# Representing exact number visually using mental abacus

Michael C. Frank

Department of Psychology, Stanford University

David Barner

Department of Psychology, University of California, San Diego

The two authors contributed equally to this work. We gratefully acknowledge the students and staff at UCMAS Gujarat, and give special thanks to Snehal Karia and his family. Thanks also to George Alvarez, Timothy Brady, Neon Brooks, Susan Carey, Steven Franconeri, Lisa Feigenson, Justin Halberda, Rebecca Saxe, and Jessica Sullivan for comments on this manuscript. This project supported by a UCSD Senate Award and NSF #0910206 to DB, and a Jacob Javits Graduate Fellowship and NSF DDRIG #0746251 to MCF.

Please address correspondence to: Michael C. Frank, Department of Psychology, Stanford University, 450 Serra Mall, Building 420 (Jordan Hall), Stanford, CA 94305, tel: (650) 724-4003, email: mcfrank@stanford.edu.

**Abstract**

Mental abacus (MA) is a system for performing rapid and precise arithmetic by manipulating a mental representation of an abacus, a physical calculation device. Previous work has speculated that MA is based on visual imagery, suggesting that it might be a method of representing exact number non-linguistically, but—given known limitations on visual working memory—it is unknown how MA structures could be stored. We investigated the structure of the representations underlying MA in a group of children in India. Our results suggest that MA is represented in visual working memory by splitting the abacus into a series of columns, each of which is independently stored as a unit with its own detailed substructure. In addition, we show that the computations of practiced MA users (but not those of control participants) are relatively insensitive to verbal interference, consistent with the hypothesis that MA is a non-linguistic format for exact numerical computation.

## Introduction

Human adults, unlike other animals, have the capacity to 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 (Feigenson, Dehaene, & Spelke, 2004). Multiple forms of evidence suggest that this human capacity is related to natural language (Barner, Chow, & Yang, 2009; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Frank, Everett, Fedorenko, & Gibson, 2008; Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004; Wynn, 1990). Language, however, may not be the sole cognitive system capable of symbolically representing exact number. Experienced users of an abacus—a physical calculation device—can learn to perform arithmetic computations mentally, as though visualizing a "mental abacus" (MA) (Hatano, 1977; Hatano & Osawa, 1983; Hishitani, 1990; Stigler, 1984; Stigler, Chalip, & Miller, 1986; Miller & Stigler, 1991).

Previous work, reviewed below, has described the MA phenomenon and has provided suggestive evidence that MA is represented non-linguistically, in a visual format. However, this proposal remains tentative for two reasons. First, early studies that directly tested the role of language in MA were compelling but imperfect, and used sometimes informal methods to test small and unusual populations of participants. Second, previous proposals fail to explain how MA could be represented in a visual format. The present study addressed these issues in a series of three experiments. We conducted detailed studies of MA processing to ask how it might be represented in visual working memory, given known limitations on the non-linguistic processing of quantity information. In addition, we used a dual-task paradigm to test the role of language in MA computations. Taken together, our results support the view that MA relies on visual resources, and in particular the ability to represent multiple groupings of objects in parallel, to create visual representations of

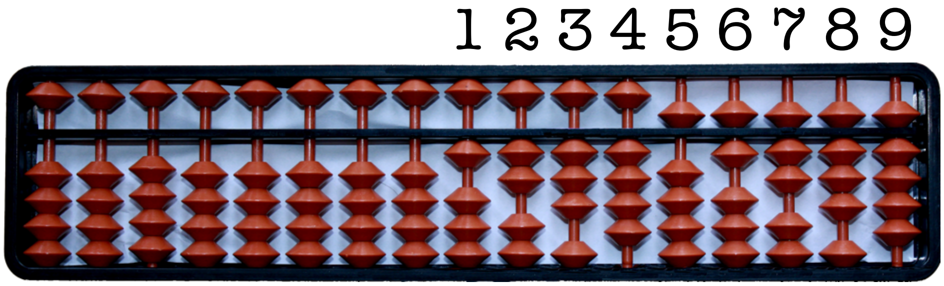


*Figure 1.* A Japanese soroban abacus of the type used by our participants. The rightmost nine columns represent the number 123,456,789.

exact number that differ fundamentally from those constructed using natural language.

*Background and previous work on MA*

The abacus has been used in Asia since 1200 for rapid precise calculation, and may have emerged from earlier Roman counting boards, which bear a similar structure (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 1). On a Japanese soroban abacus—the most commonly used type of 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. Other varieties of abacus represent number similarly, but with interesting differences: one variety of the Chinese suanpan has five bottom beads and two top beads on each column, allowing for both decimal and hexadecimal computation, while the Russian schoty (similar in appearance to the "school abacus" in the United States) is organized into rows of ten beads, color coded into a sets of 4, 2, and 4 on each row.

In addition to using the physical device, MA users are trained to visualize an abacus and to move imagined beads on this abacus in order to perform arithmetic calculations. Many users appear to move these imagined beads using their hands, and thus move their hands in the air as they perform calculations, suggesting that motor representations somehow interface with the number representations created in MA. MA is commonly used for calculations like addition and subtraction, but with practice, users can also learn routines to perform multiplication and division or even square and cube roots. Because of its incredible speed and accuracy, MA compares favorably to other methods of computation, including electronic calculators (Kojima, 1954) and alternative systems of mental arithmetic. For example, the 2010 Mental Computation World Cup was won by an 11-year-old MA user. For examples of mental and physical abacus use and an example of a participant in Experiment 2 discussing the MA procedure, see supplementary movies S1–3.[1]

Although abacus instruction is conducted verbally and begins after children learn to count, previous studies argue that MA representations are not linguistic in nature but rely on visual mechanisms (Hatano, 1977; Hatano & Osawa, 1983; Hishitani, 1990). For example, Hatano (1977) investigated how the calculation abilities of expert MA users were affected by concurrent verbal, spatial, and motor interference tasks. Consistent with his hypothesis, Hatano found that MA users could perform difficult arithmetic problems while doing concurrent tasks. However, the strength of these findings is limited, because (1) Hatano tested only a small group of MA grand masters and (2) the interference tasks were somewhat informal in nature. For example, the verbal interference task consisted of answering basic factual questions while completing addition problems, potentially allowing participants to switch rapidly between tasks during the course of the experiment. Follow-up studies tested the digit-span capacity of three national champions in mental

[1] Available at `http://langcog.stanford.edu/materials/abacus.html`.

calculation (Hatano & Osawa, 1983) and a developmental sample of intermediates and experts (Hatano, Amaiwa, & Shimizu, 1987; Lee, Lu, & Ko, 2007) and showed that experienced MA users can effectively store long strings of digits with greater accuracy than long strings of verbal material, presumably by remembering these strings as abacus images.

Echoing Hatano, subsequent studies have also reported differences in how MA users represent number. First, studies using fMRI have found different processing signatures for MA and verbal arithmetic. When asked to recall a long string of digits or do complex arithmetic tasks, MA users show selective activation of cortical areas associated with vision and visuo-spatial working memory. In contrast, untrained controls exhibit patterns of activation related to verbal processing and verbal working memory (Tanaka, Michimata, Kaminaga, Honda, & Sadato, 2002; Chen et al., 2006; Hu et al., in press). Second, according to Stigler and colleagues, there is a close correspondence between what MA users "see" in their minds eye and the structure of the physical device (Stigler, 1984; Stigler et al., 1986). They reported that MA users are far more likely to make calculation errors involving quantities of 5 (due to misrepresentation of "heavenly" 5 beads) than control participants, who make these errors less than a quarter as often. Also, they found that MA users were able to access intermediate states in calculations that are unique to abacus (e.g., when adding 5+3, the abacus passes through states representing 5, 6, 7, and 8 as each bead is moved). When shown a card depicting an abacus state, subjects could identify whether this state appeared in a subsequent mental addition problem, and did so with the same accuracy as when doing problems on a physical device. This result suggests that participants' mental abacuses pass through the same set of states as the physical device does, and together this previous work suggests that MA representations are structured like a physical abacus.

*The nature of MA representations*

Stigler (1984) and Hatano (1977) both argued that MA relies on a non-linguistic, visual representation of an abacus, but little is known about how such representations could be implemented by the visual system. Consider an MA representation of the number 49: representing this quantity requires tracking the precise location of nine beads. Representing the identity and precise location of each bead is critical not only for identifying states of the abacus (e.g., reading off values), but also for performing arithmetic computations. For example, when adding 49 + 30, it is not enough to know that four beads are present in the 10s column. A MA user must also know which four beads are in play (the bottom four) in order to select the correct motion that will transform these appropriately when the quantity 3 is to be added.[2] Thus, a user of MA must represent the location and position of each bead in the current state of the abacus to perform basic addition tasks successfully.

It is a puzzle how such states can be represented, given what is known about the processing of quantity information in the visual system. Previous studies indicate that visual working memory can represent both the location and identity of three to four items, but not more (Alvarez & Cavanagh, 2004; Cowan, 2000; Feigenson et al., 2004; Luck & Vogel, 1997). Thus, this system is insufficient for representing anything but the smallest quantities in MA. The approximate cardinality of large sets can also 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). The ANS does not track the location of individual objects, however, and although the ANS exhibits relatively little error for small sets (e.g., with fewer than 4–5 members), it can only represent the cardinality of large sets approximately. Since

[2] In this case, the user would have to subtract 2 from the "earthly" (1) beads and add the "heavenly" (5) bead, to represent the total quantity 7 in the 10s place.

representing the quantity 49 requires keeping track of the locations of each of 9 different abacus beads (and simply maintaining the information that there are 9 and not 10), it would appear that neither of these non-linguistic systems could alone represent the structure of an MA.

