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

Neon B. Brooks[1], David Barner[2], Michael Frank[3], Susan Goldin-Meadow[1]

[1]University of Chicago

[2]University of California, San Diego

[3]Stanford University

Contact Information:

Neon Brooks, Department of Psychology, Harvard University

33 Kirkland St., Cambridge, MA 02138; neonbrooks@fas.harvard.edu

Acknowledgements: 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 (SGM is a co-PI) and by NICHD grant R01-HD47450 to SGM.

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 works is poorly understood. We investigated this question by exploring the case of Mental Abacus (MA), a technique in which users imagine moving beads on an abacus to compute sums, while moving their hands as though using an abacus. Because the content of MA is transparent and readily manipulated, it offers a unique window into 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 motor planning is critical for MA, but perceptual feedback is not. We conclude that planning gestures – rather than seeing or feeling them – is critical to mental representation in MA.

INTRODUCTION

When people solve difficult problems, they often move their hands (Goldin-Meadow, Alibali & Church, 1993; Goldin-Meadow, 2003). Gesturing can affect performance on a variety of tasks (Goldin-Meadow & Beilock, 2010; Goldin-Meadow, Cook, & Mitchell, 2009; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014). For example, encouraging children to gesture when explaining how they solved math problems can improve their ability to learn from instruction (Broaders, Cook, Mitchell, & Goldin-meadow, 2007); And gesturing about mental rotation appears to facilitate spatial transformation (Chu & Kita, 2011).

The fact that gesture impacts 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 and communicative functions. Gestures produced without speech may offer a more transparent window into cognitive functioning. However, these gestures lack the framing that speech provides, making it difficult to infer the underlying mental representations with which gestures co-occur. The goal of the current 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).

MA is a mental computation technique in which users imagine manipulating beads on an abacus (Menninger, 1969, see Figure 1). In a typical MA curriculum, students first learn to use a physical abacus, and progress to using the abacus method in the absence of the physical device. MA experts mentally invoke these same 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.

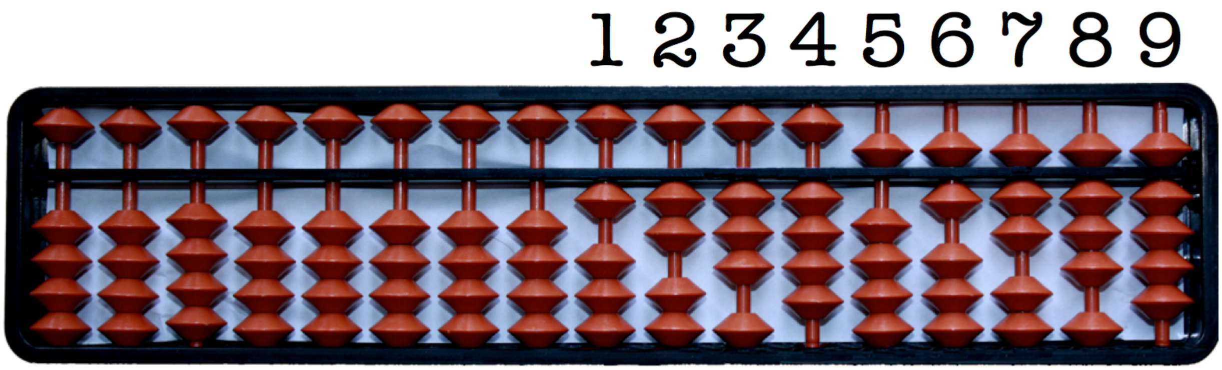

Figure 1. An abacus displaying the number 123,456,789. Values are represented on the abacus by moving beads toward the black center bar. Each column in the abacus represents a place value (e.g. ones, tens, etc.). Beads above the center bar have a value of 5, and lower beads have a value of 1.

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 they are supported by visual resources. This idea is supported by 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; Tanaka, Michimata, Kaminaga, Honda, & Sadato, 2002). 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, and appear to facilitate MA calculation: children perform significantly worse when they attempt MA under motor interference (Hatano et al., 1977; Frank & Barner, 2012). Because we can independently determine their mental representations when they correctly solve problems, MA provides an ideal opportunity to investigate the relationship between gesture and mental representations. The present study addressed this relationship by studying a population of advanced MA students in Gujarat Province, India.

