COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2015) 1–19

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12292

Moving to Learn: How Guiding the Hands Can Set the Stage for Learning

Neon Brooks, Susan Goldin-Meadow

Department of Psychology, University of Chicago

Received 11 August 2014; received in revised form 12 May 2015; accepted 26 May 2015

Abstract

Previous work has found that guiding problem-solvers' movements can have an immediate effect on their ability to solve a problem. Here we explore these processes in a learning paradigm. We ask whether guiding a learner's movements can have a delayed effect on learning, setting the stage for change that comes about only after instruction. Children were taught movements that were either *relevant* or *irrelevant* to solving mathematical equivalence problems and were told to produce the movements on a series of problems before they received instruction in mathematical equivalence. Children in the *relevant* movement condition improved after instruction significantly more than children in the *irrelevant* movement condition, despite the fact that the children showed no improvement in their understanding of mathematical equivalence on a ratings task or on a paper-and-pencil test taken immediately after the movements but before instruction. Movements of the body can thus be used to sow the seeds of conceptual change. But those seeds do not necessarily come to fruition until after the learner has received explicit instruction in the concept, suggesting a "sleeper effect" of gesture on learning.

Keywords: Gesture; Mathematical equivalence; Embodied cognition; Education

1. Introduction

Research in both cognitive and developmental psychology has demonstrated that the movements people make on a variety of tasks can influence their thinking on these tasks, from processing language (Beilock, Lyons, Mattarella-Micke, Nusbaum, & Small, 2008; Glenberg & Kaschak, 2002; Pulvermüller, 2005) to solving insight problems (Thomas & Lleras, 2007, 2009), to learning mathematical concepts (e.g., Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Goldin-Meadow, Cook, & Mitchell, 2009). In adults, producing

Correspondence should be sent to Neon Brooks, Department of Psychology, University of Chicago, 5848 S. University Ave., Room 517, Chicago, IL 60637. E-mail: neonbluebrooks@gmail.com

novel movements without speech can implicitly introduce new conceptual information that has an immediate impact on problem solving (Thomas & Lleras, 2009). Here we extend this work to children and to a learning task. We ask whether producing novel movements without speech can provide children with new conceptual information about a math problem and, if so, at what point in the learning process, and by what mechanisms, the conceptual change comes about.

Recent research has shed light on the role that motor experience can play in later cognition. For example, previous motor experience has been shown to affect how language is understood and processed—playing hockey can enhance one's ability to understand language about hockey, apparently because brain areas normally used to perform an act become highly involved in understanding language about that act (Beilock et al., 2008; see also Pulvermüller, 2005). Movements of the body can even get in the way of language processing. For example, when asked to make sensibility judgments about sentences like "close the drawer" (which implies action away from the body), adults respond slower when the response they are required to use involves a movement toward the body than when it involves a movement away from the body (Glenberg & Kaschak, 2002). Movements have also been shown to influence emotional memory: The direction in which participants were instructed to move marbles on an unrelated task affected the valence of the memories they recalled and the speed with which they recalled memories of consistent versus inconsistent valence (Casasanto & Dijkstra, 2010).

Movements of the body can also affect cognition during problem-solving tasks. Between attempts at solving Dunker's radiation insight problem, adults were given a secondary tracking task that guided their eye movements in patterns that were either compatible or incompatible with the problem's solution (Thomas & Lleras, 2007). Adults who moved their eyes in a problem-compatible manner were more successful at solving the problem, despite being unaware of the relation between the tracking task and the to-be-solved problem. These findings have been extended to arm exercises that were either compatible or incompatible with another insight problem (Thomas & Lleras, 2009); again, compatible movements facilitated problem solving relative to incompatible movements, even though the adults were unaware of the link between their exercises and the problem-solving task. Taken together, these findings suggest that guiding a problem-solver's movements can have an immediate effect on how quickly an insight is reached. We ask here whether the mechanisms at play in these earlier studies could help explain the role of gesture in learning. We explore, in particular, whether guiding a learner's movements without any accompanying speech or instruction can have a delayed effect, setting the stage for change that comes about only after instruction.

Past work has shown that telling children to move their hands as they explain their solutions to math problems can affect whether they are able to successfully solve those problems after receiving instruction. For example, Goldin-Meadow et al. (2009) taught children movements that were either completely compatible or partially compatible with solving a mathematical equivalence problem (e.g., $5 + 6 + 4 = \_\_ + 4$); the movements were produced along with a compatible spoken strategy. A third group of children was taught only the spoken strategy. The children were required to produce the spoken strategy and the movements before and after attempting to solve each problem during a math lesson. During this lesson, children were first shown a problem that was correctly solved by the experimenter, who then explained the solution to the problem without gesturing. Next, the children were presented with a problem and asked to produce their speech and gestures, attempt to solve the problem, and then produce the speech and gestures again. This sequence was repeated six times. Children who were taught to produce the gestures that were fully compatible with the problem achieved higher scores on a test following the lesson than children who were taught to produce gestures that were partially compatible, who, in turn, achieved higher scores than children who produced only the spoken strategy and no gestures at all. Thus, children who were told to move their hands in ways that instantiate a fully correct solution to the problem were more able to profit from a concurrently presented math lesson than children who were not instructed to gesture in this way.

Gesturing has also been found to have a "sleeper effect" on math learning, having an impact on learners only after subsequent instruction (Broaders et al., 2007). Broaders and colleagues manipulated gesture production *prior* to a math lesson. In the first phase of the study, children solved a set of mathematical equivalence problems on a pretest and then explained their solutions. After the pretest, children solved a second set of comparable problems and again explained their solutions, but this time half the children were told to move their hands while they gave their explanation; the other half were told to keep their hands still. Interestingly, children who were told to move their hands during this second explanation period expressed correct problem-solving strategies in gesture (but not in speech) that they had not expressed in either gesture or speech during the first explanation period on the pretest. Children who were told not to move their hands did not express new strategies during the second explanation period.

