5-2019

Breaking Down Gesture and Action in Mental Rotation: Understanding the Components of Movement That Promote Learning

Elizabeth M. Wakefield-Connell
Loyola University Chicago, ewakefield1@luc.edu

Alana E. Foley
University of Chicago

Raedy Ping
Northwestern University

Julia N. Villarreal
University of Chicago

Susan Goldin-Meadow
University of Chicago

See next page for additional authors

Authors

Elizabeth M. Wakefield-Connell, Alana E. Foley, Raedy Ping, Julia N. Villarreal, Susan Goldin-Meadow, and Susan C. Levine

Breaking down gesture and action in mental rotation:

Understanding the components of movement that promote learning

[1,2]Elizabeth M. Wakefield, [1]Alana E. Foley, [3]Raedy Ping, [1]Julia N. Villarreal,

[1]Susan Goldin-Meadow & [1]Susan C. Levine

[1]University of Chicago

[2]Loyola University

[3]Northwestern University

Author Note: Funding for this study was provided by the Spatial Intelligence and Learning Center (SBE-1041707 and SBE 0541957, Levine and Goldin-Meadow are co-PIs) through the National Science Foundation, the Heising-Simons Foundation Development and Research in Early Mathematics Education (DREME) Network (support to Levine), and by NSF BCS-0925595 and NIDCD R01-HD47450 (to Goldin-Meadow). We also thank Caroline Frazia and Kristel Dupaya for help with data collection and all of the families and children who participated in these studies.

Abstract

Past research has shown that children's mental rotation skills are malleable and can be improved through *action* experience – physically rotating objects – or *gesture* experience – showing how objects could rotate (e.g., Frick et al., 2013; Goldin-Meadow et al., 2012; Levine et al., 2018). These two types of movements both involve rotation, but differ on a number of components. Here, we breakdown action and gesture into components – *feeling* an object during rotation, using a *grasping* handshape during rotation, *tracing* the trajectory of rotation, and *seeing* the outcome of rotation – and ask, in two studies, how training children on a mental rotation task through different combinations of these components impacts learning gains across a delay. Our results extend the literature by showing that, although all children benefit from training experiences, some training experiences are more beneficial than others, and the pattern differs by sex. Not seeing the outcome of rotation emerged as a crucial training component for both males and females. However, not seeing the outcome turned out to be the *only* necessary component for males (who showed equivalent gains when imagining or gesturing object rotation). Females, in contrast, only benefitted from not seeing the outcome when it involved producing a relevant motor movement (i.e., when gesturing the rotation of the object and not simply imagining the rotation of the object). Results are discussed in relation to potential mechanisms driving these effects and practical implications.

Keywords: mental rotation; gesture; action; learning

Mental rotation is an important spatial skill. We know that males tend to outperform females on mental rotation tasks (e.g., Levine, Foley, Lourenco, Ehrlich, & Ratliff, 2016), and this disparity may contribute to the emergence of sex differences on science and math assessments in adolescence (Casey et al., 1995; Casey et al., 1997; Ganley, Vasilyeva, & Dulaney, 2014), and the underrepresentation of females in STEM fields, fields that all rely to some extent on having excellent spatial skills (e.g., Gohm, Humphreys, & Yao, 1998; Wai, Lubinski, & Benbow, 2009). However, research also shows that mental rotation ability is malleable through training and practice (Baenninger & Newcombe, 1989; Uttal et al., 2013), and studies have been conducted to determine how different forms of experience might boost mental rotation ability. The present study furthers this line of investigation, and asks whether particular training experiences are differentially effective in promoting mental rotation ability in male and female children.

One possible explanation for the sex difference in mental rotation skill is that males and females tend to employ different strategies when solving mental rotation problems. Two strategies commonly discussed in the literature are the holistic and piecemeal approaches. Some evidence suggests that males are more likely to take a holistic approach when engaging in mental rotation and to use visuo-motor imagery to rotate an entire object; in contrast, females tend to use piecemeal strategies and rotate one part of the object at a time, a less efficient, and more error prone approach (e.g., Geiser, Lehmann, Corth, & Eid, 2008; Heil & Jansen-Osmann, 2008; Janssen & Geiser, 2010; Pezaris & Casey, 1991, although see Xu & Franconeri, 2015 for an argument that no one uses a holistic strategy during mental rotation). More recent work also suggests that adult males may out-perform females because they are more *flexible* in strategy selection (Nazareth, Killick, Dick, & Pruden, 2018). Under both views, the posited strategies that males and females use reflect engagement of the motor system during mental rotation tasks, and both behavioral and neuroimaging work corroborate the involvement of the motor system (Cohen et al., 1996; Kosslyn, Thompson, Wraga, & Alpert, 2001; Parsons et al., 1995; Vingerhoets, de Lange, Vandemaele, Deblaere, & Achten, 2002).