It is not known how the form of abacus gestures relates to the underlying abacus calculations. One extreme hypothesis is that the form of MA gesture is irrelevant to MA computations. More plausibly, gesture could serve a very general role in computation, providing, for example, a rhythmic cue that helps users keep their place in a calculation. In Study 1, we investigated the relationship between gesture form (size and number of moves produced) and the gesturer's representation of the 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 confirms that gesture provides a window into MA calculations, thus allowing us to ask how gestures interact with mental representations.

In Study 2, we examined three ways that MA gesture might interact with visuospatial representations. First, visual input from gesture could directly 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. A third possibility is that neither visual nor proprioceptive feedback is critical, but that planning gesture provides motor representations of the abacus that interact with visuospatial representations. Consistent with this third possibility, results from Study 2 indicate that motor planning of abacus movements, but not feedback from gestures, plays an active role in MA problem-solving.

STUDY 1

Methods

In Study 1, we analyzed the presence of gesture, size of gesture, and amount of gesture produced on problems that varied in difficulty for individual children. We predicted that if the form of gesture matters for mental representation of the abacus, the size and number of their gestures should vary systematically as a function of problem difficulty. Study 1 tested this prediction.

Participants

Participants were 226 children (mean age: 10.8 years, 32% female) who studied at UCMAS Abacus afterschool programs in Gujarat Province, India. Data were collected over the course of two visits: 83 children were tested in the first visit, and 143 were tested the following year. Sample size was determined by the number of participants we were able to recruit and test during a 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 report results based on the combined dataset.

Procedure

The task consisted of addition problems presented on a computer. All problems contained 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 Supplemental Material for details).

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

Gesture Coding

Gesture Size

Overall, 3,611 baseline trials were coded for size (see Supplemental Material for information on how trials were sampled). 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 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 gesture sizes were averaged 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 .59 to .78. Although these reliability rates are relatively low, coders were blind to the difficulty level of the problem and to all other aspects of children' performance, except for the ordinal position of the trial within the task. All double-coded trials with disagreements 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).

Number of Moves Produced in Gesture

A total of 2,625 trials were coded for number of moves across both datasets. For each trial, the coder counted the number of abacus gestures 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 representing 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 conceivably 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 by 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: .69 - .97, ICC max: .75 - .96). Again, rates of agreement were somewhat low, as might be expected given the challenging nature of the task. However coders were blind to problem level and difficulty, so lower rates of reliability can only contribute to type I error. All trials with disagreements of at least 3 moves were reviewed and given a resolved move score. Disagreements of fewer than 3 moves were either recoded or resolved in favor of the coder whose reliability was higher across all combinations of coders.

Measures of Problem Difficulty

We calculated the Objective Difficulty of problems in two ways: (1) by Problem Level (which corresponds, beyond level 3, to number of addends in the problem), and (2) by counting the Total Number of Required Moves (i.e., moves that needed to be made on a physical abacus to solve the problem[1]). For example, 10 + 5 is a 2-addend problem that would require two moves: 10 (adding 1 bead to the tens column) + 5 (adding the 5 bead to the ones column). However, 42 + 42 would require 5 moves (see figure 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).

