COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2017) 1–22

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12527

The Role of Gesture in Supporting Mental Representations: The Case of Mental Abacus Arithmetic

Neon B. Brooks,[a] David Barner,[b] Michael Frank,[c] Susan Goldin-Meadow[a]

[a]*Department of Psychology, University of Chicago*
[b]*Department of Psychology, University of California, San Diego*
[c]*Department of Psychology, Stanford University*

Received 15 January 2016; received in revised form 4 July 2017; accepted 10 July 2017

Abstract

People frequently gesture when problem-solving, particularly on tasks that require spatial transformation. Gesture often facilitates task performance by interacting with internal mental representations, but how this process works is not well understood. We investigated this question by exploring the case of mental abacus (MA), a technique in which users not only imagine moving beads on an abacus to compute sums, but also produce movements in gestures that accompany the calculations. Because the content of MA is transparent and readily manipulated, the task offers a unique window onto how gestures interface with mental representations. We find that the size and number of MA gestures reflect the length and difficulty of math problems. Also, by selectively interfering with aspects of gesture, we find that participants perform significantly worse on MA under motor interference, but that perceptual feedback is not critical for success on the task. We conclude that premotor processes involved in the planning of gestures are critical to mental representation in MA.

Keywords: Gesture; Mental representation; Spatial transformation; Mental abacus; Motor interference; Visual imagery

1. Introduction

When people talk about how they solved a difficult problem, they often gesture (Goldin-Meadow, 2003; Goldin-Meadow, Alibali, & Church, 1993; Hegarty, Mayer, Kriz, & Keehner, 2005; Tversky, 2011). These gestures not only reflect the speaker's thoughts about a problem (Church & Goldin-Meadow, 1986) but have also been found to influence those thoughts (Goldin-Meadow & Beilock, 2010; Goldin-Meadow, Cook, & Mitchell,

Correspondence should be sent to Neon Brooks, Department of Psychology, University of Chicago, 5848 S. University Ave., Green 510, Chicago, IL 60615. E-mail: neonbluebrooks@gmail.com

2009). For example, encouraging children to gesture when explaining how they solved a math problem can improve their ability to profit from instruction on the problem (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007), and telling adults to gesture when solving mental rotation problems can facilitate spatial transformation (Chu & Kita, 2008, 2011).

The fact that gesture influences task performance demonstrates that the physical and motor representations formed while gesturing interact with mental representations of objects, space, mathematics, and language (Goldin-Meadow & Beilock, 2010). However, it is often challenging to relate the content of gestures directly to their cognitive effects. The gestures that are most frequently studied are produced along with speech, making it difficult to disentangle their cognitive functions from their communicative functions. Gestures produced without speech may offer a more transparent window onto cognitive functioning. However, these gestures lack the framing that speech provides, making it difficult to infer the underlying mental representations with which the gestures co-occur. The goal of this study is to explore how gesture relates to mental representation in a case where gesture is produced without speech, but where the underlying representations are highly constrained and well understood: the case of mental abacus (MA).

Mental abacus is a mental computation technique in which users imagine manipulating beads on an abacus (Menninger, 1969; see Fig. 1). In a typical MA curriculum, students first learn to use a physical abacus and then progress to using the abacus method in the absence of the physical device. MA experts mentally invoke abacus procedures to manipulate a visual image of an abacus (Stigler, 1984). Because MA calculations require a precise set of bead movements performed in a specific order, it is possible to infer the specific sequence of mental states that users represent while solving a problem. MA

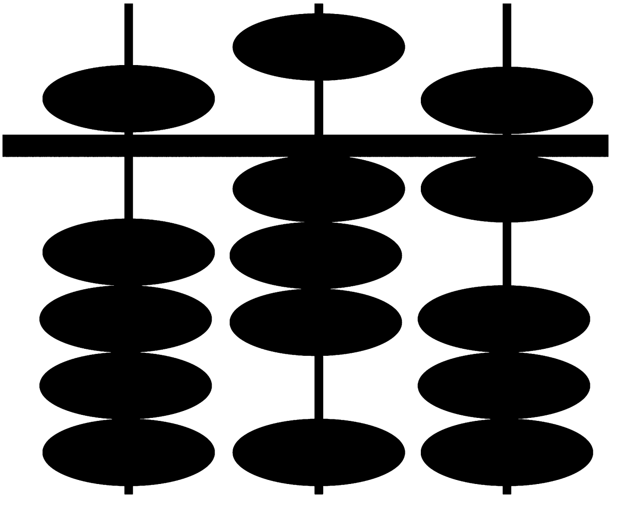

Fig. 1. A graphical depiction of a soroban abacus representing the number 536. Values on a soroban are represented by moving beads toward the center bar. Each column in the abacus represents a place value (e.g., ones, tens, etc.). Beads above the center bar have a value of 5, and lower beads have a value of 1.

training has been shown to improve children's arithmetic abilities (Barner et al., 2016) and numerical processing abilities (Du, Yao, Zhang, & Chen, 2014; Wang, Geng, Hu, Du, & Chen, 2013; Yao et al., 2015), compared to other math curricula.

The MA phenomenon has caught the interest of cognitive scientists, in part, because MA users do not seem to rely on verbal resources when solving arithmetic problems. MA experts can solve arithmetic problems while answering simple verbal questions with no reduction in reaction time (Hatano, Miyake, & Binks, 1977) and are relatively unaffected by verbal shadowing (Frank & Barner, 2012). Further, limits on MA computations are consistent with limits on visual working memory (Frank & Barner, 2012), suggesting that MA is supported by visual resources. This idea gains support from neuroimaging data, which indicate that MA is processed in regions associated with vision and spatial working memory (Chen et al., 2006; Hu et al., 2011; Li et al., 2013; Tanaka, Michimata, Kaminaga, Honda, & Sadato, 2002), as well as spatial orientation and premotor areas (Hu et al., 2011; Li et al., 2013). Together, these findings suggest that when MA users solve arithmetic problems, they perform a specific sequence of manipulations on a visual image of an abacus.

Manual gestures are very common during MA. These gestures appear to reflect the movements used to operate the physical abacus. Although the overlearned nature of these movements distinguishes them from the more spontaneous movements typically studied in the gesture literature, MA gestures also appear to have facilitative effects on cognition: Experts perform significantly worse when they attempt MA with motor interference than without it (Frank & Barner, 2012; Hatano et al., 1977). Because MA gestures reflect movement on a physical device, it is possible to identify the specific mental procedures required to solve problems. This provides an ideal opportunity to investigate the relationship between gesture and mental representations. This study addressed this relationship by testing a population of advanced MA students in Gujarat Province, India.