Because there is no obvious answer to how MA representations are constructed in the visual system, it is tempting to conclude that each column is represented by a symbol that is unconnected to the underlying semantics of the physical abacus. On this kind of account, the picture of a column with four earth beads in play is equivalent to the Arabic numeral 4: both are an abstract representation of a particular quantity that can be composed to create larger numbers like 40 or 400. In addition to the findings that we present in this study, several facts speak against this. First, Arabic numerals and MA representations are defined differently. The Arabic numeral 4 has no internal structure—nothing that says that the symbol "4" should not stand for 5 objects and the symbol "5" stand for 4, for example. In contrast, the MA representation of four gains its numeric value because of a set of rules that also defines the MA representations of other quantities. Representing the internal structure of columns in MA is necessary for supporting arithmetic computations like addition and subtraction, because these computations rely on moving individual beads. Second, as reviewed above, MA users make errors that are consistent with access to intermediate states in the abacus calculation—intermediate states that could only be available if they were representing the substructure of abacus columns. Third, MA is often tightly linked to gesture (a striking part of the MA phenomenon for observers). These movements correspond to moves on the abacus and appear to facilitate the movements of individual beads in the mental image (an observation supported by the motor interference results shown by Hatano, 1977 and in Experiment 2). Thus, the evidence does not support a view of MA representations as unanalyzed wholes.

Instead, recent work on visual working memory suggests a possible mechanism by which abacus representations might circumvent the limits of known number representation systems. According to these reports, subjects can select and represent up to three or four sets of objects in parallel (Feigenson, 2008; Halberda, Sires, & Feigenson, 2006). These sets can then be manipulated in different ways. For example, in one study subjects saw arrays that contained spatially overlapping sets of dots of different colors, and were probed to estimate the number of items for a particular color after the array disappeared (Halberda et al., 2006). When the number of sets was 3 or fewer, subjects were able to estimate the quantity of the probed set with relative accuracy, and showed signs of using the ANS. However, when 4 or more sets were presented they failed to make reliable estimates. In another study, subjects watched as different kinds of objects (e.g., candies, batteries, toy pigs) were placed into a container, while they performed a concurrent verbal interference task that prevented them from counting. Here again, subjects could perform reliable estimates when 3 or fewer kinds of things were involved, but failed when they were required to keep track of 4 or more sets at a time (Feigenson, 2008). Together, these studies suggest that normal adults can represent multiple sets in parallel using visual working memory, and can perform numerical estimates on these sets.

Supporting this view, some work suggests that objects contained in multiple sets can be tracked individually, so long as there are no more than 3–4 objects in each set. For example, Feigenson and Halberda (2008) showed that young children can represent and compare two sets of objects, binding property information to the objects in each set and tracking their locations over time. In addition, in those studies, infants' ability to track objects improved when arrays were first presented as smaller subsets divided in space, suggesting that spatial grouping cues could facilitate object tracking. Consistent with this, studies of adult visual attention find that subjects are signficantly better at attentional tracking when targets are divided across the two visual hemifields (Alvarez & Cavanagh,

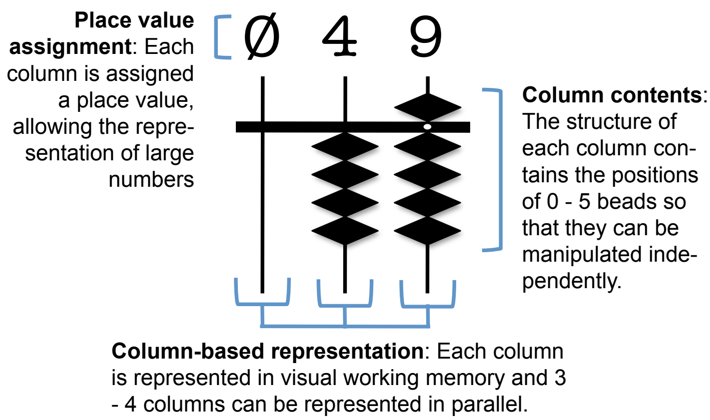


*Figure 2.* A schematic proposal for a mental abacus (MA) representation of the number 49.

2005). By organizing sets into horizontally segregated arrays, much like the abacus, subjects can optimize the number of objects they are able to track in parallel.

In keeping with these findings, Figure 2 shows a schematic proposal for how MA might represent a number like 49.[3] By treating each column of the MA as a separate set in visual working memory, users could track the locations of beads in up to three or four columns in parallel. The main studies of parallel set representation have investigated the approximate quantities represented in each set (Halberda et al., 2006; Feigenson, 2008). Nevertheless, we do not believe that the information represented about an individual column is restricted to the approximate quantity of beads present: instead, column representations must contain information about the precise quantity and locations of the beads in the column (we return to the issue of the relationship between MA and approximate number representations in the general discussion). Thus, recent work lends

[3]It is conventional in drawings of abacus representations only to represent those beads that are "in play"—thus an MA image of the number 10 involves imagining an abacus with only one bead, while an MA image of 49 involves nine beads.

plausibility to the idea that MA uses existing visual resources to store multiple, internally-structured set representations in parallel, in order to represent large exact numerosities.

*The current studies*

We explored the proposal described above—that MA representations are column-based models in visual working memory—in a series of three experiments. The goal of the studies was not to compare this hypothesis to an existing alternative, since no viable alternative hypotheses exist in the literature. Instead, our studies were exploratory in nature, testing the plausibility of the view that MA is a non-linguistic representation of number that uses existing visual resources to perform exact arithmetic computations.

To do this, we tested a population of children in Gujarat Province, India, where MA is taught in a 3-year after-school program. Because of the effectiveness of MA for arithmetic calculation—a critical component of standardized tests in the Indian educational system—MA courses have experienced huge growth in India in the past decade. Many children from throughout Gujarat province and the rest of India compete in regional, national, and international abacus competitions using both MA and physical abacus. This situation has created a large student population within which to study MA. Our studies examined both highly practiced users of MA (Experiment 2) and also children who were randomly sampled from the larger student population (Experiments 1 and 3).

Experiment 1 asked children studying MA to perform challenging addition problems in order to test the limits on MA addition and their relationship to limits on visual working memory. According to our hypothesis—that abacus columns are stored as sets in visual working memory—MA users should show limits on the number of columns they can compute over. The results of Experiment 1 are congruent with this prediction: MA is sharply limited by the number of digits in each addend—a limit that corresponds to the capacity of visual working memory (≈3–4 digits). However, there appears to be no hard limit on the number of distinct addends children can add, suggesting that the total number of computations in a problem cannot explain its difficulty.

Experiment 2 then follows up on Hatano's early work by using a variant of the adaptive addition paradigm of Experiment 1 to investigate the effects of verbal and motor interference on both MA users and untrained adults. The goal of this study was to determine the relative role of language in MA computations. Our results suggest that while language interference has some effect on MA calculation, the effect of motor interference was approximately equivalent, and most participants were still able to perform extremely well on difficult addition problems under interference. This finding is in contrast to the large effects of verbal interference on untrained control participants, for whom motor interference had no effect on computation.

Experiment 3 investigated the behavior of MA users and untrained control participants on a final task: translating a picture of an abacus to Arabic numerals ("abacus flashcards"). This study provides a second, independent test of the column limit found in Experiment 1. Also, it tests whether the encoding of visual arrays in an unrelated task—estimation—is facilitated when arrays become more abacus-like in structure. The results suggest that untrained control participants perform in ways that are remarkably similar to MA users, giving evidence that MA expertise does not fundamentally alter the method of representation of the abacus image. Instead, based on these results, we conclude that MA representations are optimally designed to exploit pre-existing visual representations.

These studies make three primary contributions. First, our studies suggest that MA representations are supported by the recently-discovered capacity of visual working memory to select multiple sets and store information about them concurrently. Second, Experiment 2 replicates and extends Hatano's claim that linguistic resources are not essential to abacus computations, and that motor representations may be more critical. Finally, our studies suggest that MA is not—as would be expected from previous literature—a phenomenon in which experts' representations differ dramatically in structure from those of novices. Instead, the power of the MA technique is that mental representations of the soroban abacus fit neatly into visual working memory, such that untrained controls store abacus images in a way not unlike highly trained MA users. In summary, our studies support a view of MA as a visual method for representing exact number that is tailored to the structure of the visual system.

## Experiment 1: Rapid addition

Our first experiment was designed to probe the limits of the MA representation. Because of the problem posed above—the inability of the ANS or visual working memory to represent the whole of the abacus—we were interested in what factors controlled the difficulty of doing particular arithmetic problems using MA. To the extent that performance is tied to particular aspects of the underlying representation, this method may allow us to differentiate hypotheses about MA.