The central role that the motor system plays in mental rotation for both males and females has led several training studies to use one of two forms of movement experience: action or gesture. *Action training* has typically involved having children perform or watch objects move, whereas *gesture training* has typically involved having children perform or watch gestures that show how an object could move. Action and gesture have led to gains in mental rotation ability for both male and female children. For example, Frick and colleagues (2013a) found that 5-year-old children made fewer errors on a mental rotation task after they produced or observed *actions* that translated and rotated cut-out pieces to determine if they fit into a hole (i.e., moving, or watching an experimenter move, the cut-out piece) than if they did not receive the action experience. Similar improvements were found when Goldin-Meadow and colleagues (2012) gave 6-year-olds *gesture* production experience in an object transformation task.

One recent study compared the effects of action versus gesture training (Levine, Goldin-Meadow, Carlson, & Hemani-Lopez, 2018). Both action and gesture training improved children's performance on the mental transformation test, with no significant differences in the gains realized one week following training. However, there were differences in the timing of the effects: Action movement training resulted in an immediate gain in performance with no further gain one week later, whereas gesture movement training resulted in a more modest immediate gain with a subsequent additional gain emerging one week later, opening the possibility that gesture training would be more effective than action training at more distal time points.

Together, these studies suggest that movement experience, whether action or gesture, facilitates improvement on mental transformation tasks, albeit perhaps on different time frames. But there are many components involved in the experiences of producing actions and gestures, some of which are specific to each of these forms of movement, and some of which are shared. Given the sex differences reported in the mental rotation literature, it may be that different components of action and gesture training help males and females differently. The present study manipulates the components that comprise action and gesture movement experiences in order to determine whether certain components of these forms of motor training are critical, and whether these components have different effects on male and female children's mental rotation skills. Using a pretest-training-posttest design, we considered the impact of (1) actually touching, and thus feeling, an object during action; (2) including a grasping handshape in either gesture or action; (3) including the object's rotation around an axis, which captures the essence of the transformation, and can occur in either gesture or action; and (4) not seeing the outcome of the rotation (a property of gesture but not action), which requires more effortful processing and presumably forces the learner to use mental imagery to visualize rotational outcome.

For Study 1, five training conditions were created, using various combinations of these movement components, and the gains children made following training were compared across these conditions. Two of the training conditions mirrored action or gesture training in previous studies: (1) Learning through *action* involved positive instances of all of the movement components just mentioned (feeling the object as it is moved; using a grasping handshape; rotating the object around an axis; seeing the object in its rotated state). (2) Learning through *gesture* involved only some of these components (using a grasping handshape; turning the hand along the path of rotation that *would* occur if the object were physically rotated). The other three training conditions were made possible through the use of a touchscreen, and allowed us to test combinations of movement components that would not be possible in the 'real' world[1]: (3) Learning through *enacting* (using a grasping handshape; turning the hand along the path of rotation that *would* occur if the object were physically rotated; seeing the object in its rotated state). (4) Learning through *tracing* (turning the hand along the path of rotation that *would* occur if the object were physically rotated; seeing the object in its rotated state). (5) Learning through *tapping* (seeing the object in its rotated state after tapping it).

To our knowledge, no study has directly compared the impact of different movement components on mental rotation training. Given the prior literature, we hypothesized that all training conditions, because they involve movement and practice, should benefit children in some way. However, the *amount* of gain might be greater for some conditions than for others, and differ by sex. If so, these differences would point to the aspects of motor experience that are critical for improving mental rotation skills for males and females.

Study 1

Method

Participants

Usable data were collected from 107 children (47 males; 60 females) between the ages of 4 and 6 years (48 – 71 mos.; M = 57 mos.; SD = 7.2 mos.). Table 1 presents a breakdown of participants by sex and training condition. Participants came from a large metropolitan city, had various racial backgrounds (56% Caucasian, 19% Black, 15% Mixed Race, 3% Asian, 7% non-reporting), and came from predominately high-SES backgrounds: 48% of guardians indicated a

---

[1] As discussed in the Method section, learning through gesture occurred using a touchscreen to keep it as close as possible to the other experimental conditions involving gesture. However, it is possible to learn through gesture *without* a touchscreen, which distinguishes this training condition from the training conditions that need a touchscreen in order to be performed.

yearly household income of over $100,000; in 89% of households, at least one caregiver had obtained a bachelor's degree. To determine a target sample size, we considered a previous experiment that had investigated the effects of three training conditions on mental rotation ability, and had an effect size of Cohen's f=0.36 (Ping, Ratliff, Hickey & Levine, 2011; Cohen, 1988). We used the R package pwr, and set alpha=0.05, power=0.80, effect size f=0.36, to do a power analysis for a design with 5 conditions. The analysis indicated that we would need at least 19.39 participants per condition. We oversampled in each condition, anticipating some children would need to be excluded for various reasons (see below), and our final analysis included 19 to 24 children per condition (see Table 1).

Informed consent was obtained from a parent or guardian of each participant, and data were collected under IRB H08224 (Mental Rotation Training in Children) approved by the University of Chicago Institutional Review Board. An additional 28 children were excluded from analyses for refusing to complete the first experimental session (n = 11), failure to comply with experimental procedures (n = 13), or because of a computer error (n = 4). Children who completed the first session but were absent for no more than one subsequent session were included in analyses. Children completed the experimental sessions individually in a quiet laboratory setting. All children received a small prize as a thank-you for their participation, and parents or guardians received $10 to cover transportation costs.