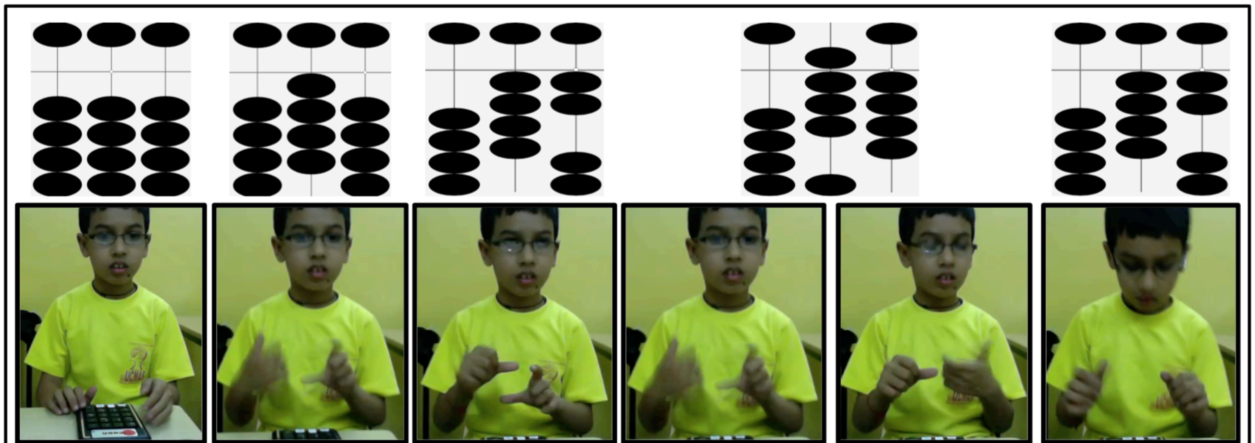

Figure 2. Example of the gestures produced as a child solves MA problems. The entire problem (42 + 42) was visible on the screen while the child solved the problem. The abacus images show how the gestures reflect the change in the state of the abacus.

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

[1] For this measure, compound moves (e.g., adding 8 by pushing down the five bead and pushing up 3 ones beads) were counted as the number of individual movements required to add the numbers (i.e., adding 8 would count as two moves).

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 =0.85) and Total Number of Required Moves (r = 0.81). In addition, since all trials start from Level 1 and increase based on a child's performance, trial number is also highly correlated with Subjective Difficulty (r = 0.85) and Problem Level (r = 0.81). In our analyses, we consider the independent effects of each of these factors on gesture outcomes; we do not explore interactions among the variables because of 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 level variable for all analyses.

Results

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 trials with only two addends, trials that are trivial for most MA users. Although MA practitioners can do computations without gesturing, no child in our study chose to do so more than a small fraction of the time, reflecting the importance of gesturing to MA.

Gesture Size

As the difficulty of the problems increased, children produced larger gestures. We ran a series of models predicting size as a continuous variable with Subjective Difficulty, Trial Number, log of Problem Level (our first measure of objective difficulty), and Number of Required Moves (our second measure of objective difficulty) as independent variables, with a random intercept of subject and random slopes where appropriate. The best-fit model contained all four variables and showed a significant effect of Problem Level ($\beta = 0.19$, t = 3.21, $p < .01$); a marginal effect of Subjective Difficulty ($\beta = 0.04$, t = 1.94, $p = .05$); and no significant effects of Trial Number or Number of Required Moves. Gesture size thus increased systematically as the number of addends in a problem, and the subjective difficulty of that problem, increased. Figure 3 shows average gesture size at each problem level for participants who found the problem relatively easy or relatively difficult.

