Mental abacus represents large exact numerosities using pre-existing visual resources

Michael C. Frank[1] and David Barner[2]

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

[2]Department of Psychology, University of California, San Diego

Word count: 3905 main text + 46 acknowledgements + 32 notes = 3984

Address for correspondence: Michael C. Frank, 77 Massachusetts Avenue, MIT Room 46-3037D, Cambridge, MA 02139

Acknowledgments: The two authors contributed equally to this work. We gratefully acknowledge the students and staff at UCMAS Gujarat. Thanks to George Alvarez, Timothy Brady, Susan Carey, and Rebecca Saxe. This project supported by a UCSD Senate Award to DB and a Jacob Javits Fellowship to MCF.

Abstract

Only humans who learn to count can represent large exact numerosities, leading some to speculate that language supports the development of new representational resources. Here we explore an alternative system of numerical representation, “mental abacus” (MA), which is believed to rely on visual imagery. We ask how attested visual mechanisms could represent MA, and whether they are transformed by MA training or are simply redeployed. We tested children in India who were proficient in MA and also untrained adults. For both groups, we found that MA involves a redeployment of pre-existing visual resources, and that each abacus column is encoded as a separate “model” in visual working memory. This contrasts with alternative accounts of perceptual expertise, which posit sharpening of the perceptual system or extensive experience with the statistical structure of a domain. We conclude that large exact numerosities can be encoded without requiring new representational resources.

Word count: 147 words

Humans, unlike other animals, can perform exact numerical computations. Although other creatures are sensitive to precise differences between small quantities and can represent the approximate magnitude of large sets, no non-human species can represent and manipulate large, exact numerosities. In development, learning a linguistic counting system always precedes conceptual understanding of large numbers (Wynn, 1990; Le Corre, 2006), and indigenous groups who lack numerical vocabulary are unable to represent the exact magnitudes of large sets (Frank et al., 2008; Gordon, 2004; Pica et al., 2004). Such studies suggest that language may drive the creation of new representational resources in humans.

Language, however, is not the sole cognitive system capable of symbolically representing exact number. Experienced users of an abacus—a physical calculation device—learn to perform arithmetic computations mentally, as though visualizing a "mental abacus" (MA) (Hatano, 1877; Hatano & Osawa, 1983; Hishitani, 1990; Stigler, 1984; Stigler et al., 1986; Miller & Stigler, 1991). MA thus provides an important case study of building complex representations using non-linguistic cognitive resources. The current study asks, given known limitations on the non-linguistic processing of quantity information, how the visual system could support the representation of large exact numerosities using MA.

The abacus has been used in Asia since 1200 AD for rapid and precise calculation (Menninger, 1969). It represents number via the arrangement of beads into columns, where each column represents a place value that increases in value from right to left (Figure 1a). On a Japanese Soroban abacus, each column is divided into two levels separated by a horizontal beam. On the bottom are four "earthly" beads and on top is one "heavenly" bead, whose value is five times greater than the individual earthly beads below. Moving beads towards the dividing beam places the beads in "play," thereby making them count towards the total number represented.

Rather than using the physical device, MA users are trained to visualize an abacus and to move mental beads in order to perform addition, subtraction, multiplication, and division, all with astonishing speed. Errors in computation suggest that the structure of this mental representation mirrors that of the physical device. MA users are more likely to confuse the number of digits in a sum than non-MA users (because they occasionally drop an entire column) and more likely to make calculation errors involving quantities of 5 (due to misrepresentation of heavenly beads) (Stigler, 1984).

MA representations could not be implemented using either of the standard routes for non-linguistic processing of quantity. Representing the number 49 in MA requires tracking the precise location of exactly nine beads. Quantities of three to four objects or sets ("models") and their locations and identities can be represented in visuo-spatial working memory (Alvarez & Cavanagh, 2004; Cowan, 2000; Feigenson, Dehaene, & Spelke, 2004; Luck & Vogel, 1997), but this system is insufficient for representing anything but the smallest quantities in MA. The approximate cardinality of large sets can be represented using the approximate number system (ANS), where error in estimation is proportional to the size of the set being evaluated (Feigenson et al., 2004; Whalen, Gallistel, & Gelman, 1999; Xu & Spelke, 2000). However, because the ANS is approximate, the degree of precision found in normal adults should be unable to support accurate MA representations of large quantities.

We asked how MA representations could supersede the limitations of these systems, with the goal of understanding how non-linguistic visual resources can support the representation of large exact numerosities. We contrasted three hypotheses. First, MA may present a case of approximate number expertise; according to the *perceptual sharpening* hypothesis, extensive MA training improves the acuity of ANS such that participants can precisely represent large MA arrays. For example, an ability to precisely represent an array of 20 beads would support MA representations of up to 9999. It is known that intensive training with either low-level perceptual tasks or higher-level activities can result in a sharpening of many visual abilities including attention, discrimination, and estimation (Fiorentini & Berardi, 1980; Green & Bavelier, 2003), but the way in which these representations are used does not change.

Second, according to the *perceptual chunking* hypothesis, MA training allows its users to chunk bead arrays into larger units, increasing the number of beads that can be represented in parallel. Previous studies report that with extensive training, learners can process "chunks" (collections of objects or parts unified by statistical regularities) more efficiently, leading to faster and more accurate processing (Brady, Konkle, & Alvarez, 2009; Chase & Simon, 1973; Mandler & Shebo, 1982; Orbán, Fiser, Aslin, & Lengyel, 2008). On this account, learning affects the efficiency with which perceptual representations are used, without affecting their acuity.