We were particularly interested in whether MA performance declines as the total number of beads in a representation increases, or whether some sort of grouping in MA representations minimizes error related to bead number. One such grouping would be the partition of the MA image into columns. We hypothesized that each column in MA could be stored as a separate set in visual working memory. A strong prediction of this hypothesis is that MA users should be able to represent only 3–4 abacus columns, since previous work has found that only 3–4 sets can be represented in parallel (Halberda et al., 2006; Feigenson, 2008).

We used a task that was well-practiced for the students in our population: addition. In order to map out each individual participants' performance on a range of different problems, we made use of adaptive paradigms that presented more difficult problems when participants succeeded and easier problems when participants made errors. The use of adaptive paradigms is an important part of psychophysics research, but these paradigms are less used in research on higher-level cognitive phenomena. In the following set of experiments we make extensive use of adaptive designs because of quirks of the population we were studying: although many MA students were extremely proficient at the technique, they were still relatively young children and could not be relied on to complete very long experiments. In addition, their level of skill varied widely. Thus, we needed a method for quickly tuning an experiment to the level at which participants would give us information about the questions of interest.

In a between-subjects design, we tested the dependency of MA computations on A) the number of abacus columns in an addition problem (*width* condition) and B) the number of operations in a problem (*height* condition). In the width condition, we manipulated the width of the addends participants were asked to solve, first testing 1 + 7, then 18 + 34, then 423 + 814, etc. In the height condition, we manipulated the number of two-digit addends presented, first testing 18 + 34, advancing to 53 + 19 + 85 and eventually to problems like 77 + 56 + 21 + 48 + 92 + 55 + 61 + 57.

*Methods*

*Participants.* All MA participants in all experiments were children enrolled in Universal Computation Mental Arithmetic System (UCMAS) franchise schools in Gujarat Province, India. Participants were chosen for inclusion in the initial subject pool on the basis of A) their completion of level 4 UCMAS training (which includes approximately a year of physical abacus training and an introduction to the MA method), B) their ability to travel to the test site, and C) their instructor's judgment that they were among the best students in their cohort. In Experiment 1, 119 children participated; they had a mean age of 10.3 years (min = 5.8, max = 16.3).

*Stimuli and Procedure.* All stimuli were presented on Macintosh laptops via custom software designed using Matlab with Psychtoolbox. Responses were entered on USB numeric keypads. Instructions were given in English, unless children had difficulty in comprehension. In that case, instructions were given by a trilingual teacher in either Hindi or Gujarati depending on the child's preference. Instructions were illustrated with examples until the child had successfully answered several trials. In general, children had considerable practice with addition and thus had little difficulty understanding the task.

On each trial, children were asked to enter the sum of a group of addends. The addends were presented simultaneously on a computer screen until the participant typed an answer or until 10 seconds had elapsed. In the *width* condition (N=51), on each trial, the participant was presented with 2 vertically-presented addends and asked to sum them, and the place value of the addends was varied from 1-digit addends up to a maximum of 8-digit addends. In the *height* condition (n=68), on each trial, the participant was presented with some number of 2-digit, vertically-presented addends and asked to sum them. The number of addends was varied from 2 addends to a maximum of 10 addends.

In each condition, the manipulated variable was adapted via a transformed staircase procedure (Levitt, 1971). These procedures are commonly used in psychophysics to estimate accuracy in a task and to find the level of difficulty for that task at which participants performance meets a particular accuracy threshold. For example, in the width condition, the staircase procedure proceeded as follows: following two correct answers, the length of addends increased by 1 digit; following one incorrect answer the length decreased by 1 digit. In the height condition, the staircase was identical except that the number of addends increased by 1 following two correct answers and decreased by 1 following an incorrect answer.

This "2 up/1 down" staircase has been shown to converge around a stimulus difficulty level for which participants give approximately 71% correct answers (Levitt, 1971). We chose this kind of staircase in order that participants would be making primarily correct answers so that the task did not appear demoralizing or unnecessarily difficulty while still measuring performance across a range of difficulties, even for students of highly varying levels of expertise.

Stimuli for the height condition were sampled randomly from the range 10–99, while those for the width condition were sampled in the same manner depending on the width of the addends. Participants received feedback following their answer and saw a message indicating that they were out of time if they did not answer within 10s. The task was timed to last a total of 5 minutes and participants generally completed between 30 and 40 trials within this time limit.

*Results and Discussion*

Participants were in general highly expert at the addition task. Representative results from seven participants in each condition are shown in Figure 3. These curves summarize the percentage of correct answers given at each level the participant was exposed to; participants in the figure are sampled uniformly from the range of participants so that those on the left are the lowest performers while those on the right are the highest performers and those in the middle are approximately evenly spaced on the dimension of task performance.

For the purposes of our analysis we were interested in the limits on performance across conditions. Thus, we needed a robust summary statistic describing individual participants' performance in this experiment. We experimented with a variety of summary measures, including the parameters of the logistic curves plotted in Figure 3. Of these measures, the one that proved most robust to participants' errors was the average number

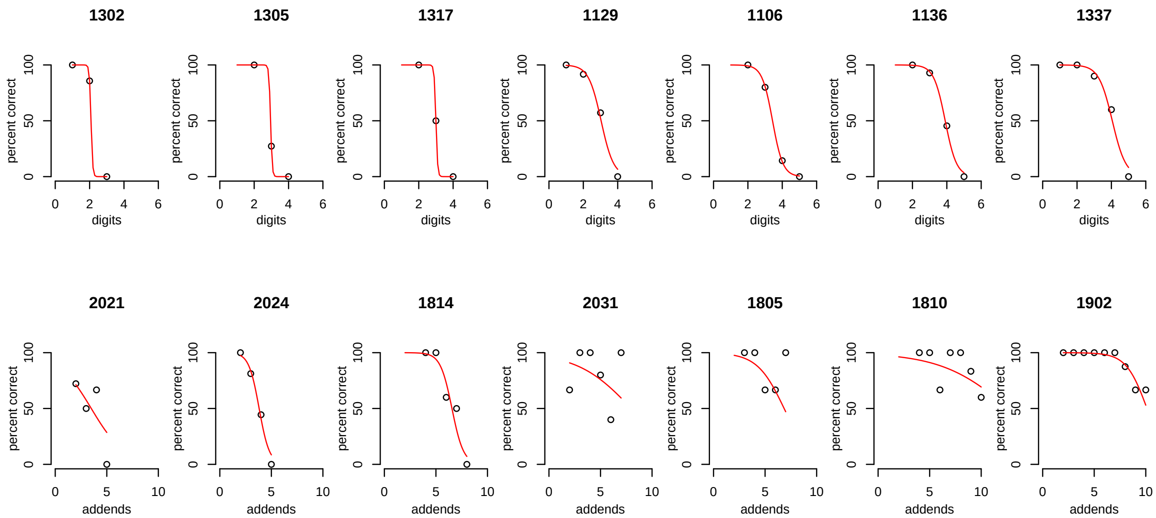


*Figure 3.* (top) Accuracy curves for representative participants in the width condition. Each subplot shows the percentage of correct trials by the number of digits presented; the red line shows the results of a logistic regression. Participants were selected by sampling uniformly along the dimension of digit thresholds (see text). (bottom) Accuracy curves for representative participants in the height condition, plotted by the number of addends presented. Participants were again sampled uniformly across the range of thresholds.

of addends presented after the staircase converged (in practice, we allowed 20 trials for convergence). As noted above, this number corresponds to an estimate of the level at which participants would be 71% correct—a psychophysical threshold value. We use this number as the primary description of an individual participants' threshold on the measure that was being manipulated: for the width condition, this was the size of the addends they could add together successfully within the time limit. For the height condition, this threshold was the number of addends they could add together successfully within the time limit.[4]

[4]Although a standard method of finding thresholds in a psychophysical task would be to extrapolate

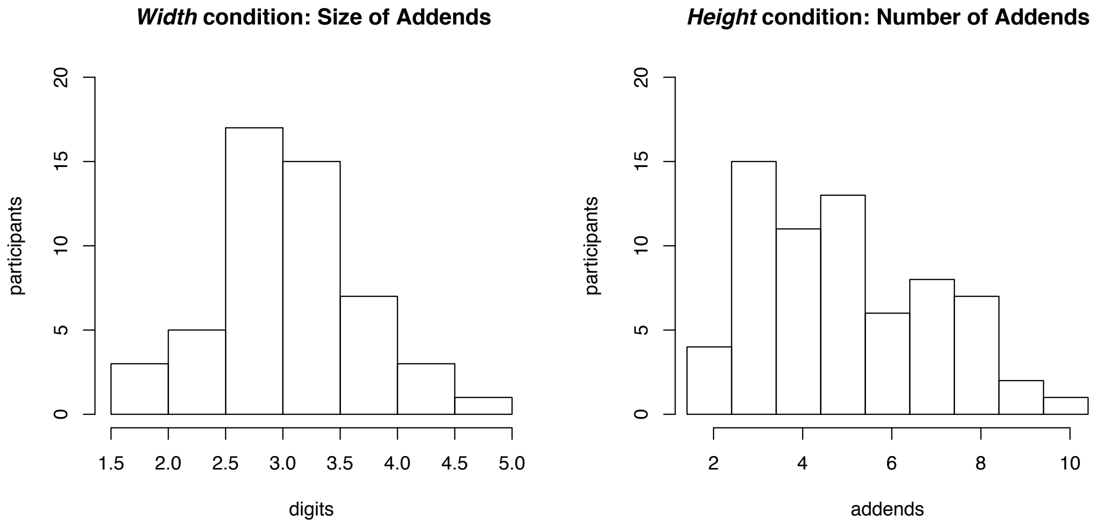


*Figure 4.* (left) Histogram of thresholds from width condition: mean size of addends presented for trials after the staircase had converged. (right) Histogram of results from height condition: mean number of addends presented for trials after the staircase had converged.