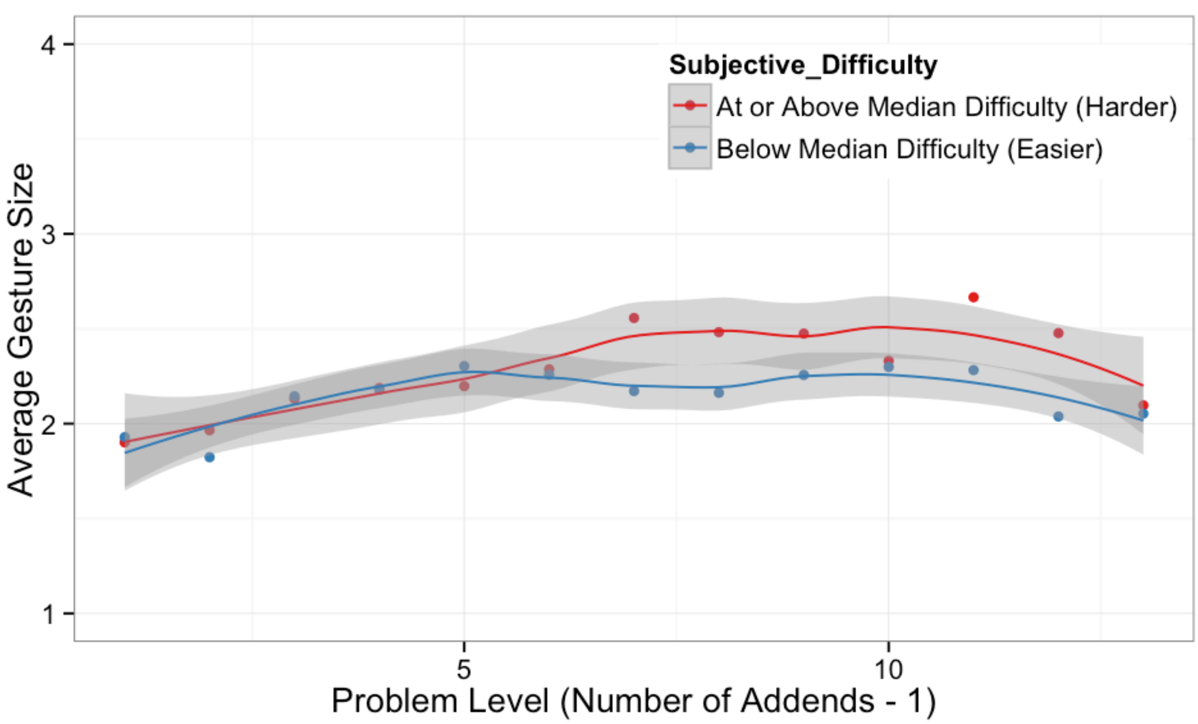

Figure 3. Gesture size as a function of Problem Level and Subjective Difficulty. At each Problem Level, we calculated the median Subjective Difficulty level of the problem, defined in relation to

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.

Moves Represented in Gesture

Children frequently produced fewer movements in gesture than the number of bead movements required to solve the problems, suggesting that some moves were skipped. Although more moves were skipped on problems that required more moves overall, we found that, when we controlled for the number of moves a problem required, participants produced a larger proportion of the required moves in gesture if the problem was subjectively challenging for them.

To examine the number of moves children gestured on a problem in relation to the number of moves required to solve the problem, 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). Fifty-eight percent of these problems required more moves than the maximum number of gestures coded, suggesting that children were producing gestures for fewer moves than the number required. In contrast, only 6% of problems required fewer moves than the minimum number of gestures coded. These cases, where more moves were observed in gesture than were necessary for the problem, may reflect the 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 moves that a child gestured 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 moves required to solve the problem. The Proportion of Moves

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 moves required; $M = .70$, $SD = .33$).

We then ran a series of mixed models to predict Proportion of Moves Gestured. We considered four contributing factors: Subjective Difficulty of the trial, Problem Level, Trial Number, and Number of Required Moves. All four factors were treated as fixed, independent variables, along with a random intercept for subject and random slopes where appropriate. The best-fit model included all 4 factors, showing significant effects of Number of Required Moves, Subjective Difficulty, and Trial Number.

Number of Required Moves. The model showed a significant negative effect of Number of Required Moves (β =-0.01, $t = 6.49$, $p < .01$): Children gestured a smaller proportion of moves for problems that were longer overall. This result is not surprising, as there may be a limit on how many moves a person can gesture in a 10-second period.

Subjective Difficulty. When controlling for the Number of Required Moves in a problem, there was also a positive effect of Subjective Difficulty (β =0.01, $t = 2.16$, $p = .03$): Once we accounted for the length of the problem, children gestured a larger proportion of moves when the problem was more difficult for them. As Figure 4 shows, for a given Number of Moves Required on the x-axis, children who found the problems easier (in green) gestured a smaller proportion of moves than those who found the problems more difficult (in red).

Trial Number and Problem Level. The model also showed a significant positive effect of Trial Number ($\beta = 0.001$, $t = 2.39$, $p = .02$): Children gestured a larger proportion of moves on later trials. There was no effect of Problem Level on Proportion of Moves Gestured in this model ($\beta = 0.0005$, $t = 0.06$, $p = .96$).

Taken together, these results show that, although the best predictor of the proportion of moves gestured was the raw number of moves required by the problem, when we control for this variable, we find that children gestured more moves in problems that were *more difficult* for them. These finding thus provide new evidence that MA experts may be recruiting specific gestures to help them solve the problems.