Table 1.

Training Condition	*n* females	*n* males	Age (*M (SD)*)
Act	11	9	57.8 (7.5)
Enact	13	9	58.4 (6.5)
Trace	15	9	58.2 (7.4)
Gesture	11	11	57.9 (7.5)
Tap	10	9	56.3 (7.3)

Note. There was no significant difference between conditions in child age, $F(4, 106) = 0.26, p = .903$, or sex, $\chi^2(4) = 0.91, p = .923$.

Materials

Testing Items. Computerized tasks were created using OpenSesame, a graphical experiment builder (Mathot, Schreij, & Theeuwes, 2012), and presented on a laptop (Lenovo IdeaPad; 15.6-inch screen). Participants were tested on a novel mental rotation task that consisted of 36 mental rotation items at pretest, at posttest, and at a one-week follow-up[2]. Warm-up items were included to familiarize children with the task. For each item, the child's task was to mentally rotate a 2-D image of an animal or vehicle (rotated in the picture-plane by +/- 67.5, 122.5, or 157.5 degrees) so it aligned with the horizontal axis. As a guide to the child, two 'balloon parties' (plastic blue or green balloons) were placed on either side of the computer. The child could indicate the direction the image would face after it was rotated by choosing which balloon party the animal or vehicle would attend when on its feet. Images were approximately 6 cm x 3 cm, presented in profile, with a clear 'head' or 'front' (adapted from a version of the Snodgrass and

[2] Three test sets were created; children received each test set once across the three assessments (pretest, posttest, follow-up), and the order in which the test sets were administered was counterbalanced across time points.

Vanderwart object database, see Rossion & Pourtois, 2004). The direction the image faced when rotated was counter-balanced.

Within the 36-item test set, items were blocked by image type. The initial 24 items contained animal images (8 unique animals presented 3 times each, once at each angle disparity), and the final 12 items contained vehicle images (4 unique vehicles presented 3 times each, once at each angle disparity). Thus, within a test set, children received 12 trials at each angle disparity. Within animal and vehicle blocks, item presentation order was randomized for each child.

Training Items. ***Screen-based Training Items***. Screen-based training items were used for children in all training conditions except the act condition (see 'Object Training Items'); items in all conditions except the act training condition were presented on a touch screen (Apple iPad). During training, the touch screen was placed on a small stand in front of the child, with the green and blue plastic balloons on either side. Children were trained on 16 animal items, rotated in the picture-plane by +/- 67.5 or 122.5 degrees (8 unique animals presented 2 times each, once at each angle disparity). The order in which training items were presented was randomized. Four additional animal items were created for pre-training, to familiarize children with the training procedure. ***Object Training Items.*** Real-world versions of the screen-based training items were created for children in the act training condition. Each animal image was attached to black foam board by a small brad so that it could be physically rotated during training. The size and color of the animal images were the same across the screen-based and real-world training sets, and the black foam board had the same dimensions as the touchscreen. The two angle disparities (67.5 and 122.5) were marked on the back of the board, allowing the experimenter to adjust each animal to the proper angle. On each training trial, a new foam board item was placed on a small stand in front of the participant, with the green and blue plastic balloons on either side.

Procedure

Children participated in two experimental sessions in a quiet laboratory space, 1-week apart. During the first session, children completed a pretest, were randomly assigned to one of five training conditions, and completed a posttest. During the second session, children completed a follow-up test. Both sessions were video-recorded. The structure of the pretest, posttest, and follow-up assessments was identical (see Materials).

For assessments (pretest, posttest, and follow-up), children were seated in front of a laptop with the balloon parties placed on either side. Blue and green stickers were attached to laptop keys, and children were taught to make their responses during the game by pressing these keys. Between responses, children kept their hands on two hand-shaped stickers in front of the laptop.

For training, children were seated in front of a touchscreen (for the training conditions using the screen-based items) or foam board (for the training condition using real-world training items), with the blue and green balloon parties placed on either side. In all training conditions, children wore a glove with the index and thumb cut off. This gave children who used screen-based training items the ability to directly interact with the touchscreen with either their index finger or index finger and thumb, depending on condition, but prevented extraneous interaction with the screen. Gloves were worn on children's dominant hand, and the experimenter wore a similar glove.

Pretest. Before the pretest, children were familiarized with the task in warm-up trials. The first trial presented a picture of a bird standing upright. Children were told: "Look, here's a bird. The bird is walking to a party, but we have to decide which party the bird is walking to. On this side [pointing to plastic green balloons], there is a green balloon party, and on this side [pointing to plastic blue balloons], there is a blue balloon party. Which party will the bird walk to?"

Children were shown how to respond using a key press, and were given feedback to help them understand how to play the game. Children received a second trial with a different animal standing upright, before receiving the following instructions, which encouraged mental rotation: "Ok, now we changed the game a little bit. We are still playing the same game, but now the animal is turned so it is not on its feet. Before you decide which party the animal is going to, I want you to try and picture in your head what the animal would look like if it were standing on its feet. If the animal was standing on its feet, would it walk to the green balloon party or the blue balloon party?" Children completed two warm-up trials with animals rotated +/- 45 degrees, receiving feedback about their choices.