In the next phase of the study, children were all given the same lesson in how to solve math problems of this type. Broaders et al. (2007) found that children who had been told to move their hands (and had added new strategies to their repertoires) during the second explanation period were more likely to profit from the subsequent math lesson than children who had been told not to move their hands (and thus had not added strategies). Telling children to move their hands prior to the lesson thus brought out previously unexpressed, and correct, ideas in gesture that were not found in speech. Providing a math lesson after this manipulation seemed to allow children to capitalize on these newly expressed, yet still implicit, ideas and learn how to solve the problem.

Broaders et al. (2007) manipulated movement prior to instruction in their learning study, just as Thomas and Lleras (2007, 2009) manipulated movement prior to the moment of insight into their problem-solving studies. However, the Broaders et al. study differed in two important respects from the Thomas and Lleras studies. Because children were told to move their hands as they explained their solutions, their movements were fully integrated with their speech; it is therefore difficult to determine the effects of gesture independent of the effects of speech. The adults in the Thomas and Lleras studies did not speak as they performed their movements. In addition, children in the Broaders et al. (2007) study were permitted to produce whatever gestures they chose, so any new strategies introduced in their gestures are likely to have been reflections of some implicit knowledge already present at the time of training. In contrast, the eye and hand movements in the Thomas and Lleras (2007 and 2009, respectively) studies were dictated by experimental protocol. In this study, we use the Broaders et al. (2007) paradigm, but we ask children to make specific, rote movements without any co-occurring speech. Even though the movements are made over math problems, their relevance to the problems is not specified by any accompanying speech.

Following Thomas and Lleras (2007, 2009), we predict that producing scripted gesture without speech during a pre-instruction period will have an impact on children's post-instruction performance (a positive effect if gesture is compatible with the problem solution, a negative effect if gesture is incompatible). If so, by isolating gesture production from the instruction period, our study can provide insight into the possible routes by which gesture influences cognition. One possibility is that producing problem-relevant gestures has an immediate impact on how children represent the problem conceptually. We test for this effect in two ways: (a) We give children a math test after the manipulation but before instruction, allowing us to measure changes in their explicit understanding of the problem (i.e., an increase in the number of correct answers to the problem and perhaps the number of correct spoken strategies produced to explain those answers), as well as changes in their implicit understanding of the problem (i.e., an increase in the number of correct explanations produced in gesture but not in speech). (b) We give children a rating task previously shown to index implicit knowledge of correct problem-solving strategies on this math task (Garber, Alibali, & Goldin-Meadow, 1998), thus providing another measure of change in their implicit understanding of the problem.

A second possibility is that gesture, when produced on its own, does not have an immediate effect on how children solve the math problems, but it may have a delayed effect, influencing how children solve problems only after they receive instruction. If so, the seeds of conceptual change may be planted when children move their hands, but those seeds may not come to fruition until much later when they can be integrated with spoken instruction, suggesting a true "sleeper effect" (either positive or negative) of gesture on learning.

Our study thus contains three phases, following the design in Broaders et al. (2007). (a) In the baseline phase, we tested children on a set of mathematical equivalence problems. We included in the study only those children who did not solve any problems correctly, and did not produce any correct speech strategies when asked to describe how they solved the problems. (b) In the manipulation phase, we followed Thomas and Lleras (2009) and taught half of the children a hand movement that was *relevant* to the problem and instantiated a strategy for solving it, while the other half of the children learned a hand movement that was *irrelevant* to the problem. Children were asked to reproduce the movements they had learned over a series of math problems, without attempting to solve any of the problems. Along with this manipulation, children were also given the ratings task as a measure of whether the manipulation was affecting their implicit understanding of how to solve the problems. We then gave children a post-manipulation test that was identical in structure to the pretest as a measure of whether the manipulation had improved the children's ability to solve the problems before they received instruction. (c) In the instruction phase, all children were given spoken instruction on how to solve the problems, which involved no gesture. Following the math lesson, children were given a third, post-training test of their understanding of mathematical equivalence comparable to the pretest and post-manipulation test.

This paradigm is designed to ask at what point in the learning process, and by what mechanisms, gesture produced without speech influences children's reasoning about math problems. It is not critical to our purposes to establish whether the *relevant* movement condition has a positive impact on children's learning outcomes, the *irrelevant* movement condition has a negative impact, or both. It is clear from past literature that gesture can have both positive and negative effects on reasoning (Goldin-Meadow & Beilock, 2010)—A issue here is whether gesture has any influence on math learning when produced outside of a rich, spoken instructional context.

To summarize, our goal is to determine whether telling learners how to move their hands, without accompanying speech and prior to receiving instruction in a math problem, has an impact on their ability to profit from that instruction, and to investigate the mechanisms by which gesture influences children's changing conceptual representations of the math problem. By providing a minimal intervention (manipulating gesture without accompanying speech or instruction) and including a number of sensitive measures of children's understanding of the problem before, and after, they receive explicit instruction on the problem, we probe the nature of gesture's influence on children's math learning in a way that allows direct comparison to work showing effects of movement on adults' representations of insight problems (i.e., Thomas & Lleras, 2009).

2. Method

2.1. Participants

Sixty-two third and fourth graders were recruited and tested at their elementary schools in the Chicago area. Three participants solved at least one problem correctly on the pretest and were thus excluded from the remainder of the study. One additional child was excluded because she gave a correct explanation on the pretest in speech. This left 58 children in the analyses (M_{age} = 8.98, range = 8–10 years, 36 girls). Thirty children were randomly assigned to the *relevant* movement condition and 28 to the *irrelevant* movement condition.

2.2. Procedure

Children were tested individually. The study contained three phases: baseline, manipulation, and instruction, each culminating in a test of the child's understanding of mathematical equivalence.