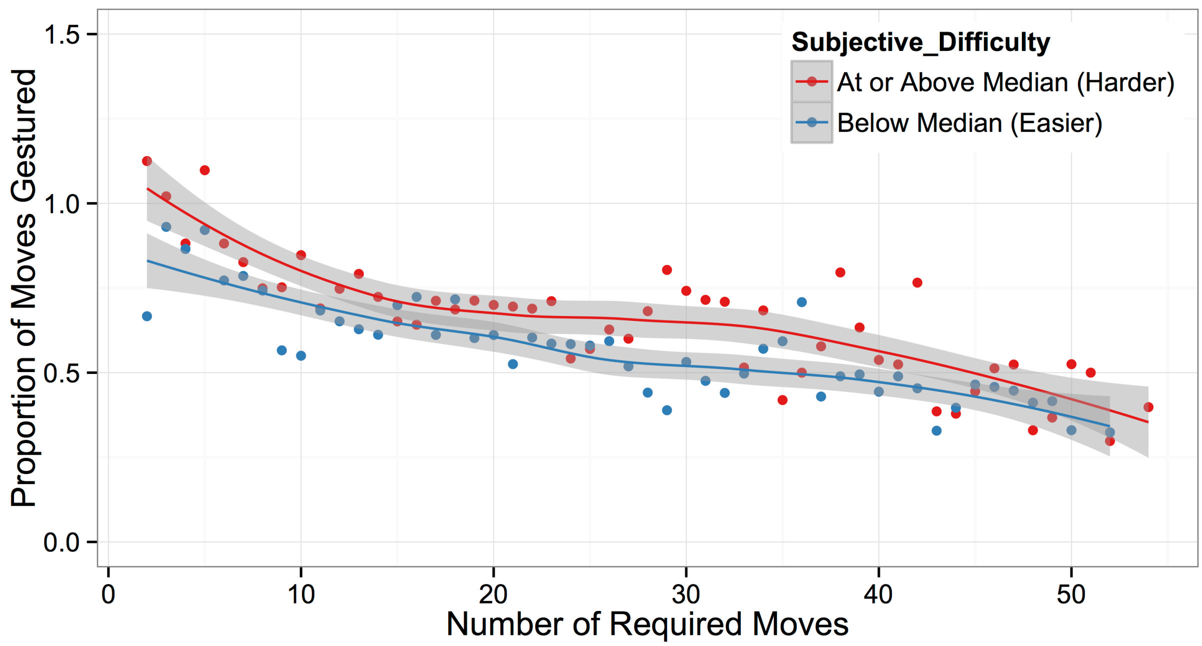

Figure 4. *Proportion of Moves Gestured as a function of Number of Required Moves. For each required number of moves, we calculated the median Subjective Difficulty level of the problem, defined in relation to a given child's threshold level and the level of the problem. Those problems falling at or above the median difficulty at each number of required moves are plotted in red; those falling below the median difficulty are plotted in blue. Although individual proportions ranged from 0 to 2.8, mean proportions at any given number of moves were never higher than 1.25.*

Summary

The findings from Study 1 show that gesture is common during MA arithmetic and reflects the specific movements required by a problem. Moreover, gesture size and the number of moves a child gestures while solving a problem 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 are actively adapting their gestures depending on the problem, rather than simply moving their hands in an arbitrary or habitual manner. However, this finding leaves open precisely *how* gesture size relates to the underlying representation of the abacus. Our second finding – that children produce more distinct moves 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 closely replicated in gesture the movements they would have produced on a physical abacus, suggesting that they were relying on gesture to help them solve these problems.

STUDY 2

Methods

In Study 2, we examined aspects of gesture that could potentially play a role in MA. We focused first on the effects of visual feedback and then on the effects of proprioceptive feedback on MA. Either type of feedback could enrich the gesturer's visuospatial representations of the abacus, making those representations easier to maintain and update. Finally, we examined the effects on MA of motor planning involved in creating gestures. MA experts might be able to solve problems without difficulty even when prevented from gesturing – as long as they are not prevented from planning gestural movements. To test the importance of these factors, we systematically eliminated each one: (1) we eliminated visual feedback by having children wear a blindfold during MA; (2) we eliminated visual and proprioceptive feedback, but preserved motor planning, by instructing children to keep their hands flat on the table during MA; (3) we disrupted motor planning along with visual and proprioceptive feedback by having children perform a motor interference task during MA.

