Title
Was the Ishango Bone—'The Oldest Numerical Device'—a Mathematical Tool? An Empirical Approach

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Parkinson-Coombs, Oisin
Courtenay, Lloyd Austin
Geis, Lila
et al.

Publication Date
2026

Was the Ishango Bone — *'The Oldest Numerical Device'* — a Mathematical Tool? An Empirical Approach

Oisín Parkinson-Coombs (oisin.parkinson-coombs@uib.no)[1,2], Lloyd Courtenay[3], Lila Geis[3], Francesco D'Errico[3] & Rafael Núñez[1,2]

[1]University of Bergen, Norway
[2]ETH Zurich, Switzerland
[3]Université de Bordeaux, France

Abstract

The Ishango bone — a 10cm, approximately 20,000 year old notched bone — has been frequently claimed as evidence of human numerical competence in the Upper Paleolithic, on the basis of patterns in the collections of notches' quantities. We test whether the notch collections are suitable for university-level numerate adults to perform numerical tasks precisely. We find that performance on a range of simple numerical tasks is significantly below the precision required for exact calculation. Moreover, in an explicit counting task, we find that increasing the size of the bone to about hand-length substantially improves performance. These results are at odds with numerical interpretations of the Ishango bone, and raises the question of why the notches were produced at such a small scale given these performance difficulties and their amelioration after moderate size increases.

Keywords: numerical cognition; tally; Ishango bone; cognitive archaeology

Introduction

The Ishango bone — a 10cm long worked bone with 16 collections of notches, discovered in what today is the Democratic Republic of Congo (de Heinzelin, 1957) and dated to approximately 19-25,000 years old (Crevecoeur et al., 2014; de Heinzelin, 1957) — has been purported to be one of the earliest human examples of mathematical notation and numerical competence (de Heinzelin, 1957, 1962). Owing solely to patterns in the quantity of notches and their arrangement, de Heinzelin initially remarked that they 'suggest knowledge of multiplication by two. . . [and] suggest that 10 was the base of the number system and the concept of prime numbers was understood' (de Heinzelin, 1962, p. 114).

These patterns (Fig. 1) are compelling and these claims have been repeated in both the literature on numerical cognition (Butterworth, 1999; Huylebrouck, 2019) and popular documentaries (BBC, 2005), history books (Darling, 2004), and even postage stamps. These numerical interpretations of the Ishango bone, along with other pre-historic artefacts, are frequently levied to support nativist accounts of number concept (Butterworth, 1999; Butterworth et al., 2018; Dehaene et al., 2025; Nieder, 2019). More recent scholarship has moderated the claims slightly while retaining the central account of the artefact as a numerical device. Some, while dismissing the invocation of prime numbers, have drawn attention to 'what seems less coincidental... [that] the last two columns contains the same total number of marks: sixty' (Everett, 2017, p. 35). Others have referred to the Ishango bone as 'the oldest known artefact to show a tallying system for numerosities' (Pantsar, 2024, p. 107), and have hypothesised that the notches may have served as 'some kind of numeric reference table' (Everett, 2017, p. 36). These interpretations reflect a broader acceptance in the literature that prehistoric notch collections readily support exact numerical practices, as seen in claims that: 'a stick engraved with forty-nine notches keeps a permanent track of this exact number. To verify whether a count is correct, one *merely* has to go through the objects one by one and move forward by one notch for each object' (Dehaene, 1997, p. 84, emphasis added).

As suggestive as the patterns of notch collections are, they alone are insufficient to support the claim of a numerical or tallying function for the artefact. To address the numerical purpose one must focus not only on these patterns, but also on the **production**, **archaeological context**, and **use** of the Ishango bone. Accounts to date have made little engagement with the production of the notches: Were all the notch collections (or notches within a collection) made at the same moment, or did they slowly accumulate over time? How are claims of doubling altered if there was a period where only the doubland collection were present on the artefact? Further, this bone is the only[1] known artefact from the site that exhibits collections of notches (Overmann, 2025). How should accounts countenance rich mathematical interpretations with the rather unique quality of such an artefact? Moreover, there has been little analysis as to how, and in what context, the artefact was used: Was it used for direct 1-1 comparisons with other collections of notches or physical items? Are scholars supposing the presence of a 20 thousand year old linguistically conventionalised countlist operating in concert with these tallies of notches? Is *merely* moving forward one notch for each object as readily achievable as it is presented? In the only line of dissent to the dominant narrative, Overmann (2019, 2023) addresses barriers to using the Ishango bone for numerical purposes, noting that the notches 'are too small for incremental one-to-one comparisons' (2019, p. 145) and