In the width condition, while most children were able to add two three-digit addends consistently (average performance on these trials across participants was 67% correct), almost none were able to add four-digit addends (average performance was 23%). Corresponding to participants' difficulty in performing four-digit problems, there was a

from the logistic curve to find the number of addends at which participants' performance was expected to be a particular level (e.g. 50%), the wide range of psychometric functions we observed—and the relatively limited number of trials we were able to ask for from our participants, many of whom were grade-school children—made this approach unreliable. In contrast, the average trial level metric that we adopted accorded very well to our intuitions about participants' performance, formed after close examination of individuals' data. Nonetheless, we believe the choice of summary measure did not qualitatively affect our results: the same basic patterns were observed for measures like parameters of the logistic curve or alternative threshold values estimated from the logistic curves.

very tight distribution of addend thresholds. Figure 4, left, shows a histogram of these average digit thresholds. Consistent with natural variation in a cognitive limit, the digit thresholds were normally distributed ($\mu = 3.1$, $\sigma = .62$, Shapiro-Wilk $W = 0.98$, $p = 0.74$), and did not show non-normal kurtosis (Anscombe-Glynn test for non-normal kurtosis, $k = 2.91$, $z = .25$, $p = .80$).

In contrast, in the height condition, there was no tight limit on the number of addends that participants were able to sum. A histogram of addend thresholds is shown in Figure 4, right; the distribution was quite different from that in the width condition. The thresholds were not normally distributed ($\mu = 5.1$, $\sigma = 1.99$, Shapiro-Wilk test for violations of normality $W = 0.95$, $p = 0.02$). In addition, threshold scores exhibited substantially lower kurtosis, indicating a wider spread of abilities (Anscombe-Glynn test for non-normal kurtosis, $k = -2.05$, $z = -2.05$, $p = .04$). Finally, participants' addend threshold scores ranged widely, from 2 to 10.

What explained this difference in threshold distributions? The diffuse distribution of thresholds in the height condition might simply reflect the varying skill levels across participants; some added quickly and accurately due to greater practice with the MA technique, while others were slower and more error prone. In contrast, the tight distribution of thresholds in the width condition was more puzzling. There is no qualitative change between how two-, three-, or four-digit addends are added on the abacus. Instead, a limit on the structure of the MA representation seems like a possible explanation for this result. This explanation would be consistent with our hypothesis about the relation between the number of columns on the abacus and the limits on the number of sets that can be stored in parallel in visual working memory, but might also be confounded with the number of beads on the abacus, as opposed to the number of sets. The next analysis follows up on this possibility by performing a separate analysis of the

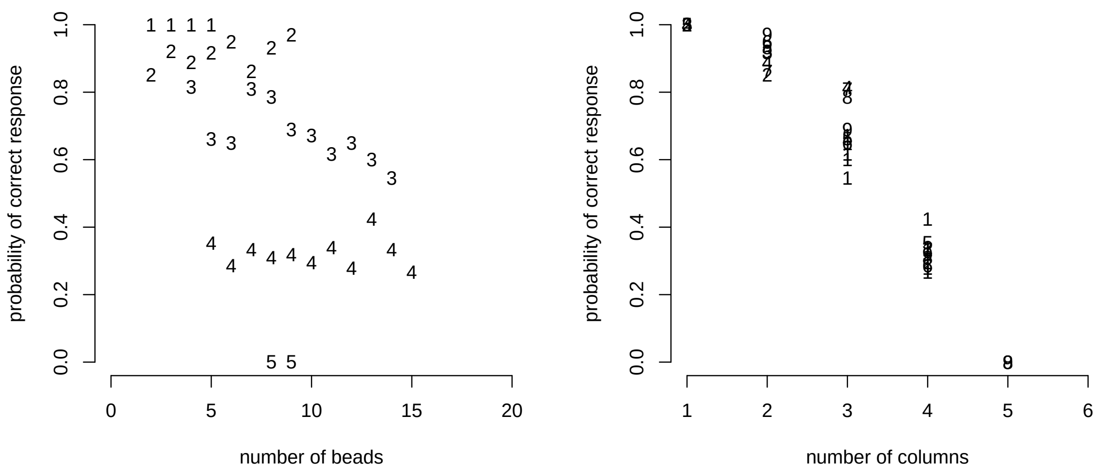


*Figure 5.* Probability of a correct response in Experiment 1 is plotted by two different predictors on the horizontal axis. (left) Probability correct is plotted by the number of abacus beads in the solution of the addition problem. The numbers used as markers on the plot correspond to the number of columns involved in a problem. For problems with a given number of columns—e.g., all instances of "4" on the plot mark all the problems with four columns—there is no relationship between number of beads and probability of success. (right) Probability correct is plotted by the number of columns in a problem, and plotting markers correspond to the number of beads. Now it is clear that number of columns predicts performance and there is no additional effect of beads.

data in the width condition.

We next investigated the distribution of thresholds in the width condition by contrasting two possible explanations of participants' trial-by-trial performance in this task. The first explanation was that problem difficulty increases as the number of abacus beads "in play" on the abacus increases. The second explanation was that the number of beads was irrelevant to problem difficulty and that the number of columns of abacus involved in the problem is the primary determinant of performance.

We plotted the probability of success in a problem by the number of beads required to represent the solution, marking the number of columns by using a numeric marker in the plot, in place of a dot or square. We found no relationship between the number of beads and probability of success (Figure 5, left). In contrast, when we plotted probability of success by the number of columns of abacus involved (the number of digits in the addends), there was a strong relationship between these two factors (Figure 5, right).

To quantify this visual impression, we used multilevel logistic regression models (Gelman & Hill, 2006). These models allowed us to model the entire dataset produced by our participants (all trials in all conditions). Unlike typical ANOVA analyses they A) are appropriate for binary response variables (like whether a response was correct or not), B) allow for the analysis of adaptive/asymmetric designs such as the one we used here, and C) allow us to test for the effects of trial-level predictors like the number of beads or columns involved in a particular addition problem. We use these models throughout the studies reported here. In each experiment we used the multilevel model to capture the effects of interest using group-level coefficients ("fixed" effects); at the participant level, each model also included separate intercepts ("random" effects) for each participant.

To test the effects of the number of abacus beads and abacus columns on participants' performance in the width condition, 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. While 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 ($\chi^2 = 504.58$, $p < .0001$). In a model with both predictors, the column predictor remained highly significant ($p < .0001$) while the bead predictor was no longer significant ($p = .21$). This result shows that all variance explained by beads is due to the correlation of number of beads with number of columns.

Summarizing the results of this analysis: while MA students varied widely in the number of two-digit addends they could successfully sum in 10 seconds, nearly all students were limited in width of the addends they could sum. This difficulty going beyond three-digit addends was consistent with a column-based limit on computations (but not a limit based on the number of beads involved in the computation), and further consistent with a theory of MA as drawing on parallel set representations in visual working memory.[5]

## Experiment 2: Verbal and manual interference

Hatano (1977) observed that MA users could answer basic factual questions while doing abacus calculations. This result suggested that some language comprehension and production could be integrated into the abacus routine, again suggesting that MA representations are primarily visual, rather than linguistic. In addition, both Hatano's observations and our own experiences suggested that MA users made considerable use of abacus-like gestures to facilitate computation. These gestures vary from person to person

[5] In fact, one advanced MA technique (which these participants had not been exposed to) involves learning to use a three-column abacus to add much larger numbers by breaking them into parts. Although Hatano and Osawa (1983) suggested that very highly trained adult MA experts may gain the ability to represent more abacus columns over the course of many years of practice, we found no evidence for a significant expansion in even the most experienced abacus users we tested.

in their degree of intensity, but they are a notable feature of the MA technique and obstructing them appears to cause difficulty in calculation. Thus, we were interested in comparing verbal and manual interference effects in order to understand the relative contributions of linguistic and gestural resources to MA computation.

Because of the relatively demanding nature of dual-task studies, which involve carrying out two complex and unrelated tasks at once, we wanted to identify a subgroup of MA participants that were expert enough to be able to do any kind of computation under interference. We thus conducted an initial screening experiment with children from local abacus schools, and asked the most expert group of children in the initial sample to return for testing in Experiment 2 and perform a set of adaptive addition tasks, similar to the height condition of Experiment 1.

Participants were tested in three different interference conditions—manual interference, verbal interference, and combined manual-verbal interference—as well as a baseline no-interference condition. In the manual interference condition, participants were asked to tap their fingers on the table as they did the addition problem (pausing only to enter the sum on a keypad). In the verbal interference task, participants listened to a story on headphones and "shadowed" it by repeating back words and phrases immediately after hearing them. Combined interference required performing both of these tasks at once.

In order to test whether the pattern of interference effects we observed was specific to MA representations or general across other strategies for mental arithmetic, we additionally tested a group of untrained control participants in an identical paradigm.