Participants

Twenty-nine abacus experts (mean age: 11.1 years, 29% female), drawn from the same population as Study 1, participated in the study. Four children did not complete all of the tasks and thus were excluded from analyses, leaving 25 children. Sample size was estimated based on a power analysis of past motor interference data on the same population. Sixteen control children, who were naïve to abacus, were tested in the United States. We excluded 6 children from the control group because they spontaneously counted on their fingers during the baseline task and thus appeared to be using a manual strategy to solve the calculation problems (and, in this sense, were not an adequate control). Another participant was excluded because his left hand was in a cast and he was unable to comply with the motor interference manipulation. The remaining 9 control participants had a mean age of 12.5 years old (66% female).

Procedure

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 by using blindfolds. On each trial, children heard an automated voice read 6 addends aloud through headphones. Children were asked to add the numbers and verbally report their answer to an experimenter.

Each trial had a 30 second time limit, which was reached on fewer than 1% of trials overall. The trial level varied as a function of performance on the previous trial. At Level 1, all addends had a value of 3 or less (for example, a child might hear "two, two, three, one, two, one"). The maximum addends for all subsequent levels was the level cubed; e.g., at Level 2, the addends had a maximum value of 2^3, or 8; at Level 10, the maximum value was 10^3, or 1000. An exponential 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 minutes.

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. Movements of the hands that did not disturb the location of the fingertips, such as shaking, were allowed.

(4) *Motor Interference task.* Children were instructed 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 home row of the keyboard as the numbers were presented orally, the bar would decrease in length and eventually disappear, causing the trial to end and an error message to appear. Children only needed to tap while listening to the addends and could stop once the prompt to report their answer appeared on the screen.

Control Participants

The procedure was identical for the control participants tested in the US, except for changes made to adjust for the control children's arithmetic skills, which were dramatically less advanced than those of the MA participants. Control participants were asked to add three addends rather than six. In addition, the maximum addend size 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 3 factors on performance. Critically, no changes were made to the experimental manipulations for the control participants.

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 4 tasks and truncated the data so that this number of trials was used to compute the threshold for all 4 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 had at least 21 trials.

Results

Abacus experts were only minimally affected by a lack of visual or proprioceptive feedback, but performed substantially worse when not permitted to plan motor movements. Figure 5 (left panel) presents the mean problem level achieved on each of the first 20 trials in each of the 4 tasks. The Baseline, Blindfold, and No-Hands conditions cluster together and are considerably higher than the Motor Interference condition. The mean threshold level in the Baseline condition was 5.20 (which corresponds to a maximum addend size of 141 at threshold), compared to 5.30 in the Blindfold condition (maximum addend size = 149), 4.69 in the No Hands condition (maximum addend size = 103), and 2.78 in the Motor Interference condition (maximum addend size = 21).

In a linear mixed-effects model[2], we found that thresholds were significantly worse on the Motor-Interference task compared to Baseline ($\beta = -2.46$, $t = 7.03$, $p < .01$). Thresholds on the No-Hands task were marginally lower than Baseline ($\beta = -0.45$, $t = 1.71$, $p = .09$), and threshold on the Blindfold task was not significantly different from Baseline (*ns*). Adding child age, gender, or the order in which the tasks were performed did not improve the fit of the model at a very liberal .2 significance level (Barr et al., 2013).

[2] This model included threshold as the dependent variable, condition as a fixed, dependent variable, a random intercept term for subject, and a random slope term for subject across conditions.