[1]There is one other notched bone from this site, however its notches do not come in collections but are randomly spread across the artefact, and this bone has never been archaeologically published (Overmann, 2025). Further, the two bones are not contemporaneous, the Ishango bone being part of the NfPr layer, slightly more recent than the SD layer in which the second bone was found (Crevecoeur et al., 2014).

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

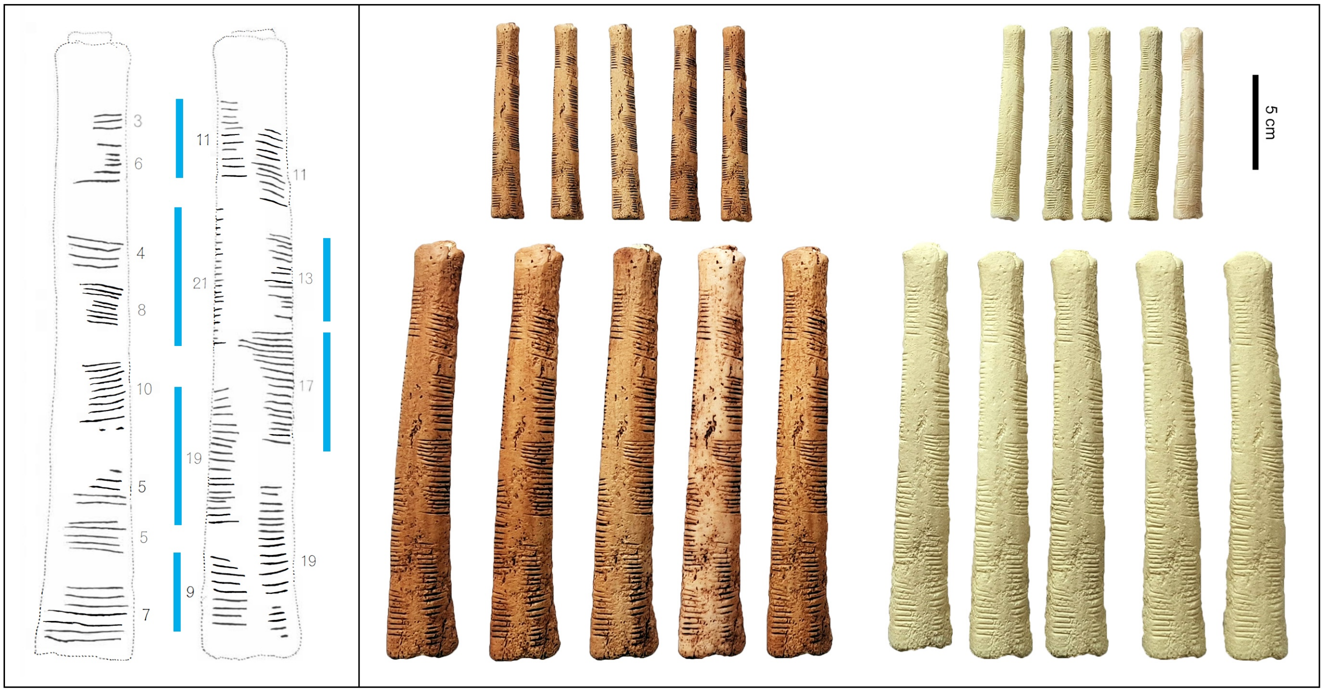

Figure 1: *Left*: Ishango bone sketches reproduced from de Heinzelin (1957), the best preserved notch collections used in experiments identified in blue. *Right*: Replicas used in experiments, at both 1:1 and 2:1 scale and with and without ochre.

would require counting and a numeral lexicon. Although, here too, this speculation is offered without any evidence as to whether these notches are in fact too small, that is, there is no argument or empirical evidence that the measured sizes of these notches actually presents a challenge to their use.