2.2.1. Baseline phase

2.2.1.1. Pretest: Math problems and explanations: Children were given a paper-and-pencil test containing six mathematical equivalence problems that varied in where the blank was positioned on the right side of the equation (e.g., $5 + 8 + 2 = \_\_ + 2$, and $5 + 4 + 7 = 5 + \_\_$). After children completed the paper-and-pencil test at their own pace, they were brought up to a whiteboard and asked to explain how they solved each problem. The experimenter wrote each problem, along with the child's answer, on the board one at a time and said, "On this problem, you said that [5] plus [8] plus [2] equals [15] plus [2]. How did you get your answer?" Responses were videotaped and later coded for strategies expressed in speech and in gesture.

Children who solved any of the pretest problems correctly were excluded from the next phase of the study. If later coding showed that a child produced a correct problem-solving explanation in speech, the child's data were excluded from the analyses. As mentioned above, three children were excluded for solving problems correctly on the pretest, and one child was excluded for producing correct speech in her pretest explanations.

2.2.1.2. Pretest ratings: After completing the paper-and-pencil test and giving their explanations, children were given a ratings task as a measure of their implicit understanding of the problems and possible problem solutions. The experimenter wrote a problem on the board, filled in a possible answer, and then asked the child, "Is this answer definitely right, mostly right, mostly wrong, or definitely wrong?" (see Garber et al., 1998, for details of the procedure; children are, in general, happy to accept more than one answer on these problems, another index of the fact that they do not have a good understanding of mathematical equivalence). In turn, the experimenter filled in four possible answers for each problem, the correct answer and three incorrect answers, each corresponding to strategies used by children who do not know how to solve the problems. For example, for the problem $4 + 6 + 3 = \_\_ + 3$, the experimenter put 10 in the blank (the correct answer); 16 (the solution generated by the "add-all" strategy in which all of the numbers in the problem are added and the sum is placed in the blank); 13 (the solution generated by the "add-to-equal-sign" strategy in which the numbers on the left side of the equation are added and the sum is placed in the blank); and 6 (a solution generated by the "carry" strategy in which one of the numbers from the left side of the equation is carried over to the right and placed in the blank). This procedure was repeated for six different problems.

2.2.2. Manipulation phase

2.2.2.1. Learning the movements: Children were introduced to the hand movements they would use by a second experimenter who wrote a problem on the board and demonstrated one of two movements. The *relevant* movement was modeled after a gesture used spontaneously by children who succeed on mathematical equivalence problems of this type (Perry, Church, & Goldin-Meadow, 1988) and consisted of first moving the left hand back and forth under the left side of the problem, then moving the right hand back and forth under the right side of the problem (see Fig. 1a). This movement treats the two

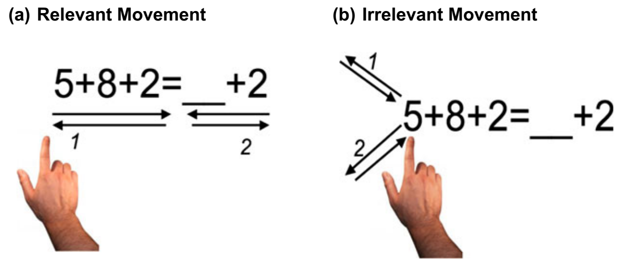

Fig. 1. Movements taught to children in the (a) *relevant* and (b) *irrelevant* movement conditions.

sides of the equation in a similar manner, highlighting the idea that the two sides should be made equivalent (the "equalizer" strategy). The *irrelevant* movement consisted of the same motions but positioned so that the motions would not provide any useful information about how to solve the problem; moving the left hand upward at a 45-degree angle away from the left side of the problem, then moving the right hand downward at a 45-degree angle away from the left side of the problem (see Fig. 1b). This movement is not associated with any known correct or incorrect problem-solving strategy that children use on this problem. Children were asked to repeat the hand movements after observing them until they could successfully reproduce the movement on two consecutive attempts. As in previous work, children were given no explanation as to why they were being asked to make these particular movements.

2.2.2.2. Movements and ratings: Once the child had successfully repeated the movement twice, the second experimenter began the manipulation phase. The experimenter wrote a new problem on the board with no answer in the blank and asked the child to repeat the hand movements that he or she had just learned. The experimenter then wrote a different problem on the board and conducted a ratings task trial as in the pretest, asking the child to rate each of four possible answers. This problem was then erased, and the child was given another movement trial. Children received six movement trials, alternating with six ratings trials.

2.2.2.3. Post-manipulation test: Math problems and explanations: After the movement and ratings tasks, the first experimenter returned to give the child another paper-and-pencil test, followed by explanations, as in the pretest.

2.2.3. Instruction phase

2.2.3.1. Training: Children next received instruction on how to solve mathematical equivalence problems, conducted by the second experimenter who was present during the movement and ratings manipulation. The experimenter wrote a problem on the whiteboard and asked the child to provide the answer. The experimenter then wrote the correct answer in the blank and provided feedback. For example, on the problem $5 + 2 + 8 = \_\_ + 8$, the experimenter said, "The answer is actually 7 [Or, if the child had put the correct answer in the blank, “That’s right, the answer is 7”]. You want to make one side equal to the other side. So 5 plus 2 plus 8 equals 15, and 7 plus 8 equals 15. So one side is equal to the other side.” This procedure was repeated on six problems. No gestures were produced by the child or the experimenter during training.

2.2.3.1. Post-training test: Math problems and explanations: The post-training test was identical in format to the pretest and post-manipulation test, and it was conducted by the first experimenter.