Third, the *perceptual redeployment* hypothesis is that MA works by integrating visuo-spatial working memory with the ANS, allowing users to store each column in the abacus as a separate model in working memory and maintain a separate estimate of quantity and position for each (Figure 1b). Adults can compute ensemble statistics such as the mean position of a set of objects (Alvarez & Oliva, 2008, 2009; Ariely, 2001) and can represent quantity information for multiple models in parallel (Feigenson, 2008; Halberda, Sires, & Feigenson, 2006). On this account, pre-existing visual resources represent numerical information precisely by working in concert.

We investigated MA in a population of children in Gujarat Province, India, where it is taught in a 3-year after-school program. We used two tasks familiar to children: (1) rapid addition, and (2) reading abacus flashcards (translating cards with representations of abacus states into Arabic numerals). We also conducted a third experiment on abacus reading with college students who had their first exposure to abacus representations in the lab. Consistent with the perceptual redeployment hypothesis, these experiments suggest that MA makes use of the visual system's ability to track several groupings or sets and maintain estimates about their quantity and position.

Experiment 1: Rapid Addition

We first tested participants on an adaptive addition task in which participants summed two vertically-aligned addends (e.g., $\begin{array}{r}278\\+596\end{array}$). Across trials, we varied the number of digits of the addends depending on success on previous trials, in order to adapt the difficulty of the task to each individual participant.

According to the perceptual sharpening hypothesis, MA training increases the precision of ANS representations, allowing users to represent sets of discrete individuals with higher accuracy. On this account, the accuracy of MA computations should be limited by the number of beads that are "in play," since the error in ANS computations should grow with the number of individuals that are being manipulated. In contrast, on the chunking hypothesis, participants have less familiarity with larger abacus configurations, and on the redeployment hypothesis, each MA column is represented as separate model in visual working memory. Thus, according to both, MA computations should be limited by the number of columns (digits) of the addends, not the number of beads.

Methods

Participants. Participants were children enrolled in UCMAS schools in Gujarat, India. Participants were chosen for inclusion on the basis of their completion of level 4 UCMAS training (which includes approximately a year of physical abacus training and an introduction to MA), their ability to travel to the test site, and their instructor's judgment that they were among the best students in their cohort. The 51 participants in Experiment 1 had mean age 11.2 years (range: 7.2–16.3 years).

Stimuli and Procedure. All stimuli were presented using Matlab and Psychtoolbox. Responses were entered on numeric keypads. Instructions were given in Hindi, English, or Gujarati (depending on child preference) and illustrated with examples.

Addends were presented simultaneously on a computer screen until the participant typed an answer, up to a maximum of 10s. The number of digits in the addends was adaptive: following two correct answers, the number of addends increased; following one incorrect answer the number decreased. Stimuli were sampled randomly, conditioned on the number of digits in the resulting representation (e.g., 100 – 999 for a 3-digit trial). Participants received feedback and saw a message indicating that they were out of time if they did not answer within 10s. The task lasted 5 minutes.

Results and Discussion

We used multilevel logistic regression models (Gelman & Hill, 2006) for our analyses. This approach allowed us to model the entire dataset (all trials in all conditions). In each experiment we modeled effects of interest using group-level (fixed) effects; each model also included separate (random) intercepts for each participant.

We used these models to ascertain whether performance was best predicted by the number of beads or number of abacus columns involved in a particular trial (Figure 2). We created separate models with group-level effects of either beads or columns and then compared their fit to the data. Because the number of beads and columns in a display were highly correlated (r = .65), this model-comparison approach provides a principled method for determining which predictor better fits the data. Although both bead and column predictors were highly significant in their respective models (both p values < .0001), the column-based model fit the data far better overall (χ^2 = 504.58, p < .0001).

Consistent with the predictions of the perceptual chunking and perceptual redeployment hypotheses, but not the perceptual sharpening hypothesis, we found that mental addition performance was best predicted by the number of MA columns implicated in the computations.

Experiment 2: Abacus Flashcards

The next experiment compared abacus flashcard "reading" (reporting the cardinality represented on an abacus) to estimation tasks in which participants reported the quantity of beads. The displays in these estimation tasks included identical abacus displays, rotated abacus displays, displays that shared the same "bead" arrangement as an abacus, and random dot arrays. Each account made separate predictions.

Sharpening hypothesis. First, performance should be limited by the number of beads for all conditions. Second, no qualitative differences should exist between estimation and abacus reading tasks for perceptually similar stimuli. For example, flashcard reading performance should not differ from estimation performance in the identical and rotated estimation displays, since the displays present identical arrays of objects. Third, for similar reasons, both the accuracy and variability of estimates should be equivalent when performing estimates for the identical and rotated estimation displays.

Chunking hypothesis. First, performance should be limited by the number of columns. Second, as in the sharpening hypothesis, because expertise consists of learning statistical characteristics of the kinds of configurations used in abacus displays, estimation tasks with shared configural features (e.g., identical, rotate, and configural displays) should be relatively easier than random dot arrays in direct proportion to their similarity to the practiced abacus representation.