After completing these warm-up trials, children completed the pretest. They were told that they would be playing the same game as before, but this time the animal would always be turned so it was not on its feet; their goal was to picture the animal on its feet before deciding whether it would be going to the green balloon party or the blue balloon party. At the beginning of the vehicle block, children were told the game was the same, but this time, a vehicle would be turned so it was not sitting on its wheels; their goal was to picture the vehicle on its wheels before deciding which party it would attend. Children were not given any feedback during pretest.

Training. After pretest, children were taught a movement strategy to use when responding to training items, which were similar to the items they had seen at pretest. Movement strategies included different combinations of action components, based on training condition (summarized in Table 2). In all conditions, four pre-training items were presented, so that children could learn their movement strategy. The experimenter modeled the movement strategy on two items, and gave children feedback on their hand positions and accuracy on the subsequent two items.

(1) *Act*. In this condition, children used the real-world training item set. They were taught to grasp the animal attached to the core board and rotate it until it aligned with the horizontal axis. The act condition contained all of the action components—feeling the object as it is moved (feeling object); seeing the outcome of the rotation (seeing outcome); using a grasping handshape (grasping handshape); rotation movement (rotating path).

In the remaining 4 conditions, children used the screen-based training items.

(2) *Enact*. In the enact condition, children were taught to move their hand as though grasping the animal on the touchscreen with their index finger and thumb, and then rotate the hand until the animal aligned with the horizontal axis. The animal rotated in response to the child's movement; thus the child controlled the rotation of the animal along the path through his or her enacted rotation. The enact condition thus involved 3 of the 4 movement components: seeing the outcome of the rotation (seeing outcome); using a grasping handshape (grasping handshape); rotation movement (rotating path).

(3) *Trace*. In the trace condition, children were taught to touch the head of the animal on the screen with their index finger and trace the path along which it needed to move to rotate until the animal aligned with the horizontal axis. In this condition, the animal also rotated in response to the child's movement; thus the child controlled the rotation of the animal along the path. The trace condition involved 2 movement components: seeing the outcome of the rotation (seeing outcome); rotation movement (rotating path).

(4) *Gesture*. In the gesture condition, children were taught to grasp the animal on the screen with their index finger and thumb, and make a rotating movement with their fingers. Unlike the other conditions, the animal did not move in this condition; the gesture showed how the animal would need to rotate to align with the horizontal axis. The gesture condition thus did

not involve seeing the outcome of the rotation, but it did involve 2 other motion components: rotation movement that *would* occur if the animal rotated in response to the child's movement (rotating path), and using a grasping handshape to carry out the imagined rotation (grasping handshape).

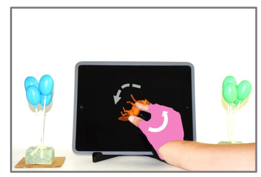
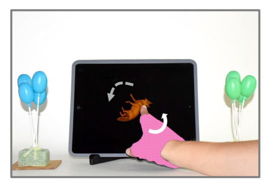
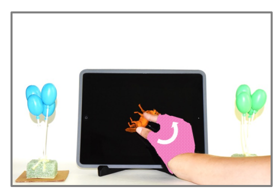
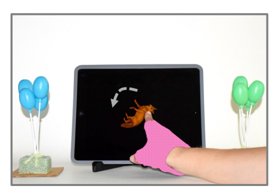
(5) *Tap*. In the tap condition, children were taught to tap the head of the animal on the screen with their index finger; this movement resulted in the animal rotating to align with the horizontal axis. The tap condition thus involved only one motion component - seeing the outcome of the rotation (seeing outcome). Although this condition engaged the motor system, it did not include any of the other 4 motion components.

Children completed 16 training trials using the movement strategy they had been taught during pre-training. All training trials showed animals at 67.5 and 122.5 degrees of angular disparity. Feedback on response accuracy was given after each trial (e.g., "The [animal] is going to the [green/blue] balloon party") and responses were recorded.

Posttest. Immediately after training, children completed a posttest on the laptop. The posttest was identical in structure to the pretest: 36 items varying by image type (animal, vehicle) and angular disparity (+/- 67.5, 122.5, 157.5 degrees). No feedback was given.

Follow-up. One week after the first session, children returned to the lab for a follow-up assessment that was identical in structure to the pre and posttests. Again, no feedback was given.

Table 2. Motion components included in each training condition. Training Example pictures: white arrow = 'Rotating Path'; grey, dotted arrow = 'Seeing Outcome'.

	Component				
Condition	Feeling Object	Grasping Handshape	Rotating Path	Seeing Outcome	Training Example
Act	+	+	+	+	
Enact		+	+	+	
Trace			+	+	
Gesture		+	+		
Tap				+	

Results

Before asking whether the action components taught during training had a differential impact on children's performance gains, we considered: (1) children's average accuracy for the three different angular disparities; (2) children's performance on animal vs. vehicle items at pretest and across time to determine whether children generalized the knowledge gained during training on animal items to vehicle items; and (3) the best model for growth (e.g., linear; quadratic) in children's performance across time.