*Methods*

*Participants.* MA Participants were selected to be among the top students in their cohort. This determination was made on the basis a pre-screening test of 346 MA students using a test of physical and mental abacus ability. Our participants either scored higher than 90% on either the mental or physical portion of the test, had completed the full UCMAS course of 3 years of training and gone on to the “grand levels” (advanced training available for high performers), or were members in a group of UCMAS students who performed demonstrations of mental arithmetic at public events. The 15 MA participants had a mean age of 13.3 years (min = 9.7, max = 16.3).

Control participants were 23 undergraduates at the University of California, San Diego who participated in exchange for course credit.

*Procedure.* The basic procedure was substantively identical to the height condition in Experiment 1: participants were given an adaptive addition task in which they added sets of two-digit addends under a 10 second time limit. There were two minor differences: first, the paradigm began with a control task of retyping a single addend, and second, the maximum number of addends was limited to eight, rather than 10.

Each participant received four five-minute blocks of trials. In the No-Interference block type, sums were computed as in Experiment 1. In the Verbal-Interference block, participants were asked to listen to and repeat a children’s story (verbal shadowing). The story that we used for the verbal shadowing task was “Ali Baba and the Forty Thieves.” It was read in Indian English, Hindi, and Gujarati by a trilingual instructor and children were allowed to pick the language in which they were most comfortable to do the shadowing task (so as not to conflate language difficulties with true interference effects). Participants were instructed on verbal shadowing via a demo by an experimenter and then given approximately one minute of shadowing practice before they began the addition task. For untrained control participants, an American English version of Ali Baba was used for verbal shadowing.

During the Manual Interference block, participants were instructed to drum their fingers on the table and then pause briefly with one hand to type in the answer. If children had difficulty drumming their fingers independently (as some of the youngest participants did) they were encouraged to tap their hands on the table at the same pace. The manual interference task was also demonstrated by the experimenter. The last block combined simultaneous manual and verbal interference. All participants performed all four blocks (no interference, verbal interference, manual interference, and both interference tasks). The second and third blocks were counterbalanced for order. An experimenter was present during testing to monitor the children for compliance with the interference tasks.

Ensuring compliance was difficult. An experimenter was present to remind participants to continue performing the manual interference task and would tap along with the participant to remind them to be consistent in performing the task. However, this was not possible during the shadowing task because the experimenter could not hear the exact source text. Accordingly, we conducted an analysis to ensure that participants were not trading off one task against the other, failing to shadow verbally while they were performing the addition task. An independent coder segmented videotapes of each participants' performance by trial and rated their shadowing on each trial on a scale of one to five from completely disfluent to completely fluent. We then split these ratings by the difficulty of the addition problem (number of addends) and whether the participant had given a correct or an incorrect answer. This analysis showed no evidence for trading off between one task and the other. Shadowing fluency was numerically very similar between trials where answers were correct vs. incorrect, and did not vary with the difficulty of the addition task.

*Results and Discussion*

Despite the unfamiliarity and difficulty of the interference tasks, the MA experts still showed surprising proficiency in adding while performing the interference tasks, with some participants at ceiling even while performing both interference tasks simultaneously.

Table 1

*Coefficient weights for mixed regression models fit separately to error rates for MA and control participants in Experiment 2. ":" indicates an interaction term.*

| | MA | | | Untrained control | | |
|---|---|---|---|---|---|---|
| Predictor | Coefficient | Std. Error | $z$ value | Coefficient | Std. Error | $z$ value |
| Intercept | 4.78 | 0.28 | 16.93 | 6.93 | 0.34 | 20.17 |
| Addends | -0.45 | 0.07 | -6.67 | -2.09 | 0.14 | -15.22 |
| Manual interference | -1.61 | 0.20 | -8.22 | 0.11 | 0.14 | 0.81 |
| Verbal interference | -1.34 | 0.19 | -7.06 | -1.22 | 0.14 | -8.50 |
| Manual:Verbal | 0.87 | 0.23 | 3.74 | -0.28 | 0.19 | -1.49 |

In contrast, control participants were far less proficient at addition overall, and were unaffected by motor interference but strongly affected by verbal interference. Figure 6 shows estimated performance at each level for each condition and group, along with a summary of the data.

Because of the complexity of this dataset (two groups, each with four within-subjects conditions), we began by analyzing each group's data separately using a separate mixed logistic regression model, as in Experiment 1. Each model included a group-level intercept term as well as effects of the number of addends in a problem, effects of verbal and manual interference, and an interaction term for performing both interference tasks simultaneously; the model also included participant-level slope and intercept terms to account for differing baseline levels of abacus skill across participants. Coefficients for the models for both groups are given in Table 1.

For MA participants, the manual and verbal interference tasks decreased

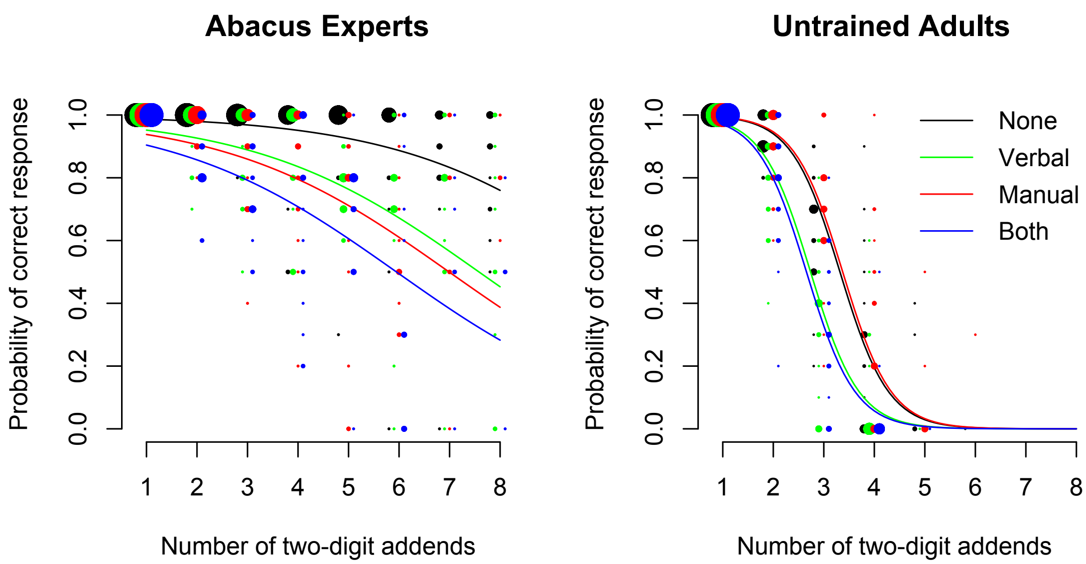


*Figure 6.* Results of Experiments 2. The probability of a correct response in an adaptive addition task with an increasing number of two-digit addends is plotted by the type of interference and the number of addends for (a) MA users and (b) controls. Size of dots reflects the proportion of participants with a given mean performance; lines reflect the best-fit curves for a mixed logistic regression model, with colors showing the different interference conditions.

participants' performance significantly from baseline ($p < .0001$ for both coefficients), but they did not differ significantly from one another. The manual by verbal interaction term had a positive coefficient value, indicating a significant sub-additive interaction. We speculate that this interaction is probably due to some fixed task-switching cost that is incurred regardless of whether there are two tasks being performed or three as well as the specific costs due to each interference task. For control participants, performance was significantly decreased by verbal interference ($p < .0001$) but there was no significant effect of manual interference and no interaction.

Next, to test for a significant group by interference-type interaction, we constructed a mixed model for both datasets. This model included terms for participant group (MA vs. control), an interaction of participant group and number of addends, and terms for motor interference, verbal interference, their interaction, and the interaction of all three with participant group. This complex model nevertheless yielded highly interpretable coefficient estimates. In the interest of brevity, we report only those that relate directly to the question of what the differences were between groups. Capturing the overall higher performance of MA participants in the task, there was a highly significant interaction between the coefficient on number of addends and participant group ($\beta = 1.55$, $p < .0001$). In addition, there was an interaction of participant group and motor interference ($\beta = -1.76$, $p < .0001$), capturing the greater effect of motor interference on MA users.

Most adults in the control group were unable to add more than two addends with any facility while under verbal interference. Our own anecdotal experience suggested that the most difficult operation under verbal interference was "carrying": when the sum in one place value exceeded 9 and needed to be applied to a higher place value (as in e.g. 27 + 19 but not 27 + 12). To capture this effect, we created a mixed model of control participants' data, identical to that reported in Table 1 except in that it included a term for the number of carries in each problem. We found that the coefficient added significantly to the fit of the model ($\chi^2(1) = 342.47$, $p < .0001$), with a coefficient estimate ($\beta = -1.24$) of approximately the same magnitude of that for addends ($\beta = -1.18$). When we added an interaction between verbal interference and number of carries, this coefficient was also significant and negative ($p = .002$, $\beta = -.38$). These analyses indicate that "carrying" is a difficult operation in verbal arithmetic, and further that carrying interacted with language interference to produce an extra interference effect. When we carried out the same analysis with the MA participant data (coefficient estimates can again be compared to those in Table 1), we found neither a significant effect of carries ($\beta = -.11$, $p = .17$) nor a significant interaction between number of carries and verbal interference ($\beta = -.07$, $p = .26$). This analysis confirms that MA users are employing different representational resources to complete their calculations than control participants.