Redeployment hypothesis. First, performance should be limited by the number of columns. Second, because abacus reading is a column-based operation and hence requires fewer separate steps than a bead-based operation (in which RT should grow linearly with the number of beads being estimated; Whalen et al., 1999), participants should have a different reaction-time slope on abacus reading than on any of the estimation tasks. Third, error in the parallel estimates of two quantities x and y grows as $\sqrt{x^2+y^2}$ (less than the $\sqrt{(x+y)^2}=x+y$ error of making a single estimate; Cordes, Gallistel, & Gelman, 2007). Thus, the scalar variability of participants' estimates should be lower in the abacus-reading condition than in any estimation condition and hence should show higher accuracy as well.

Methods

Participants. The 133 participants in Experiment 2 had a mean age of 11.2 years, with a range of 6.8–15.0 years. All participants were familiar with abacus reading from their training.

Stimuli and Procedure. Participants were presented schematic images of an abacus (flashcards) for 500ms on a computer screen and were asked to report the cardinality represented by the abacus using a numeric keypad. The task was adaptive in the number of abacus columns in the pictured quantity: if participants gave a correct answer on two consecutive trials, an extra column was added to the next trial; if they were incorrect on one trial, a column was subtracted. Participants were given feedback after each trial. There was no time limit.

Each participant also performed one of five estimation tasks (Figure 3a). In each, participants reported the number of dots on the screen. Tasks were (1) Identical: abacus flashcards identical to those in the reading task (N=24), (2) Rotated: mirror images of abacus flashcards rotated 90 degrees (N=24), (3) Configural: abacus flashcards with beam and rod structures removed but configuration of beads preserved (N=36), (4) Jittered: random dot arrays jittered within the bounding box that the beads in the corresponding abacus flashcard would have occupied plus a small constant (N=25), or (5) Random: random dot arrays (N=24). Tasks were counterbalanced for order, and each estimation task was adaptive according to the same distribution of trials as the flashcard task.

Results and Discussion. Consistent with Experiment 1, abacus reading accuracy was better predicted by a model including the number of columns on the abacus, not a model with the number of beads in play (χ^2 = 815.89, p < .0001, Figure 2b). (Abacus reading accuracy was almost identical across the five groups of children, M=.72, .72, .73, .74, and .73, respectively).

We next fit a single multi-level model to the entire dataset (all trials in all conditions), with group-level effects for each condition and the interaction of condition with number of beads in the trial. Table 1 gives the coefficient estimates and z value approximations for the group-level effects in the model.[1] There was some gain in accuracy for the spatially grouped, rectilinear conditions. Nevertheless, abacus reading exhibited greater accuracy than any of the five estimation tasks (p < .0001 for all comparisons, Figure 3b).

We fit a similar model to participants' reaction time to first key-press (Table 2). Since reaction time data are log-normally distributed, we used a linear regression to predict log reaction time. Reaction times greater than three standard deviations above the mean (1.9% of all data) were omitted. Because reaction times tend to decrease over the course of an experiment, we added a coefficient for trial-number. Abacus reading was faster than any estimation task and reaction times for abacus reading grew more slowly as a function of the number of beads in the display than for any estimation task, suggesting a qualitative difference between abacus reading and quantity estimation (p < .0001 for all comparisons, Figure 3c).

---

[1] All p values are derived from the z approximation; although it can be anti-conservative for small datasets, that this anti-conservatism is minimal for the large dataset we used (Pinheiro & Bates, 2000).

Finally, we calculated the coefficient of variation (COV; standard deviation of estimates / quantity being estimated) for each quantity, participant, and task (Figure 4). For each task we included data for all quantities above six for which there were at least five participants who produced data at that quantity. We calculated the COV for the abacus flashcard task by converting responses to the number of beads in the quantity reported. The estimation tasks and the abacus flashcard task all showed a COV that rose over small quantities and then held steady for larger quantities; however, COV for the abacus flashcard task was significantly lower. We constructed a linear mixed model of each participant's COV across tasks and used posterior simulation to calculate empirical *p* values. We found significant positive coefficients for each estimation task relative to abacus reading (all *p*s < .0001).

Experiment 2 confirmed all three predictions of the perceptual redeployment hypothesis, but did not support perceptual sharpening or chunking. Performance in abacus reading was governed by columns, rather than beads; abacus reading had a different RT slope than estimation tasks (even those making use of identical stimuli); and abacus reading had higher accuracy and a lower coefficient of variation than estimation tasks.

Experiment 3

Our final experiment tested untrained adults on abacus reading. We gave participants a brief tutorial on the abacus and then asked them to perform the reading task as well as all five estimation tasks from Experiment 2. Both the sharpening and chunking hypotheses predicted that participants' performance across tasks would be poor. The sharpening hypothesis predicted that untrained participants' COV in abacus reading should be comparable to their COV in estimation (and hence both should likely be higher than MA-trained individuals, even children). Similarly, evidence that expert children had lower COVs than adults would support this hypothesis. The chunking hypothesis predicted that differences in estimation performance due to abacus configural properties should be minimized due to limited exposure to this type of stimuli.

The redeployment hypothesis made contrasting predictions. First, it predicted that although untrained participants' performance should be poor for abacus reading, their performance on estimation tasks should exhibit a similar pattern across display types. Second, it predicted that untrained participants' COV in the estimation tasks should not be comparable to their COV in the abacus reading task (which should likely be higher due to their inexperience with MA technique). Third, it predicted that, once trained to assign place values to columns, the non-experts' abacus reading performance should be predicted by columns, not beads (leading to the same RT slope difference as the MA children).

Methods

Participants. Thirty UCSD undergraduates with no prior abacus experience participated for course credit. Due to experimenter error, one estimation task from three participants was not included in the analysis.