In past research, we noticed that children exhibit considerable individual differences in the size and amount of gesture produced during MA. It is currently not known how this variability in gesture form relates to the underlying representations of abacus calculations. One hypothesis is that the particular form of MA gesture is irrelevant to MA computations: It may simply be an artifact of past experience using the physical abacus and have no relationship with the cognitive work being done. On the other extreme, the form of an MA user's gesture might closely reflect the structure of actual mental computations and facilitate those computations. Intermediate explanations are also possible: Gesturing might play a role in facilitating computations, but the specific movements involved may not matter for achieving this end. For example, gesture might supply a rhythmic cue that helps users keep their place in a calculation.

In Experiment 1, we investigated the strength of the relationship between gesture form (gesture size and number of steps represented in gesture) and a gesturer's representation of a problem. We found that all children gestured when using MA, and that they systematically increased the size of their gestures as problems became more challenging. Importantly, they also made more gestures on problems that were more difficult for them, suggesting that they were recruiting gesture to facilitate problem-solving. This finding suggests that gesture form is closely related to mental representations of MA calculations, thus allowing us to ask *how* gestures interact with mental representations.

In Experiment 2, we examined three ways that MA gesture might interact with visuospatial representations. First, visual input from gesture could contribute to a visuospatial representation of the abacus, anchoring and constraining the form of the representation in space. Second, proprioceptive input from gesture might contribute to a spatial representation of the abacus by helping users to create and maintain representations of bead locations. A third possibility is that neither visual nor proprioceptive feedback is critical, but that premotor processes involved in planning and preparing for the execution of gesture, rather than the act of gesturing itself, interact with and support visuospatial representations. Consistent with this third possibility, results from Experiment 2 indicate that disrupting the planning of abacus movements, but not inhibiting visual and proprioceptive feedback from gestures, negatively impacts MA performance.

2. Experiment 1

2.1. Methods

In Experiment 1, we asked children to use MA to solve problems that varied in difficulty and analyzed the presence of gesture, size of gesture, and amount of gesture produced. We predicted that if the form of gesture matters for mental abacus problem solving, the size and number of gestures produced might vary systematically as a function of problem difficulty.

2.1.1. Participants

Participants were 226 children (M_{age}: 10.8 years, 32% female) who studied at UCMAS Abacus afterschool programs in Gujarat Province, India. All participants had been practicing MA for over a year. Participants' abacus level ranged from "Higher A," indicating advanced training (generally 27 months or more), to "Grand Level," the highest level included in the program (normally achieved after 41 months). All advanced students and parents at UCMAS centers near Vadodara, India, were invited to participate in a series of studies aimed at investigating the effects of abacus training on cognition. Participants and parents were not told that researchers were interested in gestures or hand movements in particular. Data were collected over the course of two visits: 83 children were tested in the first visit, and 143 different children were tested the following year. Sample size was determined by the number of participants that we were able to recruit and test during each field visit. Two visits were included in order to confirm the findings of the first visit in an independent dataset. We found the same pattern of results in both groups, and we report results based on the combined dataset.

2.1.2. Procedure

The task consisted of addition problems presented on a computer. All problems contained some number of 2-digit addends, but the difficulty of the problems presented was determined adaptively based on the child's performance. Levels 1–3 contained two 2-digit addends, with an increasing number of bead movements required to solve the problems at levels 2 and 3. From level 4 through level 13, the number of 2-digit addends presented increased by 1 for each level (see Appendix S1 in Supporting Information for more details).

Children were asked to solve the problems using their MA and to enter their answers on a keypad. Trials had a 10-s time limit. Each trial was followed by brief feedback indicating whether the child was correct, incorrect, or out of time. The task automatically ended after 10 min. The dependent variable was the child's "threshold" level of performance: the mean difficulty level of all trials completed over the full 10-min period. All trials were videotaped using built-in laptop cameras.

2.1.3. Gesture coding

2.1.3.1. Gesture size: Overall, 3,611 baseline trials were coded for size (see Appendix S2 in Supporting Information for information on how trials were sampled for coding). Each trial was given a code for gesture size on a 4-point scale: 1 for gestures smaller than those used on a physical abacus; 2 for gestures as large as or larger than the physical abacus but not requiring movement of wrists; 3 for gestures that required wrist movement; and 4 for gestures that required movements of the elbows and/or shoulders. When a trial included gestures of more than one size, the code reflected average gesture size across the trial.

Twenty-three percent of trials (821 of 3,611) were coded by more than one coder. Agreement between coders ranged from 54% to 78%, with weighted Cohen's kappas ranging from 0.59 to 0.78. Despite a substantial investment of time in developing a coding system, these reliability rates are relatively low. These low reliabilities reflect the rapid and often undifferentiated movements made by MA experts. Low reliability increases the odds of type II error, but it does not affect interpretation of positive results. Further, findings from the first visit were independently replicated in the second visit, providing evidence that the reported findings are not merely an artifact of coding anomalies.

One reason for the difficulty of the coding task was that coders were blind to the difficulty level of the problem and to all other aspects of children's performance, except for the ordinal position of the trial within the task. Including trial number as an independent variable in analyses further allows us to ensure that any effects of problem difficulty are not driven by coder expectations.

All double-coded trials on which coders disagreed were reviewed, and codes were determined by consensus. Of 2,706 codable trials, 21.8% were coded as 1s, 43.8% as 2s, 18.9% as 3s, and 15.5% as 4s. The mean gesture size was 2.28 (SD = 0.97). Appendix 2 in the Supporting Information provides additional information about the coding process, as well as additional analyses addressing the robustness of findings across coders.

2.1.3.2. Number of moves produced in gesture: We use the term "moves" to refer to the individual up- and down-movements that children *produced* in gesture when doing MA. By contrast, we refer to the individual up- and down-movements that are *required* to solve a problem on the physical abacus as "steps," since these represent the computational steps that one must go through to solve the problem regardless of modality (e.g., adding 10, subtracting 5, subtracting 3).

A total of 2,625 trials were coded for number of individual moves produced in gesture across both datasets.[1] Data were analyzed only for the 1,345 of these trials where the participant computed the correct answer, making it reasonable to assume that their gestures reflect the correct series of steps necessary to solve the problem. For each trial, the coder counted the number of abacus moves the child made in two passes. First, in the "minimum" pass, coders counted only finger movements that clearly represented specific, individual movements of abacus beads. In this pass, compound gestures that could represent moving two sets of beads at the same time (e.g., pinches, moving a top bead down at the same time as moving a bottom bead up) were counted as a single movement. Second, in the "maximum" pass, coders counted every gesture that could reasonably be related to abacus, separating compound gestures like pinches into two moves.

Because coders were counting numbers of moves rather than classifying gestures into discrete categories, we assessed reliability using the intraclass correlation statistic (ICC, Shrout & Fleiss, 1979), which can be interpreted similarly to a Pearson's r, but is a more appropriate measure for comparing measurements of the same phenomenon across two observers (ICC min: 0.69–0.97, ICC max: 0.75–0.96). Again, rates of agreement were somewhat low, as might be expected given the challenging nature of the task.