To address our first preliminary question, we ran a mixed-effects binomial logistic regression model with trial-level accuracy on pretest (0, 1) entered as the outcome variable, and angle disparity (67.5, 122.5, 157.5 degrees) entered as the predictor of interest. Participant was considered a random effect in this model (and in all subsequent mixed-effects models described in this section) to account for common variance between mental rotation items answered by the same participant. The model revealed that children's likelihood of responding correctly to a pretest item was predicted by angle disparity ($F(2, 10581) = 771.17, p < .001$).

An effect of angle disparity was expected based on the mental rotation literature (e.g., Frick et al., 2009), but children's accuracy approached ceiling on items on which the object was rotated 67.5 degrees ($M = 0.83, SE = .02$): over half of participants (57 of 107) correctly answered 11 or 12 (out of 12) items on the 67.5 degree items, whereas only 7/107 children achieved this level of accuracy on the 122.5 degree items, and only 6/107 achieved this level on 157.5 degree items. These data suggest that little growth could be achieved on 67.5 degree items after training, and two analyses confirmed that removing these items from subsequent analyses was justifiable. First, Log-linear Poisson models showed that the number of children at ceiling differed across angular disparities: A log-linear Poisson model on a 2 (angle disparity: 67.5 degrees vs. 122.5 and

157.5 degrees) x 2 (performance code: ceiling vs. non-ceiling) contingency table revealed that there were significantly more children at ceiling for the 67.5 degree items, compared to the larger angle of rotation items (χ^2, 1 = 90.80, p < .001), which did not differ from each other, based on a second model (χ^2, 1 = 0.082, p = 0.77). Second, a Cronbach's alpha scale analysis revealed higher alpha values at each time point when the 67.5-degree angle was removed from the scale (e.g., 0.858 at pretest) than when the angle was included (0.798 at pretest). We therefore only considered 122.5 or 157.5 degree items in all subsequent analyses. Further, given the overall effect of angle on performance, we controlled for angle disparity in all subsequent analyses.

To address our second preliminary question, we assessed performance on animal vs. vehicle items at pretest and across time. If children were *not* able to generalize the training they received rotating animal images to vehicle images at posttest and follow-up, we should see greater growth across time points for performance on animal trials, compared to vehicle trials. We conducted a mixed-effects binomial logistic regression model on a trial-level accuracy variable (0, 1), with image type (animal, vehicle), time (centered at pretest), and an interaction between image type and time as predictors. Analyses revealed neither a main effect of image type, p = .14, nor an image type by time interaction, p = .74, indicating that learning did *not* differ across trained versus untrained image types. In other words, the children generalized what they had learned on animal trials to vehicle trials. As a result, we combined responses to animal and vehicle items in all subsequent analyses.

To address our third preliminary question, we evaluated models of children's growth across time to determine whether growth was better fit by a linear or quadratic function. We conducted two mixed-effects binomial logistic regression models on a trial-level accuracy variable (0, 1), the first with a term for linear growth across time, and the second with terms for linear and quadratic growth across time. The first model revealed a main effect of linear growth (F(1, 7054) = 111.46, $p < .001$), indicating that the likelihood of a correct response increased linearly across time points. The second model also revealed a main effect of linear growth (F(1, 7053) = 25.02, $p < .001$), but the term for quadratic growth did not reach significance, $p > .05$, indicating that there were no significant changes in the rate of growth from one time point to the next. Additionally, the model containing the term for quadratic time was a significantly worse fit to the data than the model with only a linear term, based on Akaike Corrected Information Criteria (AICc; χ^2 (1) = 16.32, $p < .001$). Thus, in our main analysis, we examined children's growth in performance across time as a linear function.

We next turned to our main question: whether training condition differentially promotes performance gains across time and, if so, whether the effects vary by child sex.[3] As in previous analyses, a mixed-effects binomial logistic regression model was run on a trial-level accuracy variable (0, 1). Training condition (act, enact, trace, gesture, tap), time (centered at pretest), and sex (female = 0, male = 1) were added as predictors, as well as all 2-way and 3-way interactions between these predictors of interest. We also included two covariates in this and all follow-up analyses: image angle (122.5 vs. 157.5) and child age in months at the start of the first session.

[3] We examined the accuracy of children's responses during training as a function of condition and sex. A two-way ANOVA showed a main effect of condition, F(4,107) = 12.57, $p < .001$, but no effect of sex, $p = .39$, and no interaction between condition and sex, $p = .74$. Post-hoc tests using a Bonferroni correction for multiple comparisons revealed that accuracy in the gesture condition ($M = 0.81$, $SD = 0.232$) was significantly lower than accuracy in all other conditions (act: $M = 1.00$, $SD = 0.01$; enact: $M = 1.00$, $SD = 0.00$; trace: $M = 1.00$, $SD = 0.01$; tap: $M = 1.00$, $SD = 0.02$), all ps < .001, as might be expected, as the correct outcome was not visible in the gesture condition but was visible in all other conditions. Although we might expect training performance to relate to performance at posttest and the 1-week follow-up test, it was not possible to examine this relation because there was so little variability in training accuracy. In the act condition, all but 1 child responded correctly on all 16 training trials, as did all children in the enact condition, all but 2 children in the trace condition, and all but 1 child in the tap condition.