Stimuli and Procedure. Stimuli and procedures were identical to Experiment 2, but all participants were tested in all six conditions in one of two random orders. Prior to testing, each participant completed a two-page abacus training worksheet, teaching them how to read abacus representations and giving them practice with twelve flashcards.

Results and Discussion. We analyzed data using mixed models identical to those in Experiment 2 (Tables 1 and 2, Figure 3). The ordering of accuracies for the five estimation tasks was highly similar for the MA-trained and untrained groups. In both, jittered and random estimation tasks grouped together; these two tasks were more difficult than the three configural conditions, which also grouped with one another.

As expected by all accounts, untrained participants were far less proficient on the abacus reading task. Nevertheless, like MA experts, the untrained adults showed a lower reaction time slope for abacus reading than for any other task (all *p*s < .0001 for comparisons between coefficients), even though their intercept was significantly higher than for all other conditions (all *p*s < .0001). This higher intercept likely reflects a greater constant cost for conversion of abacus representations to Arabic numerals in the non-expert group.

Although accuracy data were noisier in the untrained adults, results resembled those for MA children: the column-based model fit far better than the bead-based model (χ^2 = 128.36, $p < .0001$). Even in the absence of extensive MA training, untrained adults grouped the abacus displays into columns, likely leading to the lower reaction time slope for abacus reading as well.

Finally, we examined the individual participants' COV for each condition (Figure 4). We again constructed a linear mixed model of adults' COV and found the opposite pattern of results from Experiment 2: participants' COVs in the abacus reading task were significantly higher than their COVs in the estimation tasks (all *p*s < .001). Despite their prowess with MA, the trained children still had higher COVs as a group than the untrained adults, suggesting that it is not improvement in COV (as posited by the perceptual sharpening hypothesis) that leads to MA proficiency. Without longitudinal data we cannot rule out some improvement in the COV of participants over MA training, but this improvement is not responsible for the children's expert performance.

Results confirmed all three predictions of the redeployment hypothesis, but did not support perceptual sharpening or chunking. First, better performance for stimuli that are configurally similar to an abacus (relative to random dot arrays) does not result from MA training. Second, even untrained participants with only a small amount of exposure to abacus reading represent MA via a column-based strategy (leading to a flatter reaction time slope). Third, increased performance in abacus tasks is not due to improvement in participants' COV.

General Discussion

Our data indicate that children can represent large exact numerosities by redeploying existing representations in visual working memory. We contrasted three hypotheses about the nature of mental abacus representations: (1) perceptual sharpening, in which MA expertise involves sharpening the acuity of approximate number (ANS) representations; (2) perceptual chunking, by which MA expertise involves memorizing an assortment of specific MA configurations; and (3) perceptual redeployment, by which MA is a strategy for making use of pre-existing visual representations. The effects we report—the relationship between the number of abacus columns and participants' performance, the dissociation between COV for abacus reading and for estimation tasks, the flatter reaction time slopes for abacus reading, and the presence of these signatures even in untrained adults—all suggest that mental abacus representations are formed by the redeployment of pre-existing visual resources.

How does the organization of the visual system support this redeployment? It appears that separate columns in the abacus representation are stored as parallel "models" in visual short term or working memory. Consistent with prior work, limits on the number of models that can be represented may be somewhat flexible depending on task and complexity of information being stored (Alvarez & Cavanagh, 2004). Each model seems to store some information about the "ensemble statistics" of its contents (Alvarez & Oliva, 2008, 2009; Ariely, 2001; Feigenson, 2008; Halberda et al., 2006). In MA including the quantity of objects in a model and their mean positions would be sufficient to reconstruct the digit the model represented. Although some previous work has found evidence for color as a grouping cue (Halberda et al., 2006), any pre-attentive grouping cue—position, in the case of the abacus—is likely sufficient to allow a set of objects to be captured by attention and stored as a model. The precision of such models declines sharply for sets of more than 4, offering a possible account of why MA emerged from training with the Soroban abacus, rather than from other abacus types that use more beads.

As illustrated by the poorer performance of our untrained group, accurate storage of the abacus image in visual working memory is not the only necessary ingredient of MA. Even to read flashcards proficiently, users must be practiced with the translation of columns into Arabic numerals; to carry out rapid mental addition, they must also practice routines for adding columns. Nevertheless, at the core of MA is an ability to store abacus representations precisely—a function for which the visual system is already equipped.

Although our results suggest that humans can mentally represent large exact numerosities using existing visual representations, they do not address whether typical mathematical development involves the acquisition of new resources. They also leave open the question of whether MA could be acquired by non-humans animals or humans who lack experience with counting. Nonetheless, the system of representation used in our tasks was non-verbal and bore the hallmarks of visual working memory and the ANS (see also Hatano, 1977, 1983; Hishitani, 1990; Stigler, 1984). Although language may play a role in learning MA, visual resources appear sufficient to represent large exact numerosities once MA is acquired.

This proposal offers an alternative to traditional views of expertise as arising either from improvement of a single capacity or from the synthesis of wholly novel representations. In contrast, mental abacus provides a compelling example of how complex representations—like a precise system for representing number—can be built out of simpler, pre-existing cognitive resources.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106-111.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, 106*, 7345.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 157-162.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*, 487-502.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Cordes, S., Gallistel, C. R., & Gelman, R. (2007). Nonverbal arithmetic in humans: Light from noise. *Perception and Psychophysics, 69*, 1185.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence. *Science, 284*, 970.

Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition, 107*, 1-18.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307-314.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition, 108*, 819-824.