2.1.4. Measures of problem difficulty

We calculated the objective difficulty of problems in two ways: (a) by problem level (which corresponds, beyond level 3, to number of addends in the problem) and (b) by counting the total number of steps required (i.e., the number of movements that one would have to make on a physical abacus to solve the problem[2]). For example, 10 + 5 is a 2-addend problem that would require two steps: 10 (adding 1 bead to the tens column) + 5 (adding the 5 bead to the ones column). However, 42 + 42 would require 5 steps (see Fig. 2): 40 (adding 4 beads to the tens column) + 2 (adding 2 beads to the ones column) + 50 (adding the 5 bead to the tens column) − 10 (subtracting 1 bead from the tens column) + 2 (adding 1 bead to the ones column).

Since the difficulty of a problem for a given individual depends on his or her ability level, we also created a subjective difficulty measure by subtracting the problem level of each problem from the child's own threshold level. Thus, for a child with a threshold of 8, a level 6 problem would have a difficulty level of −2; but for a child with a threshold of 5 the same problem would have a difficulty level of 1.

Not surprisingly, our subjective difficulty measure was highly correlated with both objective difficulty measures: problem level (Pearson's r = .85) and total number of

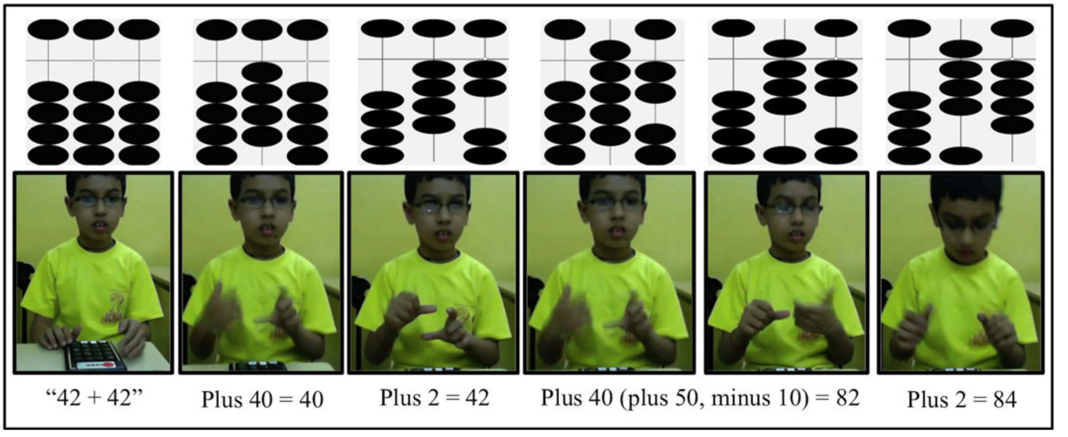

Fig. 2. Example of the gestures produced as a child solves MA problems. The entire problem (42 + 42) was visible on the screen while the child solved the problem. The abacus images show how the gestures reflect the change in the state of the abacus.

steps required (r = .81). In addition, since all trials start from level 1 and increase based on a child's performance, trial number was also highly correlated with subjective difficulty (r = .85) and problem level (r = .81). We include all of these factors in our models and test whether they account for distinct variance, but we also acknowledge that these measures of difficulty cannot be fully disentangled. Accordingly, we report regression outputs with all factors added, including those that do not significantly influence the dependent variable. We also do not explore interactions among the variables as these would be difficult to interpret because of the co-linearity among the measures. Because the trials were presented in a staircase fashion beginning with level 1, and because the first minute was coded for all children in the first dataset, lower problem levels are overrepresented in the sample. In order to approximate a normally distributed variable, we log-transformed the problem level variable for all analyses.

2.2. Results[3]

2.2.1. Gesture frequency

Gesture was nearly universal across children and trials. Coders observed at least some gesture on 95.3% of trials. Moreover, the few trials that did *not* contain gesture were the easiest problems: 77% of trials without gesture were problems with only two addends, problems that are trivial for most MA users. Although MA practitioners can do computations without gesturing (indeed, advanced MA users often do not gesture during competitions), no child in our study chose to do so more than a small fraction of the time, reflecting the importance of gesturing to MA.

2.2.2. Gesture size

We ran a mixed effects regression model predicting size as a continuous variable[4] with subjective difficulty, trial number, log of problem level (our first measure of objective difficulty), and number of steps required (our second measure of objective difficulty) as independent variables, with a random intercept of subject and random slopes by subject for all four independent variables (e.g., a maximal model; Barr, Levy, Scheepers, & Tily, 2013). Here and throughout the paper when conducting mixed effects regressions, we used the $t = z$ approximation to derive p values. The model showed a significant effect of problem level ($\beta = 0.19$, $z = 3.23$, $p < .01$); a marginal effect of subjective difficulty ($\beta = 0.04$, $z = 1.94$, $p = .05$); and no significant effects of trial number ($\beta = 0.001$, $z = 0.81$, $p > .05$) or number of steps required ($\beta = -0.004$, $z = 1.18$, $p > .05$). Gesture size thus increased systematically as the number of addends in a problem increased, and it may have also been influenced by the subjective difficulty of that problem. Fig. 3 shows average gesture size at each problem level for participants who found the problem relatively easy compared to those who found it relatively difficult.[5]

2.2.3. Steps represented in gesture

As one might expect, children produced more moves in gesture on problems that would have required more hand movements on a physical abacus: The number of steps required to solve a given problem was significantly correlated with the number of moves gestured on that problem ($r(1343) = .76$, $p < .01$). However, children frequently produced fewer movements in gesture than the number of steps required to solve the problems, suggesting that some steps were skipped. Not surprisingly, more steps were skipped on problems that required more steps overall. Nevertheless, our analysis (described next) found that, when controlling for the number of steps a problem required, participants

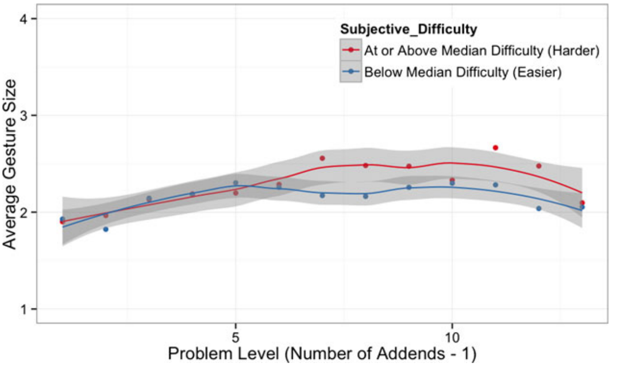

Fig. 3. Gesture size as a function of problem level and subjective difficulty. At each problem level, we calculated the median subjective difficulty level of the problem, defined in relation to a given child's threshold level. Those problems falling at or above the median difficulty for each level are plotted in red; those falling below the median difficulty are plotted in blue. Note that for the statistics reported in the text, subjective difficulty was treated as a continuous variable. It is dichotomized here in order to create an interpretable figure. Lines and shaded regions show estimates and confidence intervals from loess (locally weighted polynomial regression).

produced a larger proportion of the required steps in gesture if the problem was subjectively challenging for them.