Both verbal and manual interference produced decrements in some MA participants' performance, but if anything, manual interference was harder (despite the simplicity of the tapping task). Some participants were still able to perform close to ceiling even under interference. These data speak against an account under which either language or manual skills are critical to MA performance, although both may play some facilitatory role[6]. Comparing the magnitude of interference effects between the MA and control groups is difficult across groups with such disparate baseline skill levels. Nonetheless, two pieces of evidence suggest differences in the method of computation employed by the two groups: first, the MA group made much greater use of motor resources during computation; and second, the MA group did not show the same interaction between verbal interference and the number of "carries" shown by the control grop. Thus, evidence from this experiment does not rule out the possibility that there is some involvement of language in MA (we return to this issue in the General Discussion), but it does strongly suggest qualitative differences between the verbal algorithm used by control participants and the MA strategy.

[6]We noted anecdotally one very interesting phenomenon with respect to shadowing performance: gaps or difficulties in shadowing were almost always at the end of an abacus computation, immediately before entering the sum onto the keypad. We hypothesized that these gaps were caused by the necessity of translating abacus representations into Arabic numerals. This translation process is likely at least partially linguistic (since Arabic numerals are so closely linked to their corresponding word forms), and hence would be likely to cause greater difficulties in the verbal interference task if abacus computations were otherwise non-linguistic.

## Experiment 3: Abacus flashcards

In our final experiment, we asked whether the benefit of MA expertise is seen in forming initial MA representations, or whether the primary benefit of practice with MA comes in performing computations once representations have already been formed. Although it seems clear that MA training requires extensive practice with computational procedures, it is less clear whether it also involves a form of perceptual expertise, or whether, as we suggested in the Introduction, MA takes advantage of existing visual resources to build representations of number. However, because Experiments 1 and 2 relied on addition, these studies were unable to differentiate computational expertise from perceptual expertise. To explore this question, we contrasted the performance of MA experts and adult novices on a task that does not require arithmetic computation, but that does require forming MA representations: flashcard reading.

As part of their abacus training, children that we tested learn to rapidly read abacus flashcards: a card showing a schematic representation of an abacus is flashed and participants call out the value shown on the abacus. Here we asked whether adult novices could also perform this task with a brief training, and whether their performance differed qualitatively from that of child experts, or if instead the two groups exhibit similar limits in their ability to rapidly perceive bead arrays. Evidence that the two groups use similar perceptual mechanisms for representing abacus structures would suggest that MA training does not involve acquiring unusual perceptual expertise, but instead involves practice of computational algorithms that are defined over existing perceptual resources.

To compare how experts and novices encode abacus structure, we not only tested both groups with an abacus reading task, but also asked them to perform a series of dot-array estimation tasks. The logic of this second set of tasks was as follows. If MA experts acquire perceptual expertise when they learn MA, then they may become better not only at reading an abacus flashcard, but also at encoding other perceptual arrays, like a set of dots on a screen. For example, when asked to estimate the number of beads on an abacus structure, MA experts may be faster at this non-abacus task than adult controls. However, in contrast, if MA exertise is not perceptual in nature, then we may expect little difference in how experts and novices make such estimates.

In addition, using a dot array estimation task allowed us to probe whether abacus structure might be, in some sense, optimized for visual processing. If abacus was designed to fit the limits of visual processing, rather than requiring the development of perceptual expertise, then we may expect that a non-abacus task like estimation will be facilitated when arrays of dots are organized like an abacus, into small vertical columns.

To test this, we asked experts and novices to make estimates for five different types of dot array. Across these five conditions, we parametrically varied the similarity of the displays to an abacus, as shown in Figure 7a. The identical display tested the difference between the abacus reading task and the estimation task. The rotated and configural estimation displays tested whether the specific orientation and rectilinear arrangement of beads, respectively, contributed to estimation accuracy. The jittered estimation task tested whether the spatial extent of the abacus display was important. Finally, the random dot estimation display provided a baseline for estimation performance. By systematically varying aspects of the abacus structure in a distinct task, these conditions allow us to investigate which parts of abacus structure aid in perceptually grouping elements of complex displays.

*Methods*

*Participants.* The 133 MA participants in Experiment 3 had a mean age of 11.2 years (min = 6.8, max = 15.0). All participants were familiar with the abacus reading task from their training; these participants were sampled from the same population as those in Experiment 1 and had the same level of training. In addition to the MA participants, 30

UCSD undergraduates participated for course credit. None had any prior experience with abacus calculation or MA technique. Due to experimenter error, one estimation task from three participants in the control group was not included in the analysis.

*Stimuli and Procedure.* Each MA participant in this experiment performed two tasks, in a random order: an abacus flashcard reading task and one of the five estimation tasks. For control participants, stimuli and procedures were identical, but all control participants were tested in all six tasks in one of two random orders. Prior to testing, each control participant completed a two-page abacus training worksheet which taught them how to read abacus representations of the type used in our experiments and which gave them practice on twelve abacus-reading problems.

Example stimuli for each of the six conditions are given in Figure 7a. In the abacus flashcard task, 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 and there was no time limit for responses.

For the estimation tasks, participants simply reported the number of dots on the screen. Tasks were (1) Identical: abacus flashcards identical to those used in the reading task (N=24), (2) Rotated: mirror images of abacus flashcards rotated 90 degrees (N=24), (3) Configural: abacus flashcards with the beam and rod structures removed but the configuration of beads preserved (N=36), (4) Jittered: random dot arrays jittered within the bounding box space 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).

Each estimation task was adaptive according to the same distribution of trials as the flashcard task. For individual participants, stimuli were generated beginning with 1-column abacus flashcards (with 1 – 5 beads) and then converted into estimation displays. If two of these trials were completed correctly, stimuli sampled from a 2-column abacus display were converted into estimation displays (with 1–10 beads). This procedure matched the adaptive structure of the estimation tasks exactly to the adaptive structure of the flashcards task. Because of the inherently noisy nature of the ANS, this procedure ensured that the majority of trials were in the range of 1–10 items (since nearly all trials with 1–5 items were correct and most were incorrect for quantities above 5). Nevertheless, most participants saw at least some trials in the 10–15 range.

*Results and Discussion*

All results for both MA and control participants are given in Figure 7. We began by analyzing the abacus-reading component of the experiment, which every participant completed. Consistent with the results of Experiment 2, we found that abacus reading accuracy was better predicted by a model including the number of columns on the abacus than by a model with the number of beads in play ($\chi^2 = 815.89$, $p < .0001$). Abacus reading accuracy was highly comparable across the groups of MA children tested in each of the five estimation tasks (M = .72, .72, .73, .74, and .73, respectively). Although accuracy data were noisier in the untrained participants, results resembled those for MA users: the column-based model fit far better than the bead-based model ($\chi^2 = 128.36$, $p < .0001$). Hence, even in the absence of extensive MA training, untrained participants grouped the abacus displays into columns.

Supplementing this analysis, we conducted an error analysis of the MA users' data. The most interpretable errors came on the 57% of error trials where only a single column was misread. In these trials, there was an effect of the number of beads being read, but it