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science, 306*, 496-499.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534-537.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17*, 572-576.

Hatano, G. (1977). Performance of Expert Abacus Operators. *Cognition, 5*, 47-55.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition, 15*, 95-110.

Hishitani, S. (1990). Imagery experts: How do expert abacus operators process imagery? *Applied Cognitive Psychology, 4*.

Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology, 52*, 130-169.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology, 111*, 1-22.

Menninger, K. (1969). *Number Words and Number Symbols: A Cultural History of Numbers*. Cambridge, MA: MIT Press.

Miller, K., & Stigler, J. (1991). Meanings of skill: Effects of *abacus* expertise on number representation. *Cognition and Instruction, 81*, 29–67

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences, 105*, 2745.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science, 306*, 499-503.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for research in mathematics education*, 498-505.

Stigler, J. W. (1984). Mental abacus: the effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology, 16*, 145-176.

Stigler, J., Chalip, L., & Miller, K. (1986). Consequences of skill: The case of *abacus* training in Taiwan. *American Journal of Education, 94*, 447–479.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science, 10*, 130-137.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*, 155-193.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1-B11.

Table 1. Coefficient weights for logistic mixed models of accuracy in Experiments 2 (MA children) and 3 (adults). ":" indicates interactions.

Predictor	**E2 Coef. (Std. Error)**	***z* value**	**E3 Coef. (Std. Error)**	***z* value**
Abacus	2.80 (0.07)	41.03	3.24 (0.21)	15.34
Identical	3.54 (0.19)	18.24	3.35 (0.20)	16.82
Rotated	3.92 (0.19)	20.40	3.84 (0.20)	18.83
Configural	4.10 (0.15)	27.63	4.70 (0.23)	20.63
Jittered	5.66 (0.24)	23.88	5.29 (0.23)	23.25
Estimation	5.52 (0.24)	22.87	5.17 (0.23)	22.50
Abacus:Beads	-0.21 (0.01)	-28.16	-0.36 (0.03)	-14.48
Identical:Beads	-0.31 (0.02)	-14.48	-0.29 (0.02)	-13.96
Rotated:Beads	-0.36 (0.02)	-16.94	-0.33 (0.02)	-15.81
Configural:Beads	-0.41 (0.02)	-22.98	-0.40 (0.02)	-17.79
Jittered:Beads	-0.65 (0.03)	-21.61	-0.55 (0.02)	-22.39
Estimation:Beads	-0.63 (0.03)	-20.98	-0.53 (0.03)	-21.43

Table 2. Coefficient weights for linear mixed model analyses of log reaction time in Experiment 2 (MA children) and Experiment 3 (adults).

Predictor	**E2 Coef. (Std. Error)**	***t* value**	**E3 Coef. (Std. Error)**	***t* value**
Abacus	0.28 (0.02)	11.55	0.40 (0.06)	6.85
Identical	0.37 (0.04)	9.70	-0.43 (0.06)	-7.34
Rotated	0.21 (0.04)	5.63	-0.58 (0.06)	-9.85
Configural	-0.10 (0.03)	-3.38	-0.61 (0.06)	-10.39
Jittered	-0.15 (0.03)	-4.44	-0.79 (0.06)	-13.57
Estimation	-0.11 (0.04)	-2.98	-0.65 (0.06)	-11.16
Abacus:Beads	0.06 (0.002)	34.01	0.08 (0.007)	11.50
Identical:Beads	0.13 (0.004)	31.16	0.14 (0.006)	22.51
Rotated:Beads	0.12 (0.004)	32.16	0.14 (0.006)	22.90
Configural:Beads	0.14 (0.003)	48.18	0.14 (0.006)	23.26
Jittered:Beads	0.13 (0.004)	36.72	0.14 (0.006)	23.90
Estimation:Beads	0.15 (0.001)	39.79	0.14 (0.006)	23.52
Trial number	-0.0020 (0.0001)	-14.740	-0.001 (0.0001)	-12.731

Figure Legends

Figure 1. (a) A Japanese Soroban abacus of the type used by our participants. The right-most nine rows show the number 123,456,789. (b) The proposed structure of a mental abacus representation for the quantity 49.