Figure 1 shows performance gains by training condition, separated by males and females. The model revealed a main effect of time ($F(1, 7034) = 106.29$, $p < .001$, OR = 1.494, 95% CI [1.176, 1.897]), and a two-way interaction between time and training condition ($F(4, 7034) = 3.33$, $p = .010$), both of which we consider in the context of a significant 3-way interaction between time, sex, and training condition ($F(4, 7034) = 2.38$, $p = .049$). There was also a main effect of angle disparity, $F(1, 7034) = 213.70$, $p < .001$. There was no effect of age, $p = .414$, and all other main effects and two-way interactions were non-significant, ps > .05.

To explore the 3-way interaction, we ran two additional models, examining whether there was an interaction between training condition and time for males and for females separately.[4] We found a significant condition by time interaction for males ($F(4, 3060) = 3.58$, $p = .006$), but not for females, $p = .147$, suggesting that males differentially benefited from the training experiences. Post hoc analyses, using a Bonferroni corrected alpha of .005 for 10 comparisons, revealed that the time by training condition interaction seen for males was driven by performance in the gesture training condition: males showed significantly greater gains across time points when learning through gesture than when learning through three of the other training conditions (gesture vs. action: $t(1250) = 2.81$, $p = .005$, OR = 1.662, 95% CI [1.166, 2.368]); gesture vs. trace: $t(1274) = 2.96$, $p = .003$, OR = 1.749, 95% CI [1.207, 2.534]; gesture vs. tap: $t(1282) = 3.37$, $p = .001$, OR = 1.901, 95% CI [1.308, 2.763]); the fourth contrast was marginally significant (gesture vs. enact: $t(1306) = 2.48$, $p = .013$, OR = 1.626, 95% CI [1.106, 2.39]). All other pairwise comparisons were non-significant, all ps > .005. It is important to note that, although training condition did not differentially predict performance gains for females, a main effect of time was found in the model

[4] We also compared growth of males and females within each condition, and found that none of the p-values were above the Bonferroni corrected alpha level of .01 for 5 comparisons.

($F(1, 3972) = 70.101$, $p < .001$, OR = 1.501, 95% CI [1.182, 1.906]), indicating that females *did* benefit from training experiences. Although we are on shaky ground when we make inferences from the lack of an effect (in this case, the lack of a main effect of sex), it is worth underscoring that the gains made by females did *not* differ significantly from the gains made by males. The difference between the sexes is that males learned more from gesture training than from the other training conditions, whereas females did not differentially benefit from particular training conditions (although it does appear from raw performance gain scores that females made the greatest gains from action and gesture training).

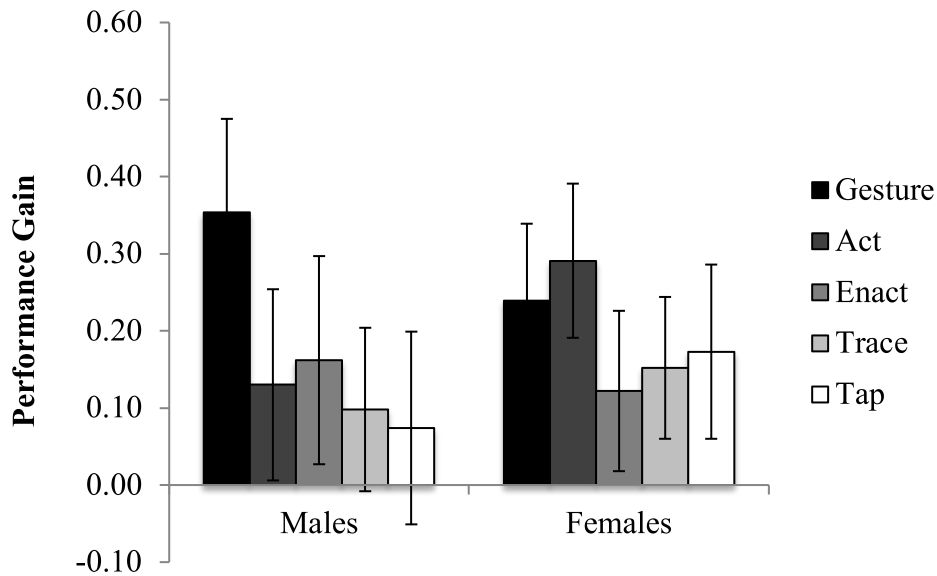


Figure 1. Model estimates of performance gain from pretest to follow-up by training condition, separated by sex. Error bars represent 95% Confidence Intervals.

Study 2

Although both males and females benefited from training experiences in Study 1, males benefited more from gesture training than the other training conditions, whereas females did not show significant differences in the gains they made across training conditions. The feature that was unique to the gesture condition and thus distinguished it from all of the other training conditions was that in the gesture condition the outcome of object rotation was *not* visible whereas it *was* visible during all other types of training (act, enact, trace, and tap). Indeed, the fact that gesturing does not result in a measurable change in the environment is a hallmark characteristic of gesture that makes it distinct from our actions on objects – presumably the gesture condition required children to *imagine* the outcome of the rotation movement on their own, rather than *see* the outcome result from their actions.