2.3. Coding

The explanations that the children produced on the pretest, post-manipulation test, and post-training test were transcribed and coded for problem-solving strategies conveyed in speech and gesture using a previously established coding system (Perry et al., 1988). Five strategies leading to correct solutions (equalizer, add-subtract, grouping, equal-addends, and grouping/equal-addends) and four strategies leading to incorrect solutions (add-to-equal-sign, add-all-numbers, carry, and add-to-equal-sign/add all) were coded, both for presence in speech and for presence in gesture. An explanation could contain more than one speech and/or gesture strategy. In some cases, no codable strategy was expressed in an explanation.

Two coders, trained on previous studies that made use of the same coding paradigm, transcribed and coded all the data. The coders were not involved in any other aspect of the study and were not told which condition participants were assigned to. Reliability was assessed on 20% of videos and was generally high: 82% ($\kappa = 0.80$) agreement between two coders for speech, and 83% ($\kappa = 0.80$) for gesture.

3. Results

3.1. Baseline performance

3.1.1. Pretest explanations: Gesture

Eighty-two percent of children produced at least one codable gesture strategy during the pretest explanation phase, gesturing on a total of 203 trials. The majority of these gestures (176 of 203, 87%) represent incorrect problem-solving strategies. Children in both conditions produced, on average, less than 1 correct strategy in gesture across the six pretest explanation trials ($M = 0.41$, $SD = 0.73$, gestures representing correct strategies in the *irrelevant* movement condition; $M = 0.61$, $SD = 0.79$, in the *relevant* movement condition). A binomial mixed-effects model[1] predicting correct gesture on a given pretest trial, with condition as a fixed effect and participant as a random effect, showed no significant differences between conditions ($\beta = 0.03$, $t = 0.87$, $p = .39$). There was also no significant difference between conditions in the number of incorrect gestures produced ($M = 2.72$, $SD = 2.14$, *irrelevant movement* vs. $M = 3.46$, $SD = 1.95$, *relevant movement*,

$\beta = 0.11$, $t = 1.29$, $p = .20$), or in the number of total gestures produced ($M = 3.13$, $SD = 2.20$ *irrelevant movement* vs. $M = 4.07$, $SD = 2.12$, *relevant movement*, $\beta = 0.13$, $t = 1.44$, $p = .15$).

3.1.2. Pretest ratings

There were no group differences in how participants rated the four strategies in the ratings measure at pretest. Following Garber et al. (1998), we assigned points to each rating (definitely wrong = 1, mostly wrong = 2, mostly right = 3, definitely right = 4) and then averaged the ratings for each child. Higher ratings thus reflect the fact that the child thought an answer was an acceptable solution for the problem. Fig. 2 (black bars) presents the mean ratings for children in the *relevant* and *irrelevant* movement conditions during the baseline phase of the study. Consistent with past findings showing that children who do not know how to solve mathematical equivalence problems typically use add-to-equal-sign and add-all strategies (Perry et al., 1988), these two strategies were given the highest ratings. Relatively few children in either group gave any "right" ratings to the correct answer (five in the *relevant* movement condition, seven in the *irrelevant* movement condition, with most children giving this response on fewer than half the trials). For our statistical analyses, we collapsed the two categories of "right" responses and the two categories of "wrong" responses, and analyzed the data binomially rather than assuming the four ratings made up an evenly spaced scale.[2] A binomial mixed-effects model predicting answers on each trial, with strategy, condition, and their interaction as

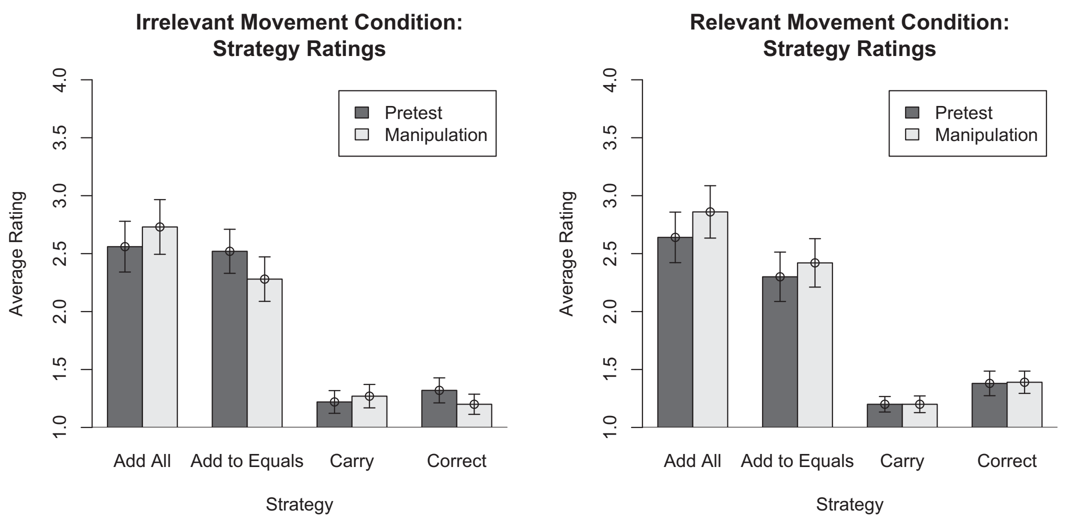

Fig. 2. Average rating for each of the strategies whose solutions were presented on the ratings task produced by children in the *irrelevant* (left graph) and *relevant* (right graph) movement conditions. Ratings during the baseline phase are represented in black; ratings during the manipulation phase are in gray (the ratings were averaged across the six problems during this phase as there were no changes over time). Error bars represent standard errors.

fixed independent variables, and participant as a random intercept, revealed a significant difference in ratings by strategy: The solutions generated by the add-all and add-to-equals strategies were more likely to be rated right than the correct solution (add-all: $\beta = 2.58$, $z = 8.26$, $p < .01$; add-to-equals: $\beta = 2.53$, $z = 8.12$, $p < .01$); there were no significant differences between ratings of the solution generated by the carry strategy and the correct solution ($\beta = 0.49$, $z = 1.20$, $p = .23$) and, critically, there were no significant differences in ratings between the *relevant* and *irrelevant* movement conditions overall ($\beta = 0.32$, $z = 0.71$, $p = .48$), and no strategy by condition interactions (all *p*s > .20).