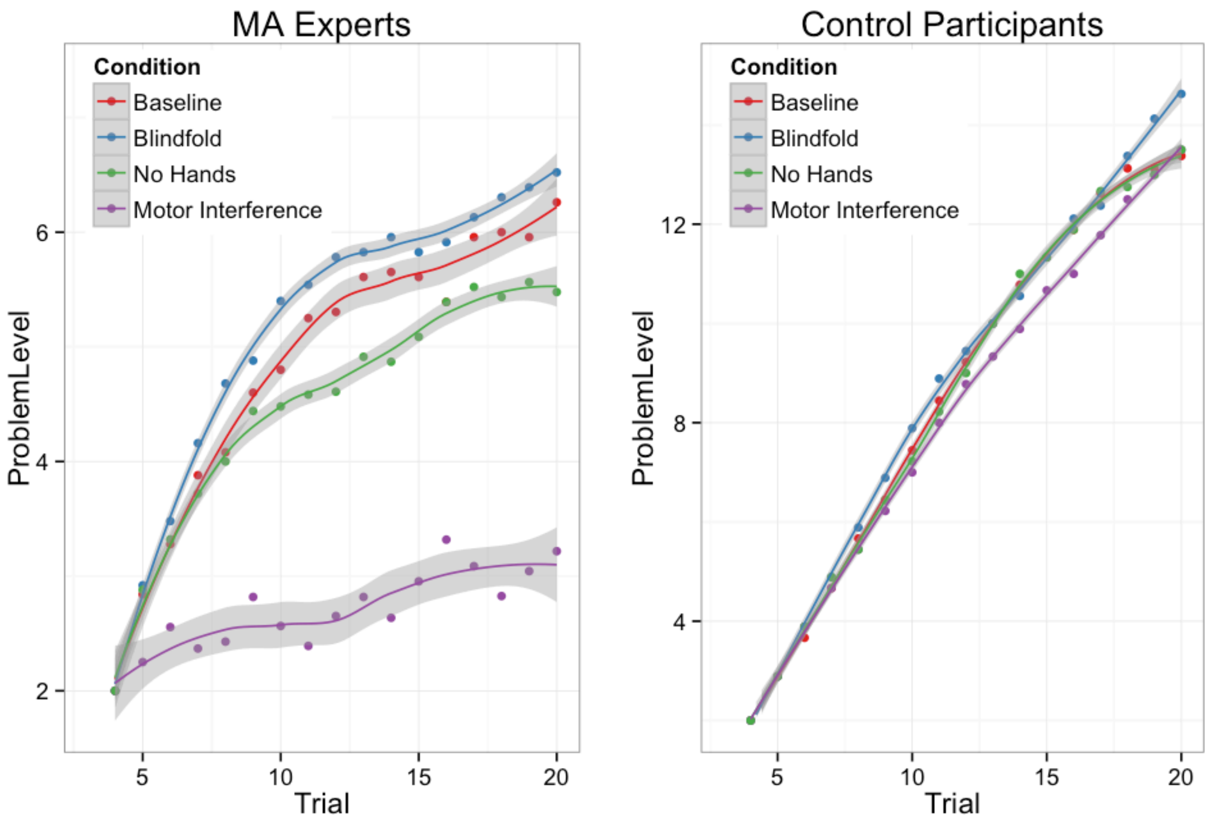

Figure 5. *Mean problem level on the first 20 trials for each of the four conditions in abacus experts (left) and control participants (right). Level of performance is not directly comparable for the control participants and the abacus participants: control participants were given shorter problems with smaller addends at all levels. However, the relation between the curves for the 4 conditions can be compared across participants.*

Control Participants

Figure 5 (right panel) presents the threshold 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 conditions, indicating that none of the manipulations had an 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 – 34 (M = 27.6, SD = 6.4). Only one participant had fewer than 20 trials. Mean threshold level was 11.72 (maximum addend size = 19.75) for Baseline; 11.87 (maximum addend size = 19.96) for Blindfold; 11.46 (maximum addend size = 19.24) for No-Hands; and 10.99 (maximum addend size = 18.24) for Motor Interference.

In a linear mixed-effects model with Threshold as a dependent variable, Condition as a fixed independent variable, and random slope and intercept terms for subject, we found no significant differences between Baseline and Motor Interference (β = -0.69, t = 1.49, p = .13), Blindfold (β =0.30, t = 0.52 p = 0.60), or the No Hands (β = -0.24, t = 0.42, p = 0.68) conditions. Adding child age, gender, or the order in which the tasks were performed did not significantly improve model fit (all ps > .50).