However, we cannot conclude from Study 1 that *not* seeing the outcome of the rotation during gesture training was the sole reason that this training benefited males more than the other training conditions. Gesture also involves *producing a relevant movement*, and it may be these components together that led to its beneficial effect. To disentangle the effect of a visible outcome from the effect of producing a relevant motor movement, in this study, we tested the hypothesis that simply *imagining* the movement of an object without gesturing the movement would, like gesture training, result in more learning than the enact condition; if so, we can conclude that the overt motor component of gesture does not play an important role in differentially improving males' mental rotation ability. By adding an imagine condition, we can explore whether males performed best in the gesture condition simply because they did not see the outcome of the motor movement, or because they also produced a relevant rotational motor movement.

Similarly, an imagine condition allows us to better understand our results for females. Females showed performance gains in *all* conditions, relative to pretest. Note that all of the conditions involved movement of some sort—even in the tap condition where children used a motor action (tapping the screen) that resulted in a change in the orientation of the object. To assess the hypothesis that movement is crucial to learning in females, it would be useful to have an imagine training condition, which does *not* involve any overt motor movement whatsoever.

To examine these hypotheses, children were given training in one of 3 conditions in which two factors were manipulated: an enact condition (+ visible outcome; + motor movement); a gesture condition (– visible outcome; + motor movement); an imagine condition (– visible outcome; – motor movement). The enact and gesture conditions were identical to the training conditions in Study 1; we chose to use the enact condition in Study 2 because it aligns most closely with the motor movement experience children have during gesture training, with the important difference that enacting results in a visible rotation outcome. In addition, because the effects of gesture have been found to unfold slowly across time (Congdon et al., 2017; Levine et al., 2018), Study 2 included a 4-week follow-up to probe the longevity of training effects. This longer follow-up time point also allowed us to test that possibility that gesture may be an ideal training tool for females, but that the time frame during which we considered training effects in Study 1 was not long enough to observe this effect – a finding that would align with previous work with adult females (e.g., Terlecki, Newcombe, & Little, 2008).

Method

Participants

Usable data were collected from 72 children (39 males; 33 females) between the ages of 3 and 6 years (45–72 mos.; *M* = 56 mos.; *SD* = 6.35 mos.). Table 3 presents a breakdown of participants by sex and training condition. Participants represented a racially diverse sample from a large metropolitan city (35% Black, 33% Caucasian, 14% Mixed Race, 13% Asian, 4% non-reporting, 1% Other) and came from predominately high-SES backgrounds: 52% of guardians who completed a demographic questionnaire indicated a yearly household income of over $100,000; in 74% of households, at least one caregiver had obtained a bachelor’s degree. To be consistent with Study 1, we used the same approach for sample size. Informed consent was obtained from a parent or guardian of each participant under IRB H08224 (Mental Rotation Training in Children) approved by the University of Chicago Institutional Review Board. An additional 21 children were excluded from analyses for failure to comply with experimental procedures (*n* = 15), or because they no longer wanted to participate (*n* = 6). Children completed the experimental sessions individually in a quiet school setting. All children received stickers as a thank-you for their participation, and each participating classroom in schools received a $20 gift card to an educational supply store as compensation.

Table 3.

Training Condition	*n* females	*n* males	Age (*M (SD)*)
Enact	10	13	56.4 (6.7)
Gesture	12	14	57.2 (6.6)
Imagine	11	12	55.5 (5.9)

Note. There was no significant relation between condition and child age, $F(2, 143) = 0.892$, $p = .412$, or sex, $\chi^2(2) = 0.09$, $p = .956$

Materials

Testing and Training Items. Testing materials and screen-based training materials were adapted from Study 1. As in Study 1, participants were assessed on 36 mental rotation items at pretest, at posttest, and at a one-week follow-up. In addition, a second, comparable four-week follow-up was included. Children's task on each item was to mentally rotate an image to determine the direction it would face when aligned with the horizontal axis. Items varied by image (animal versus vehicle), and angular disparity. Given the ceiling effect seen in Study 1 for 67.5 degree items, images were rotated in the picture-plane by +/- 122.5, 140, or 157.5 degrees. The direction the image faced when mentally rotated was counter-balanced on all assessments.

Four test sets of 36 items were created such that children received each test set once, across the four assessments, and the order in which the test sets were administered was counterbalanced within each training condition. As in Study 1, items were blocked by image type (animal; vehicle), each image was presented three times at each angle, and the order in which items were presented within blocks was randomized.

Procedure

Three experimental sessions took place in a quiet school setting. During session 1, children completed a pretest, were randomly assigned to one of three conditions for training, and completed a posttest. Children completed follow-up assessments one week and four-weeks after session 1. All sessions were video-recorded. The experimental set-up, assessment structure, and training session were modeled after Study 1. Training conditions varied on two factors: whether children performed a movement; whether children saw the outcome of the rotation during training. The enact and gesture conditions were identical to those conditions in Study 1; the new condition in Study 2 was the imagine condition.