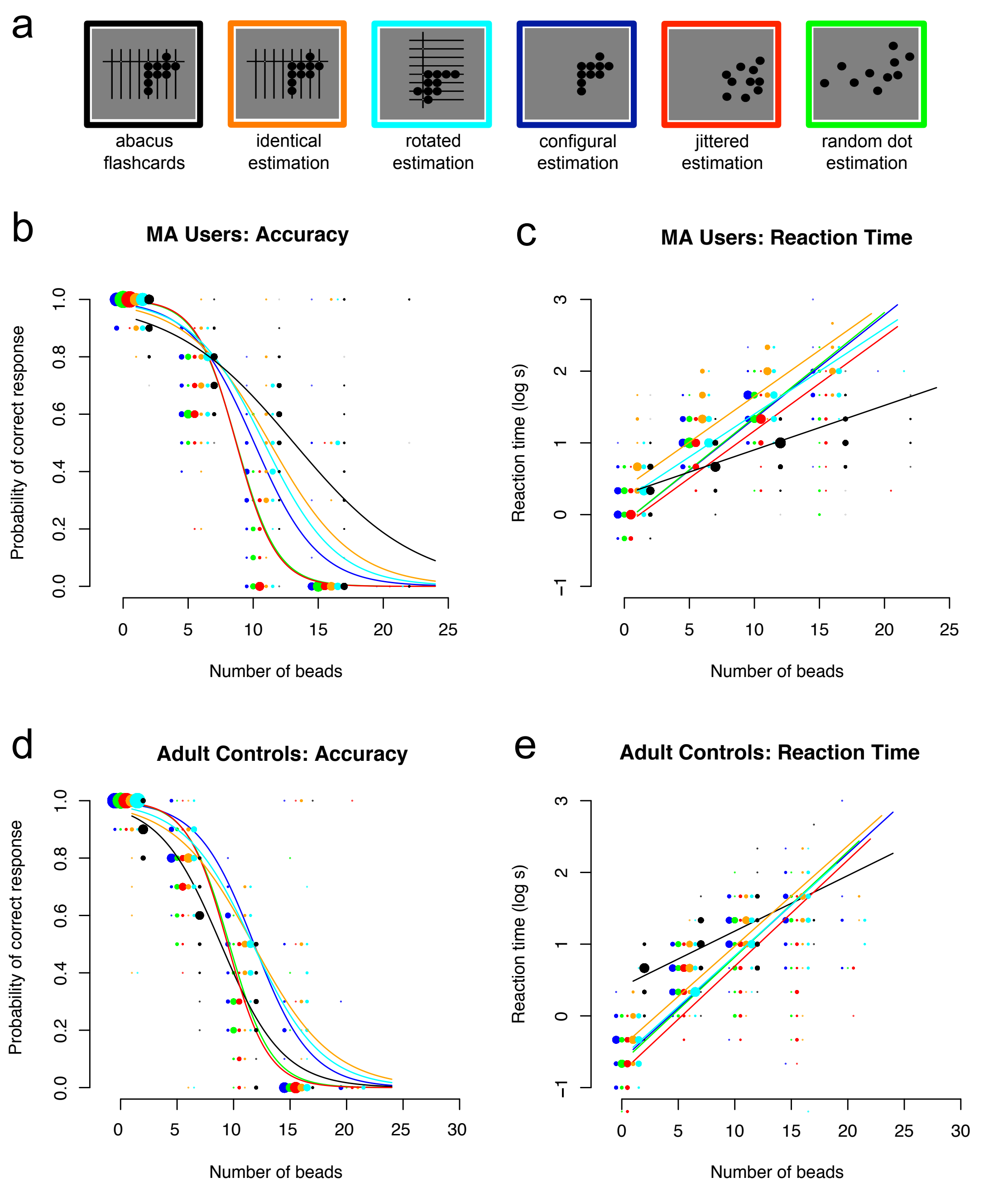


*Figure 7.* (a) Displays in Experiment 3. The color around each display corresponds to dots and lines in b–e. (b) Accuracies for MA participants. Probability of correct response plotted by number of abacus beads in the correct response. Size of dots reflects the proportion of participants with that mean performance; lines reflect best-fit curves for a logistic regression model. (c) Log reaction time for MA participants. (d) Accuracy data for untrained adults. (e) Log reaction time data for untrained adults.

Table 2

*Coefficient weights for logistic mixed model analyses of accuracy in MA users and control participants in Experiment 3.*

| 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 3

*Coefficient weights for linear mixed model analyses of log reaction time in MA users and control participants in Experiment 3.*

| 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 |

was not a linear effect that grew with quantity. Instead, children were most likely to make errors when reading columns that contained 3 beads (39% of all errors), followed by columns with 2 and 4 (20% and 28% respectively), and were least likely to make errors for columns containing 1 or 5 beads (9% and 4% respectively). While this result is congruent with a number of possible theories about the internal structure of MA columns, it argues against the operation of the ANS even within individual columns. If columns were represented using the ANS, we would predict the largest number of errors to occur in columns with 5 beads, rather than 3. We return to the implications of this analysis in the General Discussion.

Next, we compared flashcard accuracy data with accuracy data from the estimation tasks. For each group, we fit a single multi-level model to the entire dataset produced by our participants (all trials in all conditions), with group-level effects for each condition and the interaction of condition with number of beads in the trial. Coefficients are reported in Table 2, along with $z$ value approximations for the group-level effects in the model.[7] The ordering of accuracies for the five estimation tasks was almost identical for the MA-trained and untrained groups. In both groups, jittered and random estimation tasks grouped together and these two tasks were more difficult than the three configural conditions, which also grouped with one another. This suggests that the increased performance of MA users in the identical, rotated, and configural conditions relative to random and jittered conditions was not due to their extensive MA training. Instead, the advantage seen in these conditions seems to be a consequence of the perceptual properties of the stimuli. However, abacus reading was a difficult and error-prone task for the control participants,

[7] All $p$ values are derived from this $z$ approximation. While this approximation can be anti-conservative for small amounts of data, the large size of the dataset we used means that this anti-conservatism is quite minimal (Pinheiro & Bates, 2000). Using this approximation, non-overlapping standard errors can be interpreted as significant differences at $p < .05$.

while for the MA users, abacus reading exhibited greater accuracy (with lower intercept and higher slope) than any of the five estimation tasks ($p < .0001$ for all comparisons).

For each group, we also fit similar models to reaction times, measured at their first key-press (Table 3).[8] We used a linear regression to predict the natural logarithm of reaction time, choosing a log transform because individual reaction times (the input data to the mixed linear model, which operated over every trial individually rather than over means) are well-described by a log-normal distribution. Reaction times greater than three standard deviations above the mean (constituting 1.9% of the total data) were omitted from this analysis. Because reaction times tend to decrease over the course of an experiment, we added a coefficient for trial-number to the model.

For both groups, the reaction time slope for abacus flashcards was qualitatively different from that for all of the estimation tasks (which were largely undifferentiated). Both untrained adults and MA users showed a lower reaction time slope for abacus reading than for any other task (all $p$s $< .0001$ for comparisons between coefficients). The different slope for abacus reading relative to estimation for both groups is consistent with the view that abacus reading is an operation over columns rather than individual beads. Since estimation reaction times increase according to the number of items in a display (Whalen et al., 1999), the flatter slope and faster RTs for abacus reading indicate a distinct process—e.g., one that operates over columns and their configurations, rather than over individual beads. Nevertheless, for control participants, 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 quantities to Arabic numerals.

We conducted this experiment to test the nature of MA expertise, and whether it is rooted primarily in the mastery of computational procedures or also involves acquiring

[8] We made use of the first key-press rather than total input time in order to avoid the confound that larger numbers take longer to input. Thus, RTs analyzed here are due to processing time for different displays.

perceptual expertise. The flashcard reading task, easily learned by novice adults, found qualitatively similar limits in both novices and MA experts, suggesting that similar mechanisms were used by both groups. Further, we found that in both groups dot-array estimation grew more accurate as dot-arrays grew closer in structure to actual abacus configurations, suggesting that visual arrays are more easily processed as they become more similar to abacus structures. Both findings are consistent with the hypothesis that MA is adapted to the design of the human visual system, rather than requiring the acquisition of perceptual expertise. We conclude, therefore, that the difference between MA and control participants is the set of highly-practiced operations that MA participants are able to bring to bear on arithmetic problems, not the nature of the representations they can form.

## General discussion

Our studies examined mental abacus (MA), a powerful mental arithmetic technique that allows users to make extremely fast and accurate computations beyond the reach of typical arithmetic techniques. We asked two broad questions about MA. First, building on previous reports, we tested the idea that MA computations are non-linguistic by asking users to perform addition while doing concurrent linguistic and motor interference tasks, and compared this to the performance of untrained control participants. Second, we asked how MA might be represented in the visual system. Our results support the idea that columns in MA are represented as separate sets. On our view, inputs and outputs to MA computations are linguistic, but computations performed over this parallel set representation involve visual working memory resources (along with an intriguing gestural component related to the motor operations involved in individual steps of MA arithmetic). In the following two sections we sketch a more detailed picture of MA representations and then discuss the relationship of MA computations to language and the approximate number system (ANS). We end by considering some broader implications of this work.

*The nature of MA representations*

Visual working memory is believed to be limited in its capacity to approximately four objects and the approximate number system lacks the precision to capture the exact numbers used in abacus computations (Alvarez & Cavanagh, 2004; Feigenson et al., 2004; Luck & Vogel, 1997). As a result, neither system has the capacity to encode abacus structure. Thus, the status of the initial proposal (that MA representations are images) has been tentative, awaiting an account of how the detailed structure of an abacus could be represented using limited visual resources.

As noted in the Introduction, recent studies provide evidence that the visual system can represent information about multiple sets. In these studies, participants perform three to four numerical estimates in parallel (Halberda et al., 2006; Feigenson, 2008). Additional evidence suggests that multiple object tracking abilities are enhanced when targets are displaced horizontally in space (allowing each visual hemifield to form independent representations) (Alvarez & Cavanagh, 2005). Together, these studies suggest that untrained subjects can use attention to track not only multiple objects, but also the members of multiple sets. These studies raise the possibility that the columnar structure of MA allows users to select multiple horizontally-adjacent columns in parallel, to then track their contents, assign features to columns, store their values, and perform computations over their contents.

Evidence for this proposal comes from several findings reported in this study. First, in Experiment 1 we found a tight limitation on the number of columns that can be involved in an addition problem, but a much looser limit on the number of beads or addends. These signatures were also found in Experiment 3, where subjects were asked to either read abacus flashcards or makes estimates for visual arrays that varied in their resemblance to an abacus. This study also revealed a difference in RT slopes for abacus reading compared with estimating from comparable displays. The presence of a distinct RT slope in even novice adults with only brief exposure suggests that grouping the MA representation in this way is an automatic consequence of the place-value organization of abacus combined with its visual structure, rather than an overlearned strategy.