Numerical Function: An Empirical Approach

In this study, we address one horn of the issue regarding numerical interpretation: the **use** of the Ishango bone for numerical purposes. The supposed numerical functions of this object, from tallies to arithmetic, all require precision. As Overmann (2019) notes, the size of the object and notches — the artefact has 168 notches and is about the same size as a standard BIC pen — appear to present a challenge to the plausibility that it could be used to manage 1-1 comparisons. If those using the object could not do so accurately and reliably, then there would be little reason to describe this as a numerical artefact. Moreover, given the small size of the artefact, a reasonable increase in scale would not produce an unwieldy object. If such an increase substantially improved accuracy of performance, then to maintain the numerical interpretation, one must explain why the notches were produced at such a small scale. To be numerical requires precision; this does not preclude that the artefact may have been used to record approximate quantities, or may have been produced precisely without the capacity to read it precisely, and may have been an important step or opportunity to develop a properly numerical capacity. Investigating whether the artefact can support properly numerical functions does not preclude that artefacts of these natures might play an important role in the history of number, nor that these artefacts and their collections are owed an explanation. We test the hypothesis that the artefact functioned as a numerical device by examining whether modern numerate adults:

1. can use physical replicas accurately for basic numerical tasks, and
2. if their performance is substantially improved on replicas printed at double scale, roughly the size of an extended hand.

Although there is no current evidence of ochre on the Ishango bone, ochre has been found at the Ishango site (de Heinzelin, 1957) and could enhance discriminability, so we stained half of the replicas with ochre as a means of being generous to claims of numerical function.

Methods

Replica Preparation

Our team prepared replicas of the Ishango bone (Fig. 1). First, 3D-scans of the Ishango bone (now at the Royal Belgian Institute of Natural Sciences in Brussels), were made using a micro-CT (Skyscan1076), at 35 micron resolution, and a portable confocal microscope (Mahr *MarSurf CM mobile*) at magnification x20, lens 800L. The scans were then segmented to extract only the external surface of the bone, and these were used to produce 3D printed replicas of the artefact at two scales — small, 1:1; and large, 2:1 — on a FORMLABS Form 4 Pack Basic. The replicas are then washed and cured in a UV light chamber Form Cure V2 230V. The replicas were treated with a matt spray (*Prince August Base: Bone*) to reduce shine. Half of the replicas were stained with ochre (*Alumanation Manufacturing CC*), and sprayed with varnish (*Spectrum*) to fix the ochre in place to prevent wear during the experiments.

Experimental Design

This study was conducted with the corresponding ethical approvals according to the European Research Council's standards, and consisted of four experiments. There were 48 participants (19M, 29F; 46 right-handed, 2 left-handed). Participants were undergraduate and masters students from two European universities and were compensated for the participation. All experimental instructions were in English, however, for some tasks (involving alphabet recitation and counting; see below) participants used their primary language (or the English alphabet in cases where the participant's primary language did not use an alphabetic script). Participants were assigned to one of four conditions in a 2×2 between-subjects design manipulating replica scale (Small, 1:1 vs Large, 2:1) and ochre staining (Ochre vs No Ochre), with 12 participants per condition. We manipulated Scale to investigate our second research question, whether moderate increases in size and discriminability substantially improve accuracy. Participants completed four tasks in a fixed order. For each trial of each experiment, participants were presented with a covered replica on a laminated sheet with an arrow indicating the relevant collection(s) for that trial. Trials were presented in random order. Before the experimental trials, participants completed a demonstration trial with a laminated image with dot collections which resembled the replica's structure. In Experiment 1 participants were not allowed to touch the replica, but in all other experiments they were instructed that they could handle the replica in whatever manner they deemed suitable. Only the 6 best preserved notch collections were used in the experiments, these are indicated in Fig.1, and have 9, 11, 13, 17, 19, and 21 notches respectively.

Experiment 1: Perceptual Comparison Participants made rapid forced-choice judgments about which of two simultaneously presented notch collections contained more notches, without counting. Six comparisons were presented (9v21, 11v17, 9v13, 17v21, 9v11, 19v21), each presented once. Participants were instructed to respond as quickly as possible based on visual impression, but without any strict limitations on time, with each comparison presented precisely once. This experiment tested whether the artefact's grouping and spacing support relative quantity discrimination, and familiarised participants with the artefact before the tasks requiring precision.