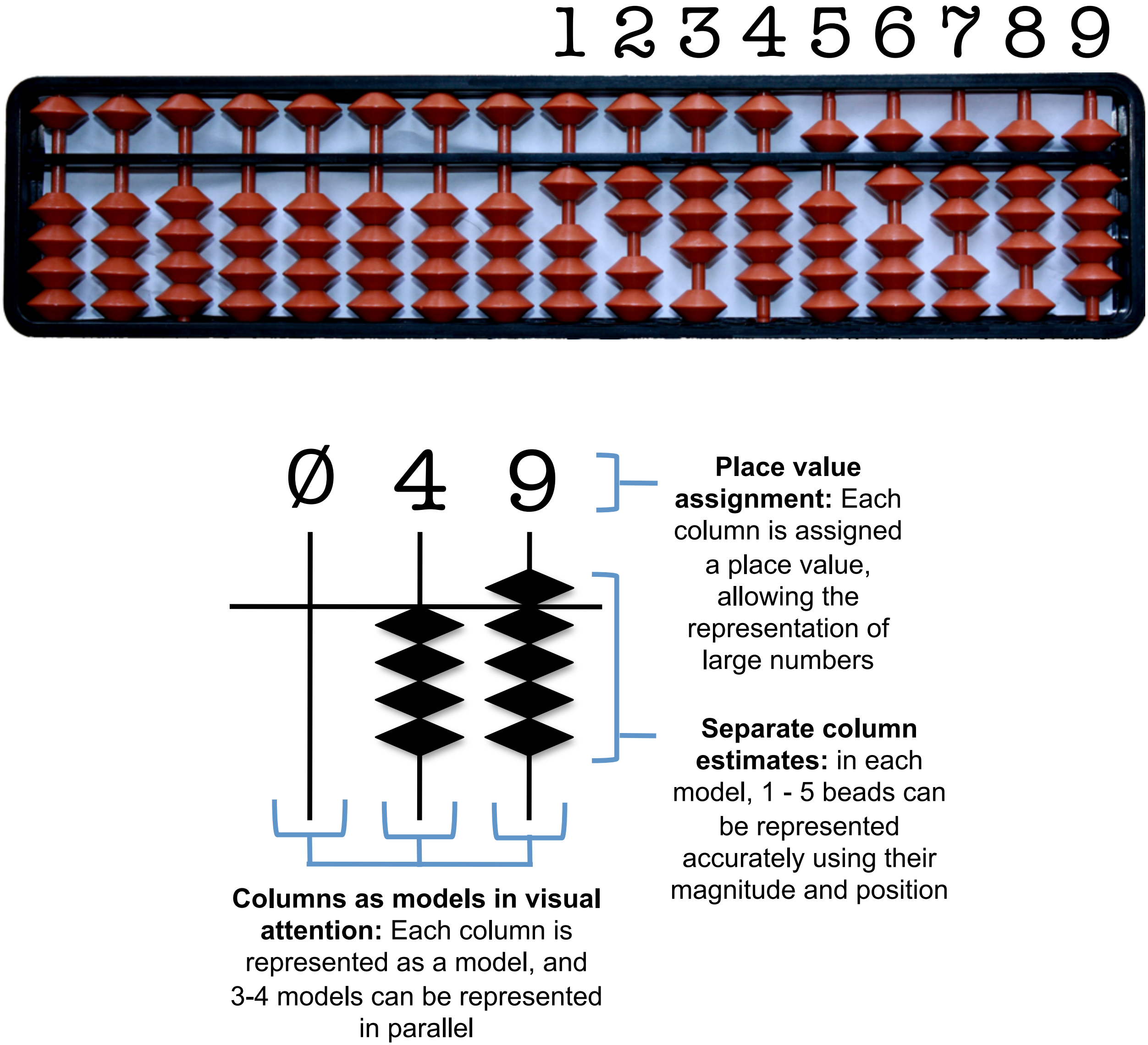
Figure 2. MA participants' probability of correct response for Experiments (a) 1, (b) 2, and (c) 3. Left-hand panels show probability of a correct response plotted by the number of beads in the display; right-hand panels show the same probability plotted by the number of columns in the display. Colors in all panels are used to denote the number of columns in a display.

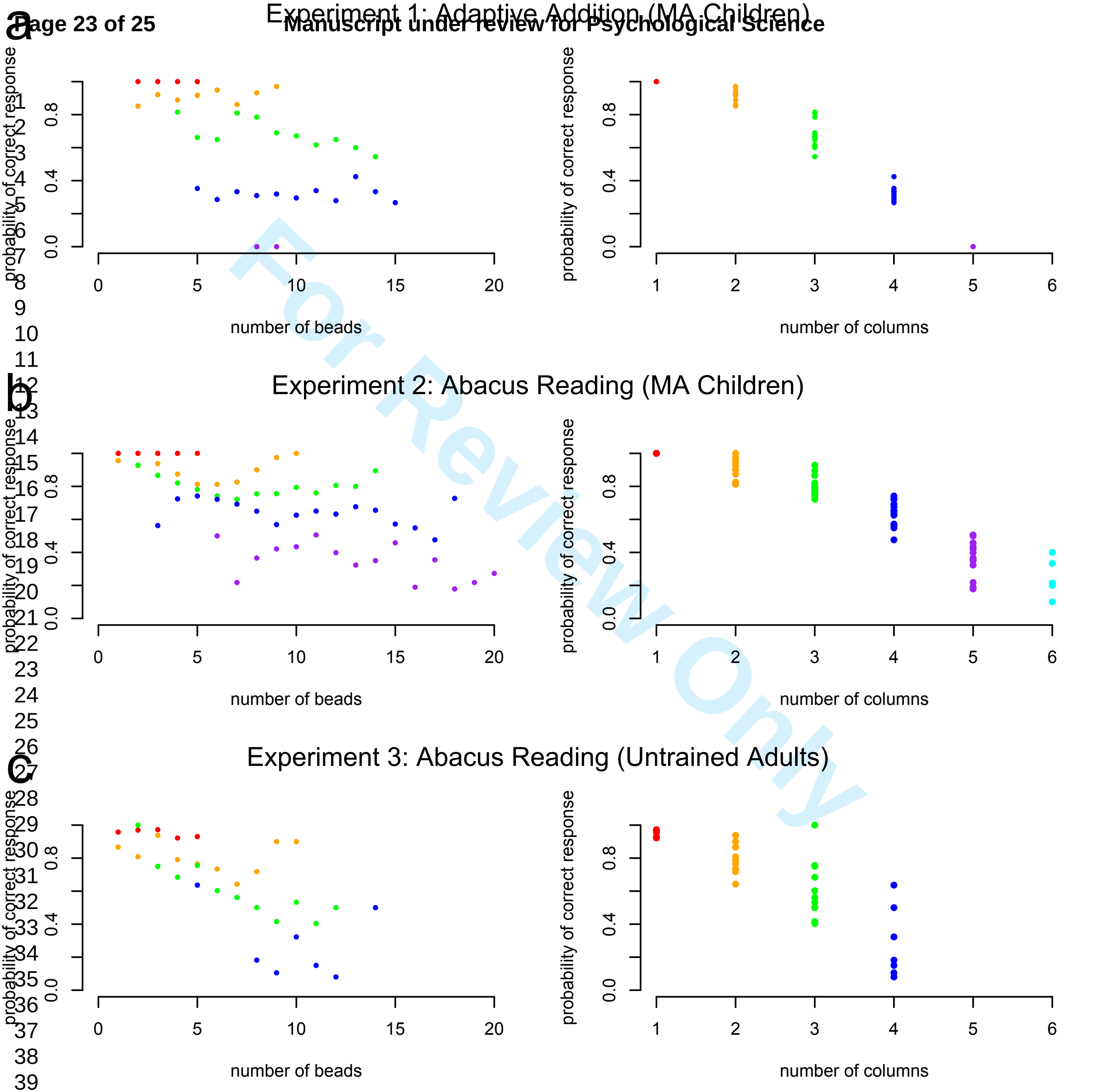
Figure 3. (a) Displays used in Experiments 2 and 3. The bounding color of each display represents the color of the dots and regression lines in b-e. (b) Accuracy data for Experiment 2. Probability of a correct response plotted by the number of abacus beads in the correct response for flashcard tasks. Size of dots reflects the proportion of participants with a given mean performance; lines reflect best-fit logistic curves for a mixed logistic regression model. (c) Log reaction time for Experiment 2. Regression lines are for a mixed linear regression model. (d) Accuracy data for adults in Experiment 3. (e) Log reaction time data for Experiment 3.