To examine the number of moves children produced in gesture on a problem in relation to the number of steps required to solve the problem on the abacus, we analyzed the gestures produced on the 1,345 coded trials in which children correctly solved the problem (we excluded the problems that the children solved incorrectly because we could not be sure of children's underlying mental representations on these problems, and thus could not calculate the proportion of moves produced in gesture). Fifty-eight percent of these problems required more steps than the maximum number of gesture moves coded, suggesting that children often produced gesture moves for only a subset of the number of steps required to solve the problem. In contrast, only 6% of problems required fewer steps than the minimum number of gesture moves coded; these few cases, where more moves were observed in gesture than were necessary for the problem, may reflect coders being overly generous in their gesture counting; children making errors and starting over; or children using gesture for purposes other than indicating bead movements on the abacus. We estimated the proportion of steps that a child produced in gesture on a particular problem by taking the mean of the minimum number of moves coded and the maximum number of moves coded, and dividing that number by the number of steps required to solve the problem. The proportion of steps gestured ranged from 0 (when the participant produced no gestures) to 2.8 (when the mean number of moves coded was greater than the number of steps required; $M = 0.70$, $SD = 0.33$).

We then ran a mixed effects model to predict proportion of steps gestured, including four contributing factors as independent variables: subjective difficulty of the trial, problem level, trial number, and number of steps required. All four factors were treated as fixed, independent variables, along with a random intercept for subject and random slopes for all four independent variables. There were significant effects of number of steps required, subjective difficulty, and trial number, reviewed in the next sections.

2.2.3.1. Number of steps required: The model showed a significant negative effect of number of steps required ($\beta = -0.01$, $z = 6.49$, $p < .01$): Children gestured a smaller proportion of steps for problems that were longer overall. This result is not surprising, as there may be a limit on how many steps a child can gesture in a 10-s period.

2.2.3.2. Subjective difficulty: When controlling for the number of steps required in a problem, there was a positive effect of subjective difficulty ($\beta = 0.01$, $z = 2.16$, $p = .03$): After accounting for the length of the problem, children gestured a larger proportion of problem steps when the problem was more difficult for them. As Fig. 4 shows, for a given number of steps required on the *x*-axis, children who found the problems easier (in blue) gestured a smaller proportion of steps than those who found the problems more difficult (in red). The beta value of 0.01 indicates that for each level of difficulty, children produced gestures for an additional 1% of the moves in the problem. For example, they produced approximately 10% more gestures for problems at a difficulty level of 5, compared to trials at a difficulty level of −5.

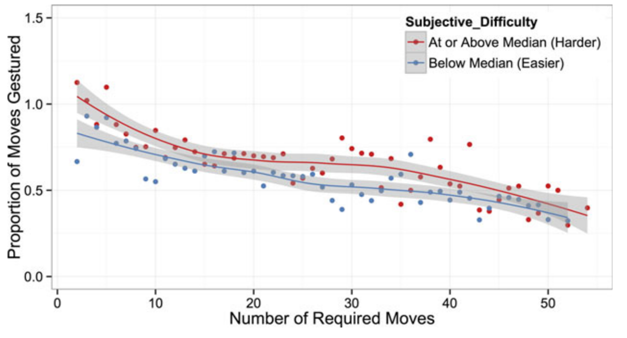

Fig. 4. Proportion of steps gestured as a function of number of steps required. For problems with each number of steps required, we calculated the median subjective difficulty level of the problem, defined in relation to a given child's threshold level and the level of the problem. Those problems falling at or above the median difficulty at each number of required steps are plotted in red; those falling below the median difficulty are plotted in blue. Note that, for the statistics above, subjective difficulty was treated as a continuous variable. It is dichotomized here for visualization purposes. Lines and shaded regions show loess estimates and confidence intervals.

2.2.3.3. Trial number and problem level: The model also showed a small but significant positive effect of trial number ($\beta = 0.001$, $z = 2.39$, $p = .02$), suggesting that children gestured a larger proportion of moves on later trials. There was no effect of problem level on proportion of steps gestured in this model ($\beta = 0.0005$, $z = 0.06$, $p = .96$).

Although the best predictor of the proportion of steps gestured was the raw number of steps in the problem, when we controlled for this variable, we found that children acted out a larger proportion of steps for problems that were *more difficult* for them. These findings thus provide new evidence that MA experts may be recruiting specific gestures to help them solve problems.

2.3. Summary

The findings from Experiment 1 show that gesture is common during MA arithmetic and reflects the specific steps required to solve a problem. Specifically, we found that (a) the size of children's gestures and (b) the number of problem steps a child gestures, both vary systematically as a function of the objective and subjective difficulty of the problem. Our first result—that gesture size varies as a function of problem difficulty—indicates that participants actively adapted their gestures depending on the problem, rather than simply moving their hands in an arbitrary or habitual manner. However, it is not clear precisely *how* gesture size relates to the underlying representation of the abacus. Our second finding—that children produce a greater proportion of problem steps in gesture on increasingly difficult problems—addresses this issue and demonstrates a close relationship between the form of the gesture and the problem itself. When solving difficult MA problems, children more closely replicated in gesture the movements they would have produced on a physical abacus compared to when they solved easier problems, suggesting that they were relying on these movements to help them solve the problems. Because we examined proportion of moves in gesture and controlled for the absolute number of steps in the problem, it is unlikely that this result can be explained by other variables such as the amount of time it took to solve the problem.

3. Experiment 2

3.1. Methods

In Experiment 2, we asked how gesture influences MA calculation by limiting the types of feedback available to MA experts as they solved problems. Visual feedback and proprioceptive feedback from enacting hand movements could enrich the gesturer's visuospatial representations of the abacus, making those representations easier to maintain and/or update. Alternately, the benefit of gesturing may come not from feedback participants get from producing movements, but from premotor activation or other higher level cognitive processes involved in planning movements. If this is the case, producing gesture itself may not be critical for MA performance, as long as MA experts are not prevented from planning gestural movements. These possibilities are not mutually exclusive: planning, executing, and receiving feedback from gestures may all facilitate MA performance. To test the relative importance of each of these factors, we sequentially eliminated each one: (a) We eliminated visual feedback, but preserved proprioceptive feedback and motor planning, by having children wear a blindfold while doing MA; (b) we eliminated visual feedback and proprioceptive feedback, but preserved motor planning, by instructing children to keep their hands flat on the table while doing MA; (c) we disrupted motor planning, along with visual and proprioceptive feedback, by having children perform a motor interference task while doing MA. We compared results of abacus experts to those of control children, tested in the United States, who were naive to abacus. In order to eliminate the possibility that non-MA gestural strategies (i.e., counting on one's hands) could explain behavior in the control group, we excluded any control participant who was observed to count on their fingers during the task ($N = 8$ out of 32).