Experiment 2: Alphabet-Notch Correspondence Participants were told that each notch collection represented a letter in the alphabet. They were asked to perform a 1-1 mapping of the alphabet to the notches in the indicated collection out loud, and state what letter the collection represented at the end. Each of the 6 notch collections indicated in Fig. 1 was presented once. This experiment tested if participants were able to accurately perform 1-1 correspondences (without numerical content) between the replica notches and a conventionalised, ordered list. This is procedurally equivalent to a counting task but without providing the final count to the participant.

Experiment 3: 1-1 Correspondence under Verbal Suppression In addition to the standard set up, participants were provided with an empty bowl, and a bowl containing white and blue beads in a 3:1 ratio. Prior to the demonstration and experimental tasks, participants learned and recited a short phrase, that they repeated throughout the task trials. Participants were instructed to place as many blue beads in the empty bowl as there were notches in the indicated collection, while repeating the practiced phrase at all times. There were 4 trials in this experiment, restricted to the notch collections (13, 17, 19, 21), due to the complexity and duration of the task. This experiment tested whether the participants were able to accurate construct a collection of equal cardinality to a notch collection, when they were prevented from using their conventional counting terms. This experiment was designed to test whether the replica supported 1-1 comparisons without the deployment of a numeral lexicon, similar to the comparison on 1-1 correspondence tasks between anumerate Pirahã adults and numerate Western adults performing a verbal interference task introduced by Frank et al. (2012).

Experiment 4: Explicit Counting Participants were asked to count all of the notches on the righthand side of the replica (as indicated in Fig. 1). Once finished, they were asked to do it again, though the replica was changed and the orientation switched so that the participant was not aware they were counting the same collections of notches. This experiment tested whether participants were able to accurately count all of the notches on one side of the artefact and whether they were consistent in their own counts, allowing that they might exclude/include a valid/invalid notch but should be consistent in doing so.

Results

Accuracy was poor across all experiments (30-63%; Tab. 1), with mean error magnitudes of 0.5-1.5 notches in Experiments 2-4 (Table 1 and Fig. 2). Even the relatively simple task of explicit counting (Exp. 4, their final experiment), after extensive experience with the replicas, failed 70% of the time.

Relative Discrimination: Experiment 1

To test the accuracy of participants' perceptual discriminations across all trials, we compared results to a simulation of expected performance in standard psychophysical tests of relative quantity discrimination (Dehaene et al., 1998) using Weber's Law with Weber fractions from 0.10 to 0.30 based on empirical estimates (Halberda & Feigenson, 2008). Binomial tests indicated that observed accuracy (63%) fell significantly below expected baseline across all simulated Weber fractions

Exp.	No. Trials	Prop. Correct	Mean Error	SD
Exp 1	288	61%	N/A	N/A
Exp 2	281	63%	0.55	0.87
Exp 3	186	36%	1.3	1.3
Exp 4	95	30%	1.6	1.8

Table 1: Accuracy and mean error in Experiments 1-4, across all conditions, after outliers were removed. For Exp. 2 and 3 we excluded trials with errors greater than 5 (2.5% and 3.1%, respectively) and for Exp. 4 with errors greater than 10 (1%).

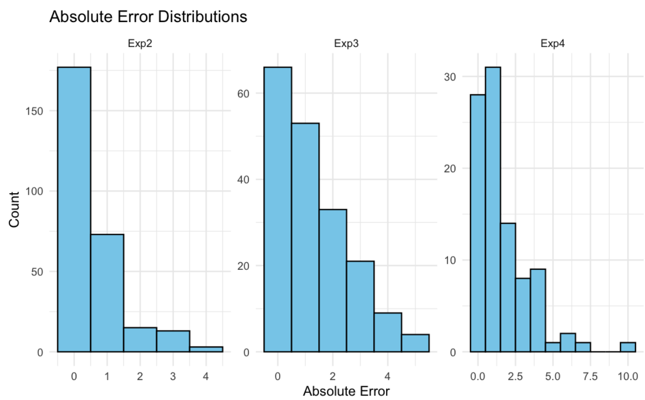

Figure 2: Distribution of errors in Experiments 2 - 4, across all condition, after outliers were removed.