3.2. Performance during and after the manipulation

3.2.1. Ratings

The differences between strategies found on the pretest ratings were maintained throughout the manipulation phase: Children gave similar ratings to the solutions generated by each strategy during the pretest and manipulation phases, preferring the *add-all* and *add-to-equal-sign* answers, regardless of whether they were in the *relevant* or *irrelevant* movement conditions. Some differences did emerge between the pretest and manipulation ratings, although they did not differ across condition: Children in both conditions were more likely to endorse the incorrect *add-all* strategy and the *correct* strategy at manipulation, compared with the pretest. There were no other significant differences in the pretest and manipulation ratings. Fig. 2 (light bars) presents the mean ratings produced by children in the *relevant* and *irrelevant* movement conditions during the manipulation phase of the study. A binomial mixed-effects model, with strategy, condition, and time point (baseline vs. manipulation) as fixed effects, and participant as a random effect, revealed no significant effect of condition ($\chi^2 = 0.72$, $df = 1$, $p = .40$); a significant effect of time point ($\chi^2 = 4.97$, $df = 1$, $p = .03$), and a strategy by time point interaction ($\chi^2 = 8.23$, $df = 3$, $p = .04$). There was no interaction between condition and strategy or between condition and time point, and no three-way interaction (all *p*s > 0.1). Individual linear models examining each strategy in turn show that the time point by strategy interaction arose from a significant increase in the ratings of the *add-all* and *correct* strategies at manipulation across both conditions (add-all: $\beta = 0.75$, $z = 2.6$, $p < 0.01$; correct: $\beta = 0.96$, $z = 2.6$, $p < 0.01$); and no differences between pretest and manipulation for the add-to-equal-sign and carry strategies. None of these models demonstrated a significant effect of condition, or a significant condition by time point interaction (all *p*s > 0.1). Although children's likelihood of rating the correct strategy as correct improved during the manipulation, responses were still well below chance levels of 50%: Only 17% of participants in the *irrelevant* movement condition and 11% of participants in the *relevant* movement condition rated the *correct* strategy as "definitely right" or "mostly right," despite the fact that the movements and ratings were interleaved during the manipulation. We conclude that correct strategy ratings were low after manipulation in both conditions, and that, accordingly, the ratings task we used was not able to pick up whatever differences there might have been between children in the two conditions before they underwent further instruction. An additional analysis of the final problem in the ratings task (after all six manipulation trials) showed no difference between conditions and no interaction between condition and strategy (ps > 0.40), suggesting no measureable differences in implicit understanding even after children produced the movement six times.

3.2.2. Post-manipulation test

As in the ratings task, children's responses on the post-manipulation test revealed little evidence of learning and did not vary across conditions. Fig. 3 (bars on the left) presents the mean number of problems (out of six) that the children in the two conditions solved correctly on the paper-and-pencil test taken at the end of the manipulation phase. Only one child in the *relevant* movement and two in the *irrelevant* movement condition gave any correct answers on the test (one child in each condition gave one correct answer, and one child in the *irrelevant* movement condition gave five correct answers), suggesting again that most of the children had not gained any explicit understanding of how to solve the problems as a result of the specific movements they produced during the manipulation phase of the study (i.e., before receiving any instruction). A mixed-effects binomial linear model with a child's response on each question as the dependent variable and condition (*irrelevant* or *relevant* movement) as a fixed, independent variable; and subject as a random intercept revealed no significant differences between conditions (p = .81). Neither form of movement training led to gains in learning at the post-manipulation test.

When looking at the *spoken* explanations that the children used to justify their answers on the post-manipulation test, we also see no evidence that the children's movements in either condition led to an increase in their understanding of mathematical equivalence. Only one child (the child in the *irrelevant* movement condition who got five correct answers) mentioned a correct strategy in speech and did so on a single trial.

Fig. 3. Mean number of problems answered correctly out of six on the paper-and-pencil test taken at the end of the manipulation phase (left) and at the end of the instruction phase (right) by children in the *relevant* (black) and *irrelevant* (white) movement conditions. Error bars represent standard error. The asterisks indicate the significant interaction between test and condition and the significant difference between conditions on the Post-Training Test.

Turning to the *gestured* explanations that children produced when justifying their answers to the post-manipulation test, we again found low performance (in this case, low rates of producing the correct strategy in gesture) and no differences by condition. Children in the *irrelevant* movement condition produced an average of 0.74 correct gestures across the six explanation trials, and children in the *relevant* movement condition produced an average of 0.79 correct gestures (compared to 0.40 and 0.61, respectively, on the pretest). We ran a mixed-effects binomial model predicting whether a child produced a correct gesture strategy on a given trial, with time point (pretest or manipulation), condition, and their interaction as fixed effects, along with a random effect of subject. There was a marginal increase in the production of correct gesture at manipulation ($\beta = 0.05$, $t = 1.70$, $p = .06$), but no effect of condition, and no interaction between condition and time point ($ps > 0.50$). Although being asked to produce movements on the math problems might have had some impact on children's propensity to produce correct strategies in gesture overall, there was no evidence that the *specific gestures* children produced at manipulation impacted their implicit understanding of the problems, as measured by production of correct gesture strategies.