3.1.1. Participants

Twenty-nine abacus experts (M_{age}: 11.1 years, 29% female), drawn from the same population as Experiment 1, participated in the experiment. Four children did not complete all of the tasks and thus were excluded from analysis, leaving 25 children in the final dataset. Sample size was estimated based on a power analysis of past motor interference data on the same population. Thirty-two control children, who were naive to abacus, were tested in the United States. Because finger counting was very common among younger control children, we limited our control sample to children who were at least 12 years of age. Eight children in this group were excluded because they spontaneously counted on their fingers during the Baseline task. Another participant was excluded because his left hand was in a cast and he was unable to comply fully with the motor interference manipulation. The remaining 23 control participants had a M_{age} of 12.4 years (30% female).

3.1.2. Procedure

3.1.2.1. Expert participants: Children were tested individually at an abacus school or, for the control participants, in a laboratory testing room. Each task required children to solve arithmetic problems on the computer. Numbers were presented in an auditory format so that participants did not have to see the addends to solve the task, allowing us to manipulate visual feedback using blindfolds. On each trial, children heard an automated voice read six addends aloud through headphones. Children were asked to add the numbers using MA and to verbally report their answer to an experimenter. Each trial had a 30 s time limit, which was reached on fewer than 1% of trials overall.[6] The trial level varied as a function of performance on the previous trial: If a child got a problem correct, she moved up a level on the subsequent problem. If she got the problem incorrect, she moved down a level. Difficulty was manipulated by changing the value of the addends in the problem. At level 1, all addends had a value of 3 or less (e.g., a child might hear "two, two, three, one, two, one"). The maximum addend size for subsequent levels was the level cubed; for example, at level 2, the addends had a maximum value of 2^3, or 8; at level 10, the maximum value was 10^3, or 1,000. This system was used in order to quickly and accurately capture children's varied ability levels. After each trial, the screen displayed feedback on the child's accuracy. The task ended automatically after 10 min.

Each child participated in four tasks, presented in a random order:

1. *Baseline task.* Children solved the problems as they typically would.
2. *Blindfold task.* Children were blindfolded as they solved problems; the experimenter read the words "correct" and "error" off the screen after each trial to provide feedback.
3. *No Hands task.* Children were instructed to keep their hands flat on the table as they solved problems; if they lifted their fingertips off the table, they were reminded to keep their hands flat. Movement of the hands that did not disturb the location of the fingertips, such as shaking of palms, was allowed.
4. *Motor Interference task.* Children were instructed to tap on the keyboard to maintain the length of a red bar, which was presented at the top of the computer screen, as they solved the problems using MA. If the child failed to tap on the keys on the home row of the keyboard (in any order) as the numbers were presented, the bar would decrease in length and eventually disappear, causing the trial to end and an error message to appear.

3.1.2.2. Control participants: The procedure was identical for the control participants tested in the United States, except for changes made to adjust for the control children's arithmetic skills, which were dramatically less advanced than those of the MA experts. Control participants were asked to add three addends rather than six. In addition, the maximum addend value at each level was determined by raising the level number to the exponent 1.2, instead of 3. For example, a level 5 problem for an abacus expert contained numbers between 1 and 125; in contrast, for a control participant, a level 5 problem contained numbers between 1 and 7. These adjustments made it possible for participants in both groups to succeed on the tasks, allowing us to examine the impact of our three factors on performance across the two groups. Critically, no changes were made to the experimental manipulations for the control participants.

3.1.2.3. Analysis: Following Frank and Barner (2012), we calculated each child's "threshold" level on each task by taking the average level reached across all trials. To account for differences in time to understand each task, we excluded any incorrectly answered trials that occurred before the first correct answer. In addition, to make the data comparable across tasks, we determined the minimum number of trials a child completed on each of the four tasks and truncated the data so that this number of trials was used to compute the child's threshold for all four tasks. The number of trials on each task examined per child ranged from 7 to 35 ($M = 25.2$, $SD = 6.41$); all children except two completed at least 20 trials on each task.

3.2. Results

Our analyses below indicate that abacus experts were only minimally affected by a lack of visual or proprioceptive feedback, but performed far worse when not permitted to plan motor movements: Thresholds were significantly lower for the motor interference task than the baseline task. Fig. 5 (left panel) presents the mean problem level achieved on each of the first 20 trials in each of the four tasks. The Baseline, Blindfold, and No Hands tasks cluster together and are considerably higher than the Motor Interference task. The mean threshold level on the Baseline task was 5.20 (which corresponds to a maximum addend value of 141 at threshold), compared to 5.30 on the Blindfold task (maximum addend value = 149), 4.69 on the No Hands task (maximum addend value = 103), and 2.78 on the Motor Interference task (maximum addend value = 21).

In a linear mixed effects model with Threshold as a dependent variable, Task as a fixed independent variable, and random task slopes and intercept terms by subject, we found that thresholds were significantly worse on the Motor Interference task compared to Baseline ($\beta = -2.47$, $z = 7.03$, $p < .01$). Thresholds on the No Hands task were marginally lower than Baseline ($\beta = -0.46$, $z = 1.71$, $p = .09$), and thresholds on the Blindfold task were not significantly different from Baseline ($\beta = 0.18$, $z = 0.90$, $p = .37$). Performance on the Motor Interference task was also significantly different from performance on the Blindfold task ($\beta = 2.65$, $z = 7.20$, $p < .01$) and the No Hands task ($\beta = 2.02$, $z = 5.98$, $p < .01$). Adding child age, gender, or the order in which the task appeared[7] did not improve the fit of the model ($ps > .10$).