($p < .001$ in all cases), including the most conservative estimate ($w = 0.30$, expected accuracy 77%).

To test whether experimental conditions affected discrimination accuracy, we fit a generalized linear mixed model (GLMM) with the formula: `Accuracy` ~ `Scale + Ochre + (1|ID) + (1|Trial_ID)`. We evaluated main effects only, as preliminary models indicated no significant Scale × Ochre interaction. We evaluated main effects only, as preliminary models indicated no significant Scale × Ochre interaction ($\chi^2(1) = 0.06$, $p = .80$). Neither Scale ($\beta = 0.18$, $SE = 0.44$, $z = 0.40$, $p = .69$) nor Ochre ($\beta = 0.41$, $SE = 0.45$, $z = 0.93$, $p = .35$) significantly predicted accuracy.

Precision Tasks: Experiments 2-4

To address our primary research questions — 1) whether numerate adults could use the replicas accurately for precise tasks, and 2) whether increasing the scale of the replicas substantially improved performance — we fit generalized linear mixed models (GLMMs) for binary accuracy outcomes and linear mixed models (LMMs) for error magnitude for Experiments 2, 3, and 4, separately. Fixed effects in all models were Scale (Large vs Small), Ochre (Ochre vs No Ochre), and Indexing strategy (described below). For accuracy models, we included an offset term to serve as a baseline — supposing that a hypothetical "proficient counter" using a "suitable numerical device" should yield a correct response for any given collection 90% of the time — such that model intercepts re-

Exp.	**Predictor**	β	**SE**	z	***p.holm***
Exp 2	Intercept	-1.64^{**}	0.52	-3.18	.001
	Scale	-0.55	0.40	-1.37	.172
	Ochre	-0.14	0.40	-0.36	1.00
	Indexing	$0.78^{\dagger}$	0.39	1.98	.079
	Random: $\sigma_{\text{ID}} = 0.97$, $\sigma_{\text{Collection}} = 0.76$				
Exp 3	Intercept	-2.68^{***}	0.52	-5.20	<.001
	Scale	-0.49	0.41	-1.21	.218
	Ochre	-0.02	0.41	-0.05	1.00
	Indexing	0.064	0.46	0.14	.888
	Random: $\sigma_{\text{ID}} = 0.83$, $\sigma_{\text{Collection}} = <0.01$				
Exp 4	Intercept	-3.03^{***}	0.65	-4.66	<.001
	Scale	-1.38^{*}	0.59	-2.33	.020
	Ochre	0.48	0.54	0.88	.755
	Indexing	0.29	0.60	0.48	.630
	Random: $\sigma_{\text{ID}} = 0.69$				

Table 2: Fixed and random effects for accuracy models (GLMMs) with binomial family and logit link: `Accuracy ~ Scale + Ochre + Indexing + (1|ID) + (1|Collection)`. Coefficients are on the log-odds scale; negative values indicate decreased accuracy. The model intercept tests whether performance differs from the 90% baseline. Reference levels: Scale = Large (2:1), Ochre = Absent, Indexing = False. p-values are Holm-Bonferroni corrected. $^{\dagger}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Exp.	**Predictor**	β	**SE**	t	***p.holm***
Exp 2	Intercept	0.35^{***}	0.09	3.93	<.001
	Scale	0.13	0.07	1.81	.154
	Ochre	-0.01	0.07	-0.09	1.00
	Indexing	$-0.14^{\dagger}$	0.07	-2.10	.079
	Random: $\sigma_{\text{ID}} = 0.19$, $\sigma_{\text{Collection}} = 0.12$, $\sigma_{\text{residual}} = 0.38$				
Exp 3	Intercept	0.67^{***}	0.13	4.99	<.001
	Scale	0.17	0.10	1.63	.218
	Ochre	-0.00	0.10	-0.08	1.00
	Indexing	-0.12	0.11	-1.04	.601
	Random: $\sigma_{\text{ID}} = 0.25$, $\sigma_{\text{Collection}} = 0.09$, $\sigma_{\text{residual}} = 0.49$				
Exp 4	Intercept	0.76^{***}	0.15	4.92	<.001
	Scale	0.38^{**}	0.13	2.88	.012
	Ochre	-0.11	0.13	-0.83	.755
	Indexing	-0.18	0.14	-1.27	.415
	Random: $\sigma_{\text{ID}} = 0.27$, $\sigma_{\text{residual}} = 0.51$				

