeys_b_/29716526. Deposited 31 July 2025.