3.2.1. Control participants

Fig. 5 (right panel) presents the mean level on the first 20 trials on each of the four tasks for the children with no abacus experience. There were no significant differences between any of the tasks for control participants, indicating that the manipulations had

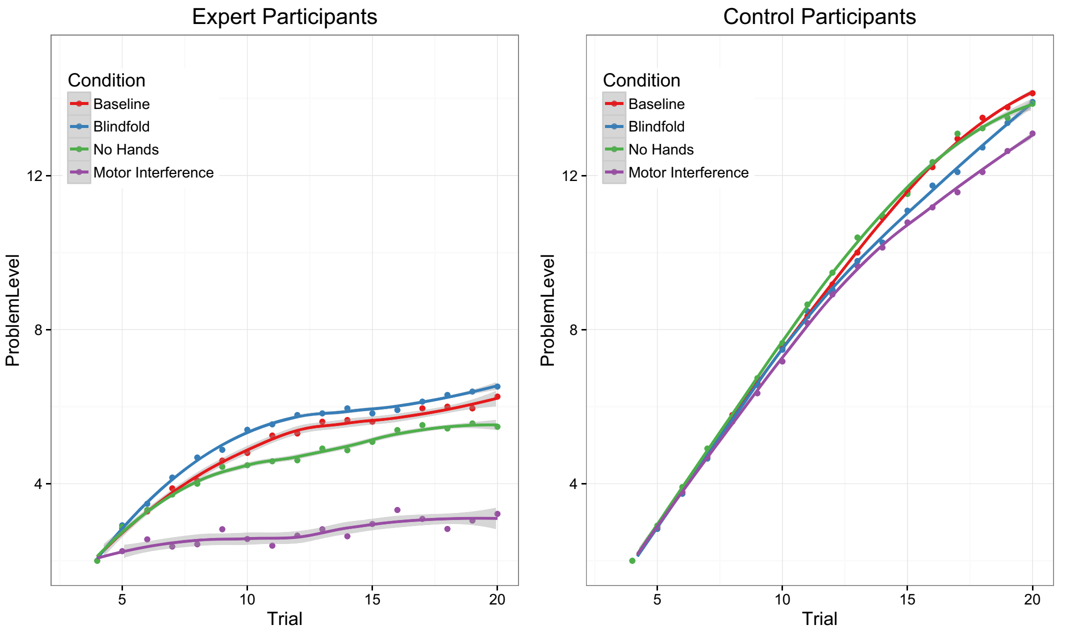

Fig. 5. Mean problem level on the first 20 trials for each of the four tasks in abacus experts (left) and control participants (right). Lines and shaded regions show estimates and confidence intervals from a local regression (loess). Level of performance is not directly comparable for the control participants and the abacus participants: Control participants were given shorter problems with smaller addends at all levels. However, the relations among the curves for the four tasks can be compared across participants. Note that analyses were conducted using participants' threshold level as the dependent variable, while these plots provide additional information showing how performance changed across trials. See Supplementary Materials for a bar graph of threshold level across tasks.

minimal impact on children who do not use a manual strategy to solve arithmetic problems. The number of trials analyzed per task for each child ranged from 13 to 39 (M = 30.0, SD = 6.5). Only two participants had fewer than 20 trials. Mean threshold level was 12.38 (maximum addend size = 21.11) for Baseline; 12.43 (maximum addend value = 21.23) for Blindfold; 12.51 (maximum addend value = 21.39) for No Hands; and 11.75 (maximum addend value = 19.91) for Motor Interference.

In a linear mixed effects model with Threshold as a dependent variable, Task as a fixed independent variable, and random task slopes and intercept terms by subject, we found no significant differences between the Baseline task and the Motor Interference ($\beta = -0.56$, $z = 1.39$, $p = .16$), Blindfold ($\beta = 0.07$, $z = 0.21$, $p = .83$), or No Hands ($\beta = 0.16$, $z = 0.41$, $p = 0.69$) tasks. The Motor Interference task also did not differ significantly from the Blindfold task ($\beta = 0.63$, $z = 1.53$, $p = .13$) or the No Hands task ($\beta = 0.73$, $z = 1.81$, $p = .07$). Adding child age, gender, or the order in which the tasks were performed did not significantly improve model fit (all ps > .10).

3.2.2. Comparing MA experts and control participants

Because the experts and control participants solved problems at very different difficulty levels, it is impossible to directly compare the results of the two groups. However, in order to confirm that the effects of motor interference on performance were specific to experts, we tested for an interaction between task and population (Indian experts or US control participants) using z-scored threshold levels for each participant on each task. First we created z-scores for mean threshold for each participant on each task separately for US controls and for Indian experts; we then combined these datasets and conducted a linear model predicting z-scored threshold by task, population (experts vs. controls), and their interaction, with a random fixed effect of subject. This model showed a significant simple effect of the Motor Interference task compared to Baseline ($\beta = -0.98$, $z = 8.26$, $p < .01$), but also, critically, a significant interaction between the Motor Interference task and population: US controls showed significantly less interference than Indian experts, compared to performance on the Baseline task ($\beta = 0.78$, $z = 4.55$, $p < .01$).[8]

3.3. Summary

The findings from Experiment 2 suggest that MA experts' reliance on gesture stems primarily from cognitive and/or premotor processes involved in planning gestural movements, rather than from the visual and proprioceptive feedback that arise when gestures are actually executed. Disrupting motor planning with a motor interference task severely disrupted MA performance, whereas allowing participants to plan gestures while preventing their production (thereby eliminating any proprioceptive or visual feedback) had a much more limited effect, on performance, if any. Although there was a trend for experts to perform somewhat worse than Baseline on the No Hands task, the effect was not significant and was far less drastic than the effect in the Motor Interference task. In control participants, none of the four tasks had a disruptive effect on performance, and, importantly, there was a significant interaction between the control and expert populations, demonstrating that the MA experts were uniquely affected by the Motor Interference task. These findings suggest that cognitive and premotor processes involved in planning gestural movements play a critical role in MA, and that any additional cognitive benefit of seeing or feeling these movements is minimal. The data from control participants demonstrate that the large effect of motor interference on MA experts' performance does not stem from the Motor Interference task simply overloading executive resources. The control participants, who did not use a motor strategy to solve addition problems, did not perform significantly worse under motor interference than on any of the other tasks, suggesting that the task was an additional cognitive load only for children who solved the problems using a motor strategy.

Although our results rule out visual and proprioceptive feedback as primary factors explaining gesture's effect on MA, they leave open a number of potential cognitive and premotor mechanisms that could play a role. One possibility is that planning the motor movements that are reflected in MA gesture creates a detailed representation of the changes to the abacus in premotor and motor cortex. These motor representations may then serve as a second code, which augments visuospatial representation of problems, improving accuracy and speed of calculations, perhaps through the construction of a forward model of the anticipated consequences of planned movement (Wolpert, Ghahramani, & Jordan, 1995). Asking participants to produce movements that are inconsistent with these plans leads to a decrement in performance. Alternatively, MA gesture may reflect higher level spatial cognitive processes, such as spatial working memory. These spatial representations may be disrupted if they receive motor signals that are not consistent with the information they represent.