Table 3: Fixed and random effects for error models (LMMs) on log-transformed error: `log(1 + |Error|) ~ Scale + Ochre + Indexing + (1|ID) + (1|Collection)`. Positive coefficients indicate increased error magnitude. Reference levels: as per Table 2. p-values are Holm-Bonferroni corrected. $^{\dagger}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

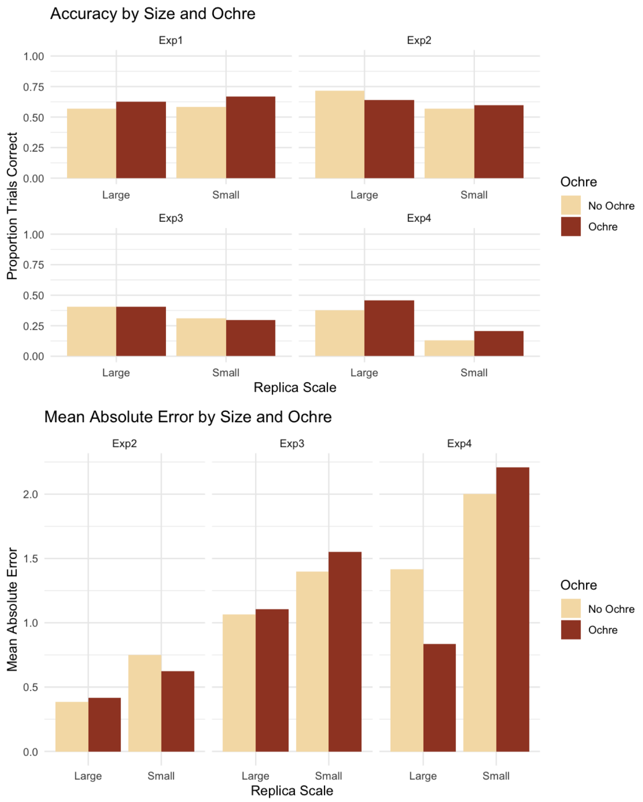

Figure 3: *Top*: Proportion of correct trials in Experiments 1 - 4 by condition. *Bottom*: Mean error of trials in Experiments 2 - 4 by condition.

flect deviations from this performance level (as opposed to chance). This baseline was selected as, with respect to the question of whether the artefact functioned as a properly numerical baseline, then it should support accurate performance with allowance for some mistakes in execution. For instance, the artefact could educate the Approximate Number System (Dehaene, 1997), stabilizing and improving its discriminations, while still failing to be a properly numerical device. Thus, we take perfect performance with minor slips, instantiated by our 90% threshold, as baseline, instead of ANS performance. For all models we included participants' ID as a random intercept. In Experiments 2 and 3 we treated Collection_ID as a random intercept rather than a fixed effect as this rendered the model intercepts interpretable relative to the hypothetical baseline, while still accounting for variation in the difficulty across notch collections. Interactions never improved fit, so we report main-effects models throughout.

Strategy Prevalence Participants spontaneously adopted different strategies: some used manual indexing (finger contact to serially track notches) while others relied on vision alone. Across experiments, 61-76% of trials involved indexing. Most participants consistently used one strategy throughout (Exp. 2: 52% always indexed, 33% never; Exp. 3: 65% always, 17% never; Exp. 4: 65% always, 21% never). Indexing improved accuracy by 11.7, 2.6, and 4.0 percentage points in Experiments 2, 3, and 4, respectively. Given that manual indexing should theoretically facilitate serial enumeration by reducing working memory load and preventing double-counting, and given the observed performance differences, we included Indexing as a covariate in all models to isolate effects of our experimental manipulations from individual differences in self-selected strategy.

Overall Performance Model intercepts indicated that accuracy was significantly below the 90% baseline in all three experiments (all $p \leq .001$; Table 2). Ochre showed no effects on accuracy or error in any experiment (all $p > .29$; Tables 2 and 3). Manual indexing significantly improved accuracy in Exp. 2 ($p = .048$) and reduced error ($p = .040$), but showed no significant effects in Exp. 3 or Exp. 4 (all $p > .21$).