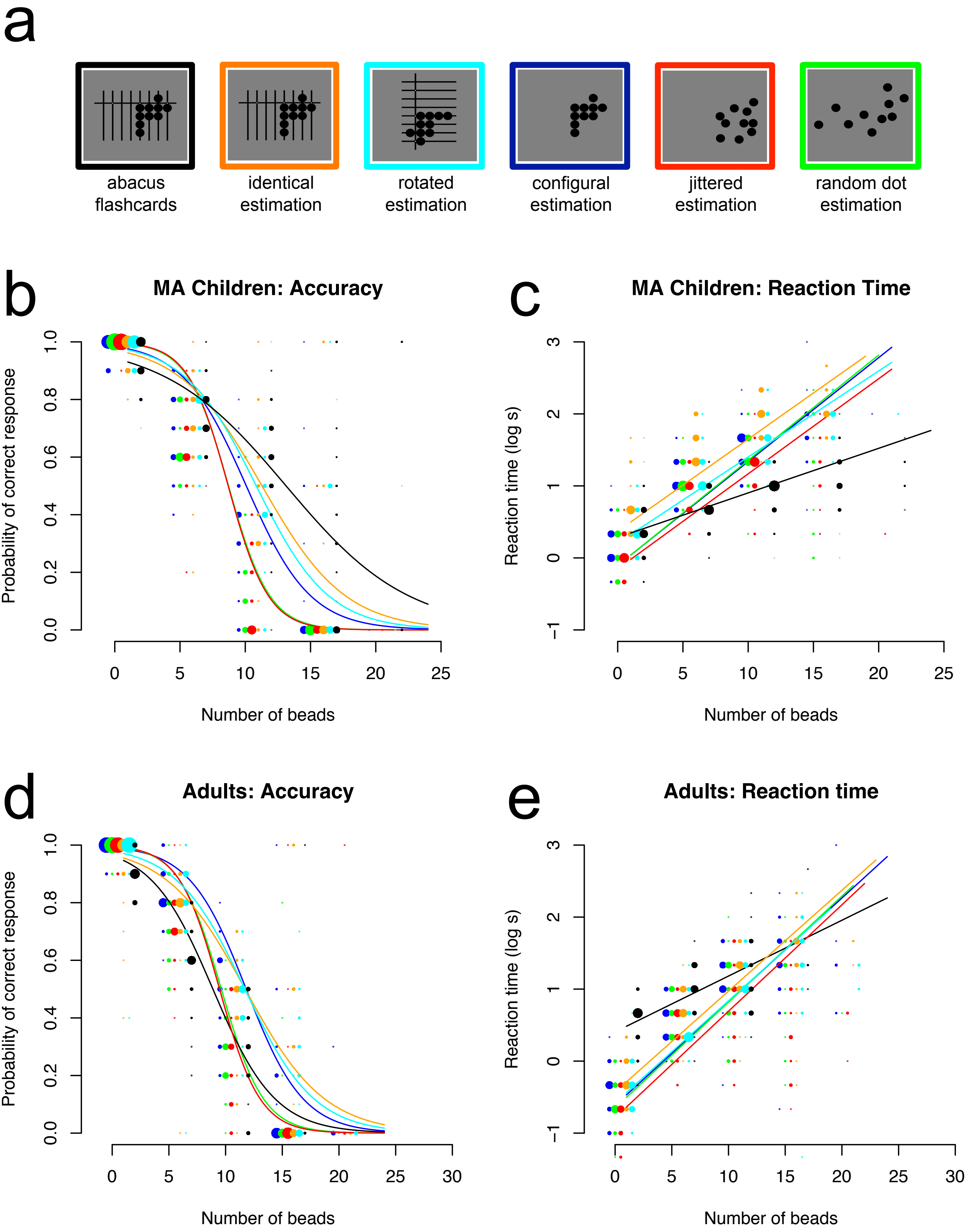
Figure 4. Mean COV for each of the six tasks in Experiments 2 and 3, calculated across participants. Error bars show standard error of the mean; they vary in size because all MA children completed the flashcard task while only a subset performed each of the estimation tasks.