We cannot rule out the possibility that proprioception plays some role in facilitating MA representations: Participants may have made undetectably small hand movements, or they may have moved other parts of their body even in the No Hands task (though these movements would likely lack the level of detail that gestures in the baseline task contained). Although it seems more likely that premotor activation involved in the production of gestures drives the success of participants on the No Hands task, it is nevertheless possible that these movements create sufficient proprioceptive feedback to influence MA directly. The fact that participants produce larger gestures during more difficult problems (as found in Experiment 1) may also suggest that either visual or proprioceptive feedback plays some useful role in MA calculation. However, other explanations for this effect (e.g., that larger gestures reflect stronger premotor representations) are also consistent with the data. Finally, the motor interference task not only eliminates motor planning of abacus gestures, but also provides contradictory proprioceptive feedback, and this contradictory feedback could be part of the reason children do so poorly. In sum, the success of MA experts on the No Hands task rules out the possibility that proprioceptive feedback from large-scale movements is necessary for MA, although it remains possible that proprioception from small movements could account for some of the observed motor interference effect.

4. General discussion

Although gesture has been shown to be a powerful cognitive tool across a wide variety of domains, understanding how externally produced gestures relate to, and influence, internal mental representations has presented a challenge for researchers. Using the case of MA, where gestures are produced without speech but where the underlying representations are highly constrained and well understood, we have discovered novel insights into the ways that gesture can interact with spatial representations in problem-solving.

In Experiment 1, we found that MA users produced gestures on nearly every problem, and that the content of their gestures was related to both the problem at hand and to the difficulty of that problem for the gesturer. We found that the size of gestures increased as problems became more difficult, as did the number of problem steps represented in gesture. These results demonstrate that the form of gestures can play a role in MA problem-solving, ruling out the possibility that gesture influences MA calculation simply by modulating attention or serving as a timekeeping device. This finding demonstrates that gesture is capable of supporting accurate mental representations of number while performing MA. This result is likely to generalize to other tasks that involve procedural motor strategies, such as mental rotation (e.g., Chu & Kita, 2008, 2011) and explanations of Tower of Hanoi problems (e.g., Trofatter, Kontra, Beilock, & Goldin-Meadow, 2015).

In Experiment 2, we asked *how* the information represented in gesture influences internal mental representations. Results from the Blindfold and No Hands tasks demonstrate that visual and large-scale proprioceptive feedback from gestural movements does not play a critical role in experts' MA computation: removing this feedback did not have a significant effect on performance. By contrast, we found that the same participants performed significantly worse on the Motor Interference task than on any of the other tasks, all of which allowed for the planning of motor movements. Together, these results suggest that actual production of physical gestures may not be critical to MA, but that computations are supported by cognitive and/or premotor processes involved in planning gestures. Critically, control participants who did not use MA did not show this pattern of effects, demonstrating that the tapping task we used interferes specifically with MA calculation, rather than affecting calculation through a more general mechanism. Interestingly, research by Wohlschläger (2001) suggests a comparable effect in mental rotation tasks: When participants were asked to plan, but not execute, hand movements that were inconsistent with the direction of a mental rotation problem, their performance on the mental rotation task was compromised. In our case, participants were asked to perform tapping movements that prevented them from producing the gestures they had planned, which led to a decrement in performance on the calculation task. This finding is also consistent with recent research showing similar cognitive effects of gesture on a Tower of Hanoi task when participants are blocked from receiving any visual feedback from their gestures (Cooperrider, Wakefield, & Goldin-Meadow, 2015).

Studies 1 and 2 together show that planning gestures that reflect specific bead movements can facilitate mental representations of MA. How does gesture planning have this impact on performance? One possibility—suggested by Hegarty et al. (2005)—is that planning gestures can facilitate the mental animation of static images. When solving problems about mechanical systems based on static diagrams, participants showed no decrement in performance when not permitted to gesture, but they performed significantly worse when required to tap their fingers in a spatial pattern (Hegarty et al., 2005). Taken together, the Hegarty et al. (2005) findings and our own findings from Experiment 2 suggest the possibility that planning gesture may have a widespread benefit for any task that requires visually tracking objects as they undergo transformations. Motor plans produced during MA may provide an additional representation of the positions of beads as they move, or they may strengthen existing movement representations by informing estimates of bead locations as the state of the abacus changes.

Another possible way in which motor planning might facilitate MA computation is by providing a format for storing and recalling information about specific MA procedures. For example, adding 9 on the abacus typically requires adding 1 to the tens column and subtracting 1 from the ones column. Over time, planning the finger movements associated with this procedure may provide a secondary code to rely on when enacting changes on the MA.

If this is the case, the benefits of gesturing on performance should be stronger for experts than for abacus novices, and they should also be stronger for gesturers who produce more complex finger movements, enacting multiple steps at once in their gestures. If this mechanism is responsible for gesture's facilitation of MA, we might not expect to see similar patterns of facilitation on tasks that do not involve a well-practiced motor component.

Interestingly, the relationship between visuospatial representations and gesture that we observe in MA does not seem to apply to all cases where gesture is produced. A large body of research has found that performance on many tasks can be impaired when gesturing is inhibited even if it is not replaced with a secondary task (as in our No Hands task). For example, participants who are prevented from gesturing produce speech that is less fluent (Pine, Bird, & Kirk, 2007; Rauscher, Krauss, & Chen, 1996), less rich in imagery (Alibali, Spencer, Knox, & Kita, 2011; Rimè, Shiaratura, Hupet, & Ghysselinckx, 1984), and less focused on perceptually present information (Alibali & Kita, 2010) than when they are permitted to gesture. Further, when participants are prevented from gesturing as they explain their solutions to math problems, they remember fewer items on a concurrent working memory tasks (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner, Nusbaum, & Goldin-Meadow, 2004) than when they produce gesture. This raises the question: Why does inhibiting gesturing impact performance in some cases, like explaining math problems, but not in others, like MA?

There are a few relevant task differences that could explain the different roles gesturing plays across these tasks. First, the gestures MA experts produce reflect the movements that would be used to calculate on a physical abacus. In contrast, many of the gestures produced in studies where inhibiting gesturing itself impacted performance were metaphorical, reflecting different strategies for solving problems (e.g., Goldin-Meadow et al., 2001); or were produced during natural conversation (e.g., Rimè et al., 1984). Further, the gestures MA experts produce correspond to heavily practiced movements and could therefore be richly associated with specific visuospatial procedures and outcomes. Practice may allow the benefits of gesturing to be internalized in experts so that the gestures themselves need not actually be produced. In contrast, the gestures produced when participants explain a math problem or describe a scene tend to be generated on the spot. However, both Hegarty et al. (2005) and Wohlschläger (2001) found similar results with movements that were not overlearned. Practice may strengthen the relationship between motor planning and action, but it is unlikely to fully account for it. Thus, either or both of these features—the degree to which a gesture is concrete or metaphorical, or the degree to which a gesture is (or represents) a well-practiced movement—could influence how important actually producing gesture is to cognition.