Scale Effects Small scale significantly impaired accuracy only in Experiment 4 ($\beta = -1.35$, $p = .022$; Table 2), with no significant effects in Experiments 2 or 3 ($p > .17$ in both; Table 2). Similarly, small scale only significantly increased the magnitude of errors in Exp. 4 ($\beta = 0.35$, $p = .007$; Table 3), although there were marginal scale effects in both Exp. 2 and 3 (both $p < .10$; Table 3). To quantify the practical impact of scale in this condition, we computed estimated marginal means from the fitted models. These predictions represent expected performance for a typical participant at each scale, averaging over factors (Ochre, Indexing) and accounting for random variation across participants and collections. Model predictions indicate that increasing replica scale from small (1:1) to large (2:1) improved predicted accuracy by approximately 25 percentage points (from 14% to 39%) and reduced predicted error by approximately 0.77 notches (from 1.60 to 0.83).

Discussion

While the Ishango bone has been claimed to provide evidence of tallying and arithmetical competence as early as the Upper Paleolithic, here we presented an empirical investigation into the plausibility of its **use** for precise numerical purposes. Across all experiments performance was poor, with accuracy substantially below baseline. Our results indicate that numerate, university-level adults struggle to perform accurately when using replicas of the Ishango bone. This raises questions to the plausibility that the Ishango bone was used to store numerical facts or could have been used for precise arithmetic. Moreover, in a simple and explicit counting experiment, moderately modifying the scale of the bone replicas yielded a significant improvement in both accuracy and error magnitude, with model predictions of accuracy almost tripling and errors approximately halving when scale was doubled. Even with the addition of ochre to enhance discriminability we see no improvement in accuracy or error magnitudes, indicating that the difficulty in using the bone replicas is not easily remedied by visual modifications. To hold that the Ishango bone was constructed for a numerical purpose thus requires an account

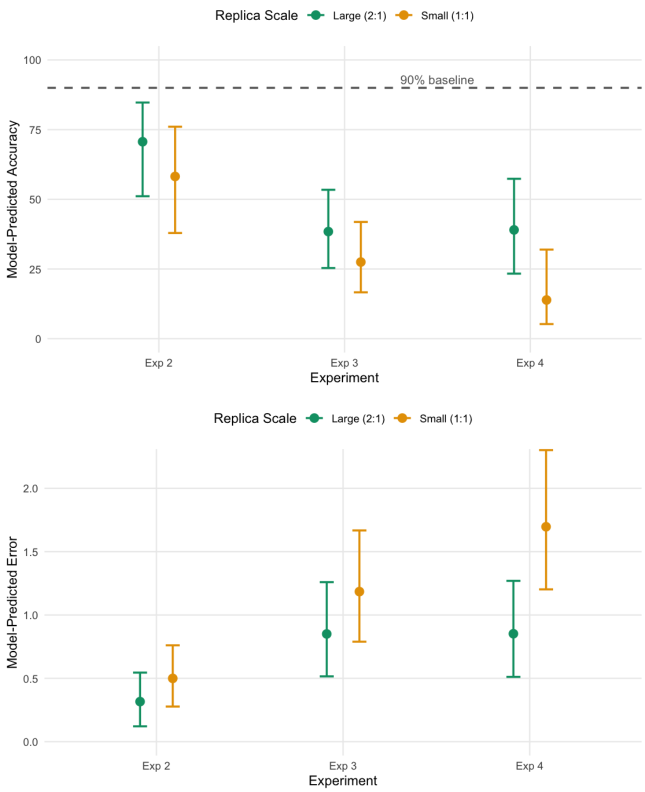

Figure 4: Points show estimated marginal means from the fitted mixed-effects models for large (2:1; green) and small (1:1; orange) replicas in Experiments 2-4; error bars indicate 95% confidence intervals. Predictions represent expected performance for a typical participant, averaging over other experimental factors and accounting for random variation across participants and collections. The dashed line in the accuracy panel indicates the 90% baseline.

of why it was made at its small scale given the further costs to accuracy and errors it invokes.