3.2.3. Overall effects of manipulation on performance

Perhaps not surprisingly, asking children to move their hands in relation to the math problems had little impact on their ability to solve those problems. Across four separate measures—rating possible answers, coming up with answers on a written test, giving spoken explanations for the answers, and giving gestured explanations for the answers—children showed little evidence of knowing how to solve the problems as a result of producing their movements, whether those movements were *relevant* or *irrelevant* to the problem. Because performance was low across the board (i.e., children in both groups were performing at the floor), the data cannot definitively rule out the possibility that there were differences between the groups after the manipulation. We can, however, be confident that neither movement in the manipulation led to success on any of these measures.

3.3. Performance after training

In contrast to their performance on the pretest and post-manipulation test, children in the *relevant* movement condition performed significantly better on the post-training test than children in the *irrelevant* movement condition. Fig. 3 (bars on the right) presents the mean number of problems (out of six) that children in the two conditions solved correctly on the paper-and-pencil test taken at the end of the instruction phase. After exposure to identical instruction, children in the *relevant* movement condition were correct on an average of 2.50 ($SD = 2.57$) of the post-training test problems, compared to 1.27 ($SD = 2.09$) for children in the *irrelevant* movement condition. The pattern of results was bimodal, with 32 participants across both conditions (19 [68%] in the *irrelevant* movement condition and 13 [43%] in the *relevant* movement condition) solving no problems correctly on the post-training test, and 15 participants (5 [18%] in the *irrelevant* movement condition and 10 [33%] in the *relevant* movement condition) solving at least five problems correctly, accounting for 83% of all participants. Since participants were, for the most part, either incorrect on all problems or correct on all problems, we were unable to include subjects as a random effect in the model; the data were consequently analyzed with arcsine-transformed proportions of correct answers as the dependent variable. A linear model with condition, task (post-manipulation test or post-training test), and their interaction, as fixed, between-subjects variables showed no main effect of condition ($\beta = 0.03$, $t = 0.31$, $p = .76$), a significant main effect of test ($\beta = 0.22$, $t = 2.21$, $p = .03$) and, critically, a significant interaction between test and condition ($\beta = 0.32$, $t = 2.14$, $p = .03$), demonstrating that the difference between the *relevant* and *irrelevant* movement conditions was significantly larger after the training than before it. The fact that rates of learning were relatively low overall is not surprising given the minimal nature of the intervention compared to previous work.

We then looked at the *spoken* explanations that the children used to justify their answers on the post-training paper-and-pencil test and found that children in the *relevant* movement condition produced more correct speech strategies over the six trials on the post-training test (2.32, $SD = 2.70$) than children in the *irrelevant* movement condition (1.30, $SD = 2.34$). However, the difference was not statistically reliable. We used a binomial linear model predicting whether correct speech was produced on each trial, with condition, test (post-manipulation or post-training), and their interaction as fixed independent variables. The model revealed a significant main effect of test (participants in both conditions produced more correct spoken explanations after instruction than before it, $\beta = 2.36$, $t = 2.15$, $p = .03$), but no main effect of condition ($\beta = 16.2$, $t = 0.01$, $p = .99$), and no interaction ($\beta = 17.2$, $t = 0.01$, $p = .99$).

Looking next at the *gestured* explanations that the children produced, we again found that children in the *relevant* movement group produced more correct strategies in gesture over the six trials on the post-training test (1.28, $SD = 1.38$) than children in the *irrelevant* movement group (0.67, $SD = 0.96$), but again the difference was not statistically reliable. We used a binomial mixed-effects regression predicting whether correct gesture was produced on each trial; with condition, test (pretest,[3] post-manipulation, and post-training), and their interaction as independent fixed effects; and with a random effect of subject. The model revealed a significant main effect of test ($p < 0.01$), but no significant differences by condition and no interactions between condition and test (ps > 0.20).

4. Discussion

We found that guiding child learners' movements on a set of math problems prior to instruction, and without any co-occurring speech, had an impact on their ability to profit from that instruction, in line with previous findings with adults solving insight problems. Our findings thus open the possibility that there are parallel mechanisms underlying the effect that movement has on problem solving in adults and on math learning in children.

Our study went beyond the studies with adults in that we examined *when* movement had a demonstrable effect on understanding. Children in both conditions showed very little improvement in performance after the manipulation, and there were no significant differences between the movement conditions in children's implicit or explicit understanding of the problems until after they had received spoken instruction. We included two measures, administered during and after the manipulation but before instruction, of children's conceptual understanding of mathematical equivalence problems, and found that the children failed on both: (a) Although children in both groups showed improvement in their ability to recognize a correct solution as correct immediately after producing their movements during the manipulation, both groups continued to perform well below chance levels; that is, they displayed no implicit understanding of how to solve the problems prior to instruction. (b) Neither group of children demonstrated an ability to solve problems correctly, nor to produce correct spoken or gestured explanations of their problem-solving strategies, after the manipulation but before instruction; that is, they displayed no explicit understanding of how to solve the problem prior to instruction. Note that our ratings measure, although shown in previous work to be a good index of implicit knowledge on this math task (Garber et al., 1998), was not sensitive enough to detect any changes in the children's understanding of the problem brought about by the movement manipulation—those changes only became evident after the children were later given instruction in the task. In other words, even though we were unable to detect any change in implicit knowledge using our post-manipulation measures, differences on the post-instruction task make it clear that, at some level, the movement task had influenced children's implicit understanding of the problems.