Our findings also suggest that the particular layout of MA—rectilinear, horizontally extended, and segregated into columns—is optimized for visual processing. When participants were asked to estimate the number of "beads" in a visual array, both MA users and controls performed better as the structure of the arrays become more similar to actual abacus structures. Previous studies report that configural cues similar to those provided by the abacus, greatly facilitate numerical estimation, mirroring our findings (Mandler & Shebo, 1982; Atkinson, Francis, & Campbell, 1976; Van Oeffelen & Vos, 1982). For example, in a study by Atkinson et al. (1976), participants were asked to make estimates for columns of dots including up to 12 items. Accuracy was perfect for arrays up to 4, but fell off quickly thereafter. However, critically, estimates for sets of up to 8 were also perfect if items were arranged into rows of 4 or fewer dots, and the rows were presented at a 90 degree angle to one another. Thus, it is likely that the rectilinear structure and frame of the abacus make it considerably easier for the image to be grouped into sets and maintained in memory.

Our proposal leaves open the precise internal structure of abacus columns in MA, however. According to the proposal just described, either ANS or visual working memory alone might represent the beads within individual columns (e.g., by representing columns as approximate values from 1–5, or as arrays of objects, each able to receive its own attentional index). While our current data do not provide conclusive answers about the internal structure of columns in the MA representation, they nonetheless speak to the question, suggesting that the ANS is not likely used to represent the contents of columns.

Instead, our data suggest that MA users track individual beads within each column, perhaps shifting attention from one column to the next as they do computations.

First, in analyses of the data from Experiment 3, we found that MA users' errors when reading flashcards did not increase linearly as the number of beads on a column increased, but instead were affected by configural cues related to the position of particular beads. Thus, an estimate based purely on approximate number cannot explain childrens errors. Second, when MA users do arithmetic, they track the location of individual beads in space, and move them in their visual image. Evidence that users do this comes not only from the errors that they make and their access to intermediate steps of MA computation (Stigler, 1984), but also from the robust and often highly detailed gestures that children make when doing computations: their fingers move in space as they move beads up and down the columns of their virtual device (for discussion, see Brooks, Goldin-Meadow, Frank, & Barner, in prep). These abilities cannot be explained if column contents are represented in the ANS alone.

In addition, we found that the ANS abilities of the MA users in Experiment 3 were quite similar to those of control participants. This result is interesting to consider with respect to an ongoing discussion about the role of the approximate number system in symbolic math expertise. (Halberda, Mazzocco, & Feigenson, 2008) showed that those children who had more accurate approximate number representations also performed better in standardized tests of symbolic mathematics, even controlling for a host of other cognitive factors. This finding has since been replicated in another population (Gilmore, McCarthy, & Spelke, 2010), but the causal direction underlying the correlation is puzzling. Is it the case that practice with exact numerosities improves approximate magnitude estimation, or is approximate magnitude precision important for "checking your work" in symbolic mathematics? Our findings suggest a bound on any causal connection between symbolic practice and improvements in ANS accuracy: even extensive mental abacus practice—enough to allow children to be considerably better at mental arithmetic than adults—does not make children adult-like in their estimation abilities.

To summarize, we propose that MA representations are supported by parallel set representations, which operate over the columns of the abacus. Individual columns preserve information about the identity and location of individual beads, suggesting that they are not approximate number representations but instead have some more complex substructure. This substructure is not specified by the current data and may be a fruitful topic for future work.

*The relationship of MA to language*

Our proposal is that MA involves performing mental arithmetic in a non-verbal format, unlike standard mental arithmetic techniques which rely on phonological working memory. One component of this argument, made in Experiments 1 and 3 of this paper, involved showing how such representations might be possible, given known limits of visual working memory. An equally important step in the argument, however, is to show experimentally that MA users remain proficient at arithmetic even when doing a concurrent verbal task. This was the goal of Experiment 2.

In this experiment, inspired by earlier work of Hatano (1977), we showed that experienced MA users were surprisingly competent at doing mental sums while simultaneously repeating an auditorily presented story (and were equivalently affected by a simple tapping task). In contrast, we found that untrained adults were much more affected by verbal interference than MA users, but not affected at all by motor distraction. In addition, controls but not MA users showed differential verbal interference effects for sums that contained many "carries" (operations that spanned place-values). Together, these results confirm Hatano's results and provide support for the contention that MA relies on primarily non-verbal structures.

Language likely does play a role in the acquisition and use of MA, however. First, to our knowledge, abacus use is always taught after children learn verbal numbers. Second, when individuals use MA, both the inputs and outputs of computations are either Arabic numerals or their verbal counterparts. As a result, verbal interference should have some modest effect on the use of MA, to the extent that it interferes with this translation process. What we have argued here is that, although the inputs and outputs to MA are linguistic, the format of representations during MA computation is not. MA is like an electronic calculator in this sense: its internal states use a representation that is convenient for calculation but not linked to language, while its inputs and outputs are base-10 numerals, linked to language. In the case of a calculator, the internal representation is binary or hexadecimal numbers stored in electronic registers; in the case of MA, the internal representation is soroban columns stored in visual working memory.

The language interference effects shown by MA users in Experiment 2 could therefore have several explanations that are consistent with a visual theory of MA representations. First, verbal interference could have interfered with (linguistic) input and output processes for MA computations. As noted above, we saw anecdotal support for this explanation from pauses in shadowing that corresponded to when participants were outputting the results of MA calculations. Second, verbal shadowing could simply impose a general task-switching or task-monitoring cost because of the complexity of the shadowing task. In neither case would the relatively modest decrement verbal interference effect that we observed be evidence for MA representations being linguistic in nature.

Summarizing this discussion, MA is a representation of exact number that draws on visual (and perhaps motor) resources to complete computations. Inputs and outputs for these computations are often linguistic, but due to their unique structural signatures and their relative resistance to verbal interference, we do not believe that the representations themselves critically depend on language.

*Conclusions*

Over human history—from Babylon, to Roman times, to China and Japan—almost every form of counting board and abacus has organized counters into small sets that fall within the limits of visual working memory. Even apparent exceptions like the Russian schoty prove the rule. The schoty features 10 beads per column, but groups the beads into a 4-2-4 structure using color cues in order to allow users to identify bead configurations easily (Menninger, 1969). Our study suggests that the soroban abacus is optimally designed to take advantage of grouping cues that permit the rapid encoding of objects in visual working memory. Even slight deviations from the soroban structure cause both trained and untrained individuals to be slower when encoding sets.

The mental structure of MA allows users to store multiple columns of beads simultaneously in working memory. By assigning each column a distinct place value, users can represent and manipulate large exact numerosities using visual resources. Although MA interfaces with language, numerical content nonetheless appears to be represented in visual working memory, suggesting that language is not the sole mental format for representing precise numerosities. More generally, the example of MA suggests that humans can make use of a range of cognitive resources in constructing symbolic systems.

**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*(2), 106-111.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637.

Atkinson, J., Francis, M., & Campbell, F. (1976). The dependence of the visual numerosity limit on orientation, colour, and grouping in the stimulus. *Perception*, *5*(3), 335–342.

Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive psychology*, *58*(2), 195–219.

Brooks, N. B., Goldin-Meadow, S., Frank, M. C., & Barner, D. (in prep). The relationship between gesture and mental imagery: a case study of mental abacus representations.

Chen, F., Hu, Z., Zhao, X., Wang, R., Yang, Z., Wang, X., et al. (2006). Neural correlates of serial abacus mental calculation in children: a functional mri study. *Neuroscience letters*, *403*(1-2), 46–51.

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), 1-18.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307-314.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences*, *105*(29), 9926.

Frank, M., Everett, D., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from pirahã language and cognition. *Cognition*, *108*(3), 819–824.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*, 394–406.

Gordon, P. (2004). Numerical cognition without words: Evidence from amazonia. *Science*, *306*(5695), 496.

Halberda, J., Mazzocco, M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572-576.

Hatano, G. (1977). Performance of expert abacus operators. *Cognition*, *5*(1), 47-55.

Hatano, G., Amaiwa, S., & Shimizu, K. (1987). Formation of a mental abacus for computation and its use as a memory device for digits: A developmental study. *Developmental Psychology*, *23*(6), 832–838.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, *15*(1-3), 95-110.

Hishitani, S. (1990). Imagery experts: How do expert abacus operators process imagery? *Applied Cognitive Psychology*, *4*(1).

Hu, Y., Geng, F., Tao, L., Hu, N., Du, F., Fu, K., et al. (in press). Enhanced white matter tracts integrity in children with abacus training. *Human Brain Mapping*.

Kojima, T. (1954). *The japanese abacus: its use and theory*. Tokyo, Japan: Charles E.

Tuttle Company.

Lee, Y., Lu, M., & Ko, H. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, *17*(3), 336–344.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*(2), 467–477.

Luck, S., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–280.

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. *Ethics & Behavior*, *8*(1), 29–67.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*(5695), 499.

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

Stigler, J. W. (1984). Mental abacus: the effect of abacus training on chinese children's mental calculation. *Cognitive Psychology*, *16*(2), 145-176.

Stigler, J. W., Chalip, L., & Miller, K. (1986). Consequences of skill: The case of abacus training in taiwan. *American Journal of Education*, *94*(4), 447–479.

Tanaka, S., Michimata, C., Kaminaga, T., Honda, M., & Sadato, N. (2002). Superior digit memory of abacus experts: an event-related functional mri study. *Neuroreport*, *13*(17), 2187.

Van Oeffelen, M., & Vos, P. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, *10*(4), 396–404.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130-137.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1-B11.