Alternatively, the critical difference between paradigms may not be the nature of the gestures produced, but the consequences of telling participants not to gesture. Constructing motor plans for gesturing may always be sufficient to change mental representations, but in some contexts, people may not plan gesture unless they actually produce gesture. Unlike MA, the studies that show effects of inhibiting gesture all involved gestures produced together with speech. When people are told not to gesture while speaking, they may also fail to plan physical movements, even if planning these movements would have been helpful to them. Indeed, gesturing without speech is relatively rare and may only occur when spatial resources are especially taxed. Perhaps only in extreme cases, like MA and mechanical reasoning, will participants continue to plan gestures even when told not to execute them.

In sum, using MA as a case study, we examined how gestures interface with visuospatial mental representations. We found that gesture is a robust component of MA, and that the gestures experts produce as they add sums on a MA offer a window onto their computations. Our results indicate that these gestures do not merely reflect computation—they actively interface with visuospatial representations of the abacus, although it is planning the gestures, rather than seeing or feeling them, that has the most influence on cognition. These findings provide new evidence that cognitive and premotor processes involved in the planning of hand movements, even when they are not enacted, can play a critical role in spatial reasoning. More research is needed to fully understand the processes by which gesture planning influences spatial reasoning, and to understand whether gestures that accompany speech contribute to reasoning through similar processes.

Acknowledgments

We thank Snehal Karia and the staff, students, and parents of UCMAS Gujarat for their vital contributions to this research. We also thank George Alvarez for his guidance and support. Thank you to Mahesh Srinivasan, Jessica Sullivan, Eleanor Chestnut, Talia Konkle, Ally Kraus, Jonathan Gill, Katharine Tillman, Ann Nordemeyer, Molly Lewis, Pierina Cheung, Ali Horowitz, Bria Long, Jordan Suchow, Viola Störmer, and Sean Barner for help with data collection; and to Hannah Fullmer, Michael Helizer, Chris Kyriazis, Samantha Chen, Emily Escovar, Anna Jones, Asher Stuhlman, and Leyla Tarhan for help with gesture coding. This work was supported by NSF REESE grant #0910206 (D.B. as co-PI), by NSF Science of Learning Center funding SBE-0541957 (S.G.M. is a co-PI) and by NICHD grant R01-HD47450 to S.G.M.

Notes

1. See Appendix 2 in the Supporting Information for information on how trials were sampled for each analysis.
2. For this measure, compound steps (e.g., adding 8 by pushing down the five bead and pushing up 3 ones beads) were counted as the number of individual finger movements required to add the numbers (i.e., adding 8 would count as two steps).
3. All code and data for both studies are available at https://github.com/langcog/abacus_gesture.
4. Because size is not truly a continuous variable, we also ran a logistic regression with size (1–2 vs. 3–4) as a binomial dependent variable, with the same four independent variables and random slopes for each variable. This model showed a significant effect of subjective difficulty ($\beta = 0.33$, $z = 2.11$, $p = .04$) but no other

significant effects. If anything, this model more strongly supports the effect of subjective difficulty.

5. See Appendix S3 of the Supporting Information for additional analyses.
6. The time limit in Experiment 2 was much longer than in Experiment 1 in order to allow time for the experimenter to correctly enter the child's answer.
7. Task order was analyzed by considering the ordinal number (1–4) of the task in the four-task sequence. This analysis does not account for potential interactions between specific tasks as our design did not allow for inferences about such interactions (there were 24 possible task orders).
8. Because the experts also spanned a greater age range than the novice participants, we repeated the comparison analysis including only the eight expert participants who were 12 years of age or older. Despite the small sample size, this model also yielded a significant interaction between the motor interference task and population ($\beta = 1.07$, $t = 4.58$, $p < .01$), suggesting that the inferior motor interference performance of the experts was not driven by the age of control participants.

References

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture*, *10*(1), 3–28.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138–1144.

Barner, D., Alvarez, G., Sullivan, J., Brooks, N., Srinivasan, M., & Frank, M. C. (2016). Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development*, *87*(4), 1146–1158.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 539–550. https://doi.org/10.1037/0096-3445.136.4.539.

Chen, F., Hu, Z., Zhao, X., Wang, R., Yang, Z., Wang, X., & Tang, X. (2006). Neural correlates of serial abacus mental calculation in children: A functional MRI study. *Neuroscience Letters*, *403*(1–2), 46–51.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*(4), 706–723. https://doi.org/10.1037/a0013157.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71.

Cooperrider, K., Wakefield, E., & Goldin-Meadow, S. (2015). More than meets the eye: Gesture changes thought, even without visual feedback. In R. Dale, C. Jennings, P. P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 441–446). Austin, TX: Cognitive Science Society.

Du, F., Yao, Y., Zhang, Q., & Chen, F. (2014). Long-term abacus training induces automatic processing of abacus numbers in children. *Perception*, *43*(7), 694.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141*(1), 134–149. https://doi.org/10.1037/a0024427.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*(2), 279–297.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674. https://doi.org/10.1177/1745691610388764.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–272. https://doi.org/10.1111/j.1467-9280.2009.02297.x.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516–522.

Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition*, *5*(1), 47–55. https://doi.org/10.1016/0010-0277(77)90016-6.

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation*, *5*(4), 333–356.

Hu, Y., Geng, F., Tao, L., Hu, N., Du, F., Fu, K., & Chen, F. (2011). Enhanced white matter tract integrity in children with abacus training. *Human Brain Mapping*, *32*, 10–21.

Li, Y., Hu, Y., Zhao, M., Wang, Y., Huang, J., & Chen, F. (2013). The neural pathway underlying a numerical working memory task in abacus-trained children and associated functional connectivity in the resting brain. *Brain Research*, *1539*, 24–33.

Menninger, K. (1969). *Number words and number symbols: A cultural history of numbers*. Cambridge, MA: MIT Press.

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747–754.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226–231.

Rimè, B., Shiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*, 311–332.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420.

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

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.

Trofatter, C., Kontra, C., Beilock, S., & Goldin-Meadow, S. (2015). Gesturing has a larger impact on problem-solving than action, even when action is accompanied by words. *Language, Cognition and Neuroscience*, *30*(3), 251–260.

Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science*, *3*(3), 499–535.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*(4), 395–407. https://doi.org/10.1016/j.jml.2004.01.002.

Wang, Y., Geng, F., Hu, Y., Du, F., & Chen, F. (2013). Numerical processing efficiency improved in experienced mental abacus children. *Cognition*, *127*(2), 149–158.

Wohlschläger, A. (2001). Mental object rotation and the planning of hand movements. *Attention, Perception, & Psychophysics*, *63*(4), 709–718.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880.

Yao, Y., Du, F., Wang, C., Liu, Y., Weng, J., & Chen, F. (2015). Numerical processing efficiency improved in children using mental abacus: ERP evidence utilizing a numerical Stroop task. *Frontiers in Human Neuroscience*, *9*, 245.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1: Additional procedural information.

Appendix S2: Additional Gesture Coding Details.