Our study extends past work showing that specific child-produced movements can influence children's learning from math instruction in two respects. First, we demonstrate that movements can affect learning even when they do not refer to ideas already in the learner's repertoire, thus going beyond Broaders et al. (2007), where children were encouraged to gesture in whatever way made sense to them, presumably activating implicit ideas already present in the learner's repertoire. It is relatively easy to imagine how an idea already present in a learner's repertoire could be strengthened to become the learner's predominant response. It is more difficult to imagine mechanisms that bring brand new ideas into the learner's repertoire. Our findings suggest that gesture can be such a mechanism (see Cook, Duff, & Goldin-Meadow, unpublished data, who hypothesize that gesture's power to change thought stems from its ability to introduce new non-declarative knowledge into a learner's repertoire). Second, we demonstrate that movements can influence learning even when the movements are produced prior to instruction and without any accompanying speech, thus going beyond Cook and Goldin-Meadow (2006), Cook, Mitchell, and Goldin-Meadow (2008), Goldin-Meadow et al. (2009), and Novack, Congdon, Hemani-Lopez, and Goldin-Meadow (2014), where gesturing took place during the lesson itself and was integrated into the lesson via speech.[4] Our study thus makes it clear that the mechanisms by which gesture influences cognition need not, in all cases, rely on the tight coupling between hand movements and speech (cf. McNeill, 1992). This result is particularly striking given new findings that the mechanisms by which *seeing* gesture influences cognition may rely on the co-occurrence of gesture with speech, at least in a teaching situation (Congdon, Novack, Brooks, & Goldin-Meadow, unpublished data).

Our work provides a link between theories of embodied cognition (Barsalou, 1999; Glenberg, 1997; Wilson, 2002; Zwaan, 1999), particularly recent work showing that the way people solve problems can be altered by manipulating the way they move their eyes (Grant & Spivey, 2003; Thomas & Lleras, 2007) or their arms (Thomas & Lleras, 2009), and work showing that children's gestures can influence their ability to learn from instruction (e.g., Broaders et al., 2007; Goldin-Meadow et al., 2009). We find that children's movements can influence their learning outcomes even when they are produced without speech, and even when they are produced outside of the immediate context of learning a strategy to solve the problems, suggesting that similar mechanisms could account for the effects of movement on insight problem solving in adults and math learning in children. However, we found a "sleeper" effect of movement in child math learning that was not present in adult problem solving—manipulating children's movements had a measureable effect on their representations of mathematical equivalence problems only after they had later received instruction in the math problem. This sleeper effect suggests that, rather than directly influencing children's representations of the math problems, the movements produced during the manipulation may have had their effects by influencing children's understanding or interpretation of the spoken instruction they received later in the experiment.

Why did the movements produced in the present study fail to have a direct impact on children's understanding of the math problems given that directing a problem-solver's movements during an insight task does have a direct effect on adults' insight problem solving (Thomas & Lleras, 2009)? One obvious possibility is the age of the participants in the studies. Third- and fourth-grade children are limited in their ability to spontaneously deploy meta-cognitive skills to resolve inconsistencies (e.g., Markman, 1979). Thus, even though the children in our study produced movements directly over the problems, their limited meta-cognitive skills may have prevented them from considering a possible connection between the two until they later received instruction. In contrast, the adults in the Thomas and Lleras (2009) study may have been able to use their more advanced meta-cognitive skills to draw a connection between their own later reasoning and the movements they had made earlier in the study. However, Thomas and Lleras (2009) found no evidence that the adults in their study were aware of a connection between their movements and the problem to be solved, making it less likely that meta-cognitive awareness was at play.

Another possible reason for the difference between the two studies is the relative difficulty of the tasks. The insight problem in the Thomas and Lleras (2009) study can be solved by adults without instruction; the children in our study (as in all of the previous studies showing an impact of gesturing on solving these math problems, Cook & Goldin-Meadow, 2006; Cook et al., 2008; Goldin-Meadow et al., 2009; Novack et al., 2014) needed instruction to succeed on the problems. The children came to the learning situation with active misconceptions about how to solve the math problems, and they may not have been motivated to consider alternative strategies until given a lesson providing evidence that these misconceptions are incorrect (although it is worth pointing out that many children in our study were willing to endorse multiple solutions for a single math problem, and many produced more than one type of spoken strategy when explaining their answers to the pretest). Future research is needed to determine whether children can make use of movement produced without speech on an easier problem-solving task without requiring additional instruction. If so, the same underlying mechanism may be responsible for the effects observed here and those reported in Thomas and Lleras (2009).

Given that the children's movements did not seem to have an immediate effect on their thinking about the problems, how did they influence learning from instruction? One way that movement prior to instruction could affect subsequent learning from instruction is through the activation and integration of the remembered movement into the instruction. If seeing the same problem type again during instruction activates children's memory of the movement, then that movement is, in a sense, being produced during the math lesson (even if children are not explicitly aware of the connection). The reactivated memory trace of the movement could then influence how children interpret the lesson, serving as an implicit prime and influencing how children attend to the problems. If so, the relevant movements are likely to have helped children take advantage of the instruction they eventually received, but the irrelevant movements are also likely to have made it more difficult to take advantage of the instruction. Indeed, previous work has demonstrated that gesture can have both a beneficial and a detrimental effect on thinking depending on the nature of the information it represents (Goldin-Meadow & Beilock, 2010).

With respect to beneficial effects, the movements that the children were taught to produce in the *relevant* movement condition were based on gestures that children commonly produce when they explain their correct answers to mathematical equivalence problems of this type (e.g., Alibali & Goldin-Meadow, 1993; Perry et al., 1988). These movements were likely to be inconsistent with the incorrect ideas the children had about mathematical equivalence problems, which may have led them to become aware of new information or weigh information about elements of the problem differently, thus setting the stage for them to profit from explicit instruction. Indeed, after instruction, children in the *relevant* movement condition not only produced the equalizer strategy that was instantiated in their hand movements, but they also generalized that knowledge and produced other, correct strategies in speech (five children produced *grouping*, one produced *add-subtract*; comparable numbers for the *irrelevant* movement group were 1 and 0). In future work, we may be able to assess the level at which children process the rote movements they are taught by teaching them a correct strategy to use during manipulation (e.g., grouping) that is different from the correct strategy later taught during the lesson (e.g., equalizer). If children are succeeding by connecting the strategies used in the manipulation and lesson phases of the study, we might expect the lesson to be more effective when the manipulation and lesson strategies are the same (e.g., both equalizer as in our current study) than when the two strategies are different (e.g., grouping during the manipulation and equalizer during the lesson).[5]

With respect to detrimental effects, the movements that the children were taught to produce in the *irrelevant* movement condition may have hurt learning by making children in this condition less likely to encode relevant features of the problem during the manipulation phase of the study by drawing their attention away from these features. Encoding problem features has been shown to be an important prerequisite for learning problems of this type (McNeil & Alibali, 2004). Because the irrelevant movement was produced solely on the left side of the equation, it may have focused children's attention on that part of the problem, making them less likely to notice that the right side of the equation contained not only a blank, but also another addend. In future work, eye-tracking methods could be used to gain insight into how children attend to the problems during the manipulation and lesson phases of the study.