Summary

The findings from Study 2 show that MA experts' reliance on gesture stems not from the visual and proprioceptive feedback that these gestures provide, but rather from planning the gestural movements themselves. Disrupting motor planning with a motor interference task severely disrupted MA performance, whereas allowing motor planning but preventing gesture production (thereby eliminating any proprioceptive or visual feedback) had virtually no effect on performance. These findings suggest that planning gestural movements plays a critical role in MA, but neither seeing nor feeling those movements is necessary for calculation using this method. Data from naïve participants demonstrate that these effects do *not* stem from the motor interference task overloading executive resources. These participants, who did not use a motor strategy to solve addition problems, performed no worse under motor interference than in any of the other conditions, suggesting that the task was an additional cognitive load only for children who solved the problems using a motor strategy.

GENERAL DISCUSSION

Understanding how externally produced gestures interact with internal mental representations has presented a challenge. Using the case of mental abacus (MA), where the mental representations of problem solvers are well known and where gestures are produced without speech, we provide novel insights into the ways that gesture can interact with spatial representations in problem solving.

In Study 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 moves represented in the gestures. These results suggest that the form of the gestures MA experts produce plays an important role in problem solving, eliminating the possibility that gesture influences calculation simply by modulating attention or serving as a timekeeping device.

In Study 2, we asked how the information represented in gesture relates to internal mental representations. Results from the Blindfold and No Hands tasks demonstrate that feedback from gestural movements does not play an active role in MA computation. But planning for those movements does facilitate computation: MA experts performed significantly worse than Baseline on the Motor Interference task. Control participants did not show this effect, demonstrating that tapping *per se* does not interfere with calculation.

Studies 1 and 2 together show that planning gestures that reflect specific bead movements can facilitate MA computation. What role does motor planning play? One possibility – suggested by Hegarty, Mayer, Kriz, and Keener (2005) – is that planning gestures facilitates 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 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 Study 2 suggest a widespread benefit of motor planning of gesture for tasks that require 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 an existing movement representation by providing detailed information about bead locations as the state of the abacus changes.

Interestingly, there is a large body of research that has found negative effects on performance by inhibiting gesture without restricting planning. For example, participants who are prevented from gesturing produce speech that is less fluent (Rauscher, Krauss & Chen, 1996; Pine, Bird, & Kirk, 2007), less rich in imagery (Rimè, Shiaratura, Hupet, & Ghysselinckx, 1984; Alibali, Spencer, Knox, & Kita, 2011), 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 gesture impact performance in some cases, like explaining math problems, but not in others, like MA?

One possible explanation is that there are differences in the characteristics of the gestures that participants produce in these studies. 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 the studies where inhibiting gesture impacted performance were metaphorical, reflecting, for example, strategies for solving a math problem (Goldin-Meadow et al., 2001). Further, the gestures MA experts produce represent heavily practiced movements. In contrast, the gestures produced when participants explain a math problem or describe a scene tend to be generated on the spot. Either 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 much gestures depend on visual and/or proprioceptive feedback.

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 speakers are told not to gesture, they may also fail to plan motor movements, even if planning these movements would have been helpful to them. The fact that gestures in MA, and when reasoning about mechanical systems, are produced without speech may indicate that individuals have a strong inclination to gesture on these problems – perhaps only in these cases will they continue to plan gestures even when the gestures cannot be expressed.

In sum, we found that the gestures MA experts produce offer a window onto their computations. Further, these gestures actively interface with visuospatial representations of the abacus, although it is planning the gestures, rather than seeing or feeling them, that gives gesture its influence. These findings provide new evidence that the motor planning behind gesturing may be critical for tracking locations of visual representations over movements and transformations.

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.

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, *136*(4), 539–550. doi:10.1037/0096-3445.136.4.539

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.

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), 706–23. doi: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), 102–16. doi:10.1037/a0021790

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology. General*, *141*(1), 134–49. doi: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. doi:10.1177/1745691610388764

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–72. doi: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), 516–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11760141

Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition*, *5*(1), 47–55. doi: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.

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

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From Action to Abstraction: Using the Hands to Learn Math. *Psychological Science*, (February). doi:10.1177/0956797613518351

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.: Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion* 8, 311–332 (1984)

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.

Trafton, J. G., Trickett, S. B., Stitzlein, C. A., Saner, L., Schunn, C. D., & Kirschenbaum, S. S. (2006). The Relationship Between Spatial Transformations and Iconic Gestures University of Pittsburgh. *Spatial Cognition and Computation*, *6*(1), 1–29.

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. doi:10.1016/j.jml.2004.01.002