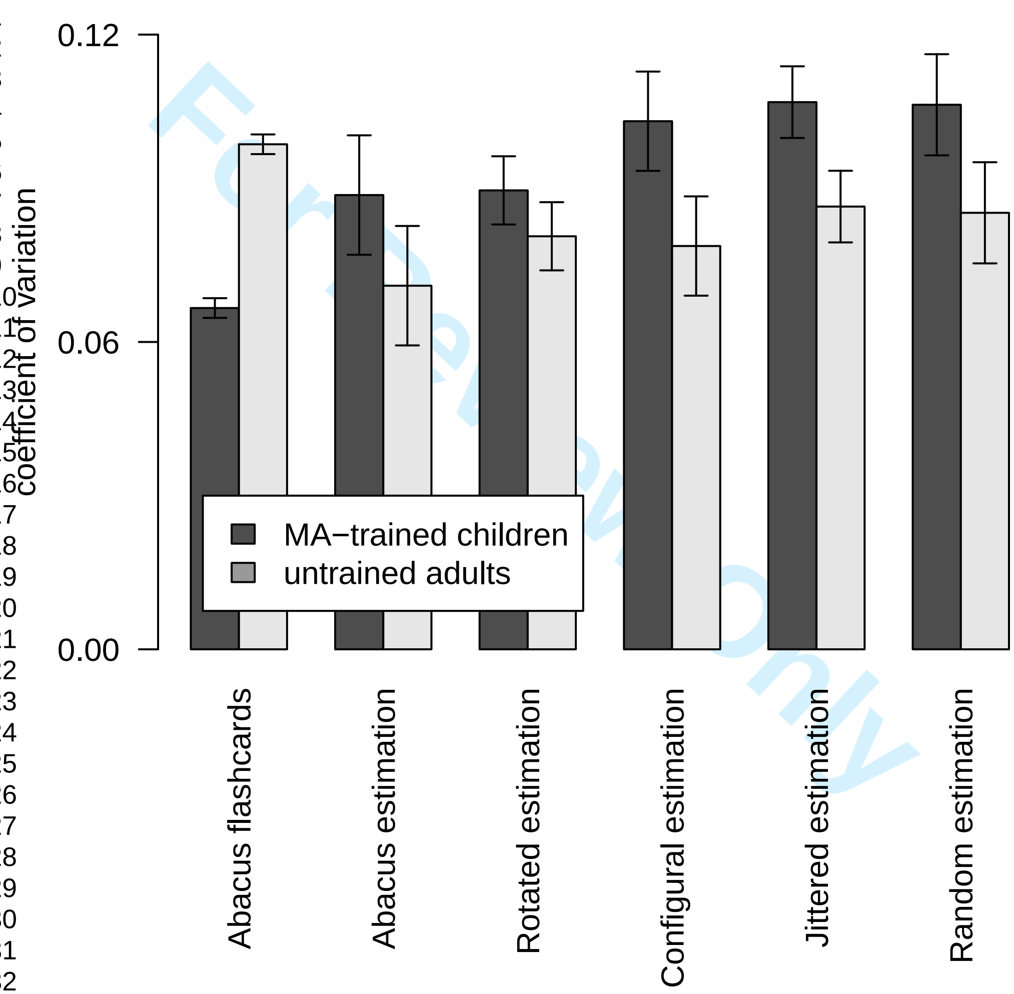
a

b


1 2 3 4 5 6 7 8 9
Ø 4 9
Place value assignment: Each column is assigned a place value, allowing the representation of large numbers
Separate column estimates: in each model, 1 - 5 beads can be represented accurately using their magnitude and position
Columns as models in visual attention: Each column is represented as a model, and 3-4 models can be represented in parallel


a
Experiment 1: Adaptive Addition (MA Children)
probability of correct response
0.0
0.4
0.8
0
5
10
15
20
number of beads
probability of correct response
0.0
0.4
0.8
1
2
3
4
5
6
number of columns
b
Experiment 2: Abacus Reading (MA Children)
probability of correct response
0.0
0.4
0.8
0
5
10
15
20
number of beads
probability of correct response
0.0
0.4
0.8
1
2
3
4
5
6
number of columns
c
Experiment 3: Abacus Reading (Untrained Adults)
probability of correct response
0.0
0.4
0.8
0
5
10
15
20
number of beads
probability of correct response
0.0
0.4
0.8
1
2
3
4
5
6
number of columns


a
abacus flashcards
identical estimation
rotated estimation
configural estimation
jittered estimation
random dot estimation
b
MA Children: Accuracy
Probability of correct response
Number of beads
c
MA Children: Reaction Time
Reaction time (log s)
Number of beads
d
Adults: Accuracy
Probability of correct response
Number of beads
e
Adults: Reaction time
Reaction time (log s)
Number of beads


0.12
0.06
0.00
coefficient of variation
MA−trained children
untrained adults
Abacus flashcards
Abacus estimation
Rotated estimation
Configural estimation
Jittered estimation
Random estimation