Whether the impact on learning stems from the beneficial effects of a task-compatible movement, the detrimental effects of a task-incompatible movement, or both, it is clear from our findings that movements produced without speech, and outside of an instructional lesson, can influence children's math learning. This finding could have important practical implications in the world of touch-screen learning tools (Ottomar, Landy, & Goldstone, 2012). Learning applications for children and adults are becoming increasingly popular on touchscreen devices such as the iPhone and the iPad. At times, the movements needed to produce an outcome on the screen are compatible with the principle underlying the problem to be solved. But unless the creator of an application has paid attention to the relation between the movements needed to produce an outcome on the screen and the principle underlying the concept to be learned, the movements are just as likely to have an arbitrary (and perhaps incompatible) relation to the principle. Our research suggests that task-compatible movements are likely to be more beneficial to learners than task-incompatible, and perhaps even task-neutral, movements.

In sum, we have shown that encouraging learners to move their hands in ways that instantiate a novel, and compatible, approach to a problem can make them more open to subsequent instruction in that problem than those who learn an incompatable movement. However, the results of this study suggest that these movements do not have a measuable effect on children's implicit or explicit representations of the problem until *after* they receive spoken instruction in how to solve the problems. Movements of the body can thus be used to sow the seeds of conceptual change. But those seeds do not necessarily come to fruition until after the learner has received explicit instruction in the concept, suggesting a "sleeper effect" of gesture on learning.

Acknowledgments

This work was supported by NICHD R01-HD47450 and NSF BCS-0925595 to SGM and by NSF Science of Learning Center funding SBE-0541957 (SGM is a co-PI). We thank Nika Castillo for help in data collection and analysis, and Lucy O'Keefe, Naureen Hemani-Lopez, and Daniel Lenzen for help with coding.

Notes

1. We used binomial regressions rather than ANOVAS with proportion of correct gestures, correct answers, etc., as the dependent variable because binomial regressions

require fewer assumptions; in particular, an ANOVA approach assumes a normal distribution of proportions when, in fact, many of our dependent variables were heavily skewed.

2. Analyzing the ratings results as a continuous scale does not change the outcome of any analysis.
3. Recall that we excluded from our sample participants who produced correct explanations in speech on the pretest. However, some participants produced correct explanations in gesture on the pretest; we therefore included this variable in our post-training analyses for gesture but not for speech.
4. In Cook and Goldin-Meadow (2006), Cook et al. (2008), Goldin-Meadow et al. (2009), and Novack et al. (2014), the learners' gestures were produced during the math lesson itself, alternating with explicit instruction from the experimenter about how to solve the problems. Moreover, the gestures were accompanied by speech that was relevant to the task and, in this way, were integrated into the instructional context. In one condition of the Cook et al. (2008) study, children were told to produce movements without speech; however, the same movements that the children produced before and after each problem during the lesson were used by the instructor immediately after the child's movements, and the instructor produced these movements along with task-relevant speech. The movements were thus modeled for the children with speech during the lesson, highlighting their relevance to the math problem.
5. Goldin-Meadow et al. (2009) taught children to produce the grouping strategy in gesture while saying the equalizer strategy in speech, and found that the manipulation led to learning. It is important to recall, however, that in the Goldin-Meadow et al. study, the learners' gestures were produced during the math lesson itself and were accompanied by speech that was relevant to task.

References

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, *25*, 468–523.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, *22*, 577–600.

Beilock, S. L., Lyons, I. M., Mattarella-Micke, A., Nusbaum, H. C., & Small, S. L. (2008). Sports experience changes the neural processing of action language. *Proceedings of the National Academy of Sciences*, *105* (36), 13269–13273.

Broaders, S., 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.

Casasanto, D., & Dijkstra, K. (2010). Motor action and emotional memory. *Cognition*, *115*(1), 179–185.

Cook, S. W., Duff, M. C., & Goldin-Meadow, S. (2015). Rethinking memory and learning: Gesture as a vehicle for non-declarative knowledge. Manuscript Submitted for Publication.

Cook, S. W., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development*, *7*(2), 211–232.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*, 1047–1058.

Garber, P., Alibali, M. W., & Goldin-Meadow, S. (1998). Knowledge conveyed in gesture is not tied to the hands. *Child Development*, *69*, 75–84.

Glenberg, A. M. (1997). What memory is for? *Behavioral & Brain Sciences*, *20*, 1–55.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

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

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–272.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*, 462–466.

Markman, Ellen. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, *50*, 643–655.

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*, 451–466.

McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago 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*, *25*(4), 903–910. doi:10.1177/0956797613518351.

Ottmar, E., Landy, D., & Goldstone, R. L. (2012). Teaching the perceptual structure of algebraic expressions: Preliminary findings from the Pushing Symbols intervention. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Pasadena, CA: Cognitive Science Society.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*, 359–400.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin and Review*, *14*, 663–668.

Thomas, L. E., & Lleras, A. (2009). Covert shifts of attention function as an implicit aid to insight. *Cognition*, *111*, 168–174.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88.