The inferences from the results of these experiments to interpretations regarding the use and purpose of the Ishango bone relies on the appropriateness of the performance of modern, numerate, university-level participants in these experiments as model of Upper Paleolithic humans performance. Although the experiments were set up to be generous to the numerical hypothesis in the experimental environment, language of execution, and the order of experiments, modern numerate humans differ substantially from the hypothetical Ishango practitioner in their familiarity with notched objects. If the Ishango bone was a numerical device, then one must suppose that the Upper Paleolithic humans using it for such were familiar with handling it as such. Our experimental participants were naïve in this regard, as evidenced by the differences in strategy usage, with many frequently relying on visual inspection alone. These differing strategies not only decrease the statistical power of our experiments, but also leave open the question that perhaps both the poor performance and effects of scale were the result of participants unfamiliarity. In future work we intend to repeat these experiments under a training paradigm such that participants have a lot more familiarity with the replicas prior to the experimental trials, and are instructed and practiced in the procedure of indexing.

A separate issue that could undermine the strength of the theoretical inference regards the baselines against which we are comparing performance. Although we set 90% as a plausible baseline for accuracy, this is not empirically validated. Moreover, the baseline for our error models is a mean error of 0, which also may not be valid. Comparing participant performance for these tasks on the Ishango replicas with other material counting devices, like abaci, at the same scale would provide a more appropriate comparison.

Though to the modern observer the putative numerical patterns in the notch collections of the Ishango bone are suggestive of arithmetic competence, they alone can not substantiate this interpretation. Our work highlights the need to analyse the Ishango bone, and other prehistoric notched artefacts, from the perspective of their capacity to support precision, as required in arithmetic and numerical records. To claim that a notched object is numerical one needs to identify the specific cognitive tools available and capacities required to operate appropriately, and then to test whether the object as constructed supports such. Moreover, if moderate modifications substantially improve performance then one must further account for the object's specific construction in the face of such results. This methodological approach is intended to complement studies of the production of the notches, as well as the archaeological context of the Ishango bone, to inform our interpretations of its supposed numeric function. With respect to the Ishango bone it appears that it does not support precision in numeric tasks, and the question of its function and the origin of the pattern of notches remains a subject of study.

References

BBC. (2005). The story of one. *British Broadcasting Corporation*.

Butterworth, B. (1999). *The mathematical brain*. Macmillan.

Butterworth, B., Gallistel, C. R., & Vallortigara, G. (2018). Introduction: The origins of numerical abilities.

Crevecoeur, I., Skinner, M. M., Bailey, S. E., Gunz, P., Bortoluzzi, S., Brooks, A. S., Burlet, C., Cornelissen, E., De Clerck, N., Maureille, B., et al. (2014). First early hominin from central africa (ishango, democratic republic of congo). *PloS one*, *9*(1), e84652.

Darling, D. (2004). *The universal book of mathematics: From abracadabra to zeno's paradoxes*. Wiley.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in neurosciences*, *21*(8), 355–361.

Dehaene, S., Sablé-Meyer, M., & Ciccione, L. (2025). Origins of numbers: A shared language-of-thought for arithmetic and geometry? *Trends in Cognitive Sciences*.

Dehaene, S. (1997). *The number sense*. Oxford University Press.

de Heinzelin, J. (1957). Les fouilles d'ishango. Institut des Parcs Nationaux du Congo Belge.

de Heinzelin, J. (1962). Ishango. *Scientific American*, *206*(6), 105–118. http://www.jstor.org/stable/24936573

Everett, C. (2017). *Numbers and the making of us: Counting and the course of human cultures*. Harvard University Press.

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive psychology*, *64*(1-2), 74–92.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the" number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental psychology*, *44*(5), 1457.

Huylebrouck, D. (2019). *Africa and Mathematics: From Colonial Findings Back to the Ishango Rods*. Springer International Publishing. https://doi.org/10.1007/978-3-030-04037-6

Nieder, A. (2019). *A brain for numbers: The biology of the number instinct*. MIT press.

Overmann, K. A. (2019). *The material origin of numbers*. Gorgias Press. https://doi.org/10.31826/9781463240691

Overmann, K. A. (2023). *The materiality of numbers: Emergence and elaboration from prehistory to present*. Cambridge University Press.

Overmann, K. A. (2025). The ishango bone. In *Cultural number systems: A sourcebook* (pp. 53–58). Springer.

Pantsar, M. (2024). *Numerical cognition and the epistemology of arithmetic*. Cambridge University Press.