(1) *Enact* (+*movement, +visible outcome*). Children were taught to grasp the animal on the screen with their index finger and thumb, and then rotate the animal until it aligned with the horizontal axis. The animal rotated as the child enacted its rotation.

(2) *Gesture* (+*movement, –visible outcome*). Children were taught to grasp the animal on the screen with their index finger and thumb, and make a rotating movement with their fingers. Unlike the enact condition, the animal did not rotate; children thus did not see the outcome of the rotation.

(3) *Imagine* (*–movement, –visible outcome*). Children were taught to imagine the animal turning until it aligned with the horizontal axis; they thus did not see the outcome of the rotation. Unlike the other two conditions, children did not produce a motor movement during training.

Results

Before considering our main analysis of interest, we examined two preliminary questions: (1) whether the change in angle disparities used during assessments (substituting the 140-degree items for the 67.5-degree items used in Study 1) resulted in lower rates of ceiling performance than in Study 1; and (2) whether children's growth in performance across time was better fit by a linear or quadratic function. To address the first question, we conducted a mixed-effects binomial logistic regression model with trial-level accuracy on pretest (0, 1) as the outcome variable and condition and angle disparity as predictors of interest. As in Study 1, this and subsequent mixed-effects models controlled for participant as a random effect. As expected, there *was* a significant effect of angle ($F(2, 10057) = 92.42$, $p < .001$): children had the lowest overall accuracy on 157.5 degree items ($M = 0.19$, $SE = 0.02$), followed by 140 degree items ($M = 0.24$, $SE = 0.03$), and then 122.5 degree items ($M = 0.37$, $SE = 0.03$). Only one child was at ceiling for 140 degree items, and no children were at or near ceiling for 122.5 or 157.5 degree items. In subsequent analyses, all items were included and angle was controlled for as a covariate.

To address the second question, we ran two mixed-effects binomial logistic regression models on a trial-level accuracy variable (0, 1), with the first model containing a term for linear growth across time, and the second containing terms for linear and quadratic growth across time. The first model revealed a main effect of linear growth ($F(1, 10078) = 110.60$, $p < .001$), indicating that the likelihood of a correct response increased linearly across time points. The second model revealed a main effect of linear growth ($F(1, 10077) = 33.75$, $p < .001$), as well as a main effect of quadratic growth ($F(1, 10077) = 8.29$, $p = .004$), indicating that there was significant deceleration in the rate of growth from one time point to the next. However, as in Study 1, the model containing the term for quadratic growth was a significantly worse fit to the data, based on AICc ($\chi^2(1) = 34.44$, $p < .001$). Thus, based on parsimony and model fit, we examined linear growth across time in our main analysis.

Our main question of interest was to determine which components of gesture training (–outcome visibility, +motor movement) had a positive impact on mental rotation skill in males and in females.[5] A mixed-effect binomial logistic regression model was run on a trial-level accuracy variable (0, 1) with condition (enact, gesture, imagine), time (centered at pretest), sex (female = 0, male = 1), and all 2- and 3-way interactions between these predictors. Child age in months and item angle disparity were included as covariates in this and all follow-up analyses. A significant main effect was found for time ($F(1, 10044) = 24.34$, $p < .001$, OR = 1.071, 95% CI [0.961, 1.193]). Significant 2-way interactions were found between time and condition ($F(2, 10044) = 5.80$, $p = .003$), and between time and sex ($F(1, 10044) = 6.13$, $p = .013$), and the 3-way interaction between time, sex, and condition was also significant ($F(2, 10044) = 8.60$, $p < .001$). There was also a main effect of angle disparity, $F(2, 10044) = 94.47$, $p < .001$. There was no

[5] As in Study 1, we examined the accuracy of children's responses during training as a function of condition and sex. A two-way ANOVA showed a main effect of condition, $F(2,67) = 19.498$, $p < .001$, but no effect of sex, $p = .766$, and no interaction effect between condition and sex, $p = .077$. Post-hoc tests using a Bonferroni correction for multiple comparisons revealed significantly higher accuracy in the enact condition ($M = .997$, $SD = .013$), compared to the gesture condition ($M = .685$, $SD = .258$) and the imagine condition ($M = .669$, $SD = .228$), both *p*s < .001, both of which did not have the correct outcome visible during training. As in Study 1, it was not possible to examine the relation between training accuracy and posttest or follow-up test accuracy because there was so little variability in training performance: in the enact condition, all but one child responded correctly on all 16 training trials, and the remaining child responded correctly on 15 of the 16 trials.

effect of age, p = .527, and all other main effects and the 2-way interaction between condition and sex were non-significant, ps >.05.

Given the 3-way interaction, we explored the 2-way interaction between condition and time for males and females separately.[6] Simple-effects tests revealed that females in the gesture condition showed significantly greater gains across time than females in the imagine condition (t(3189) = 3.20, p = .001, OR = 1.275, 95% CI [1.099, 1.479]) and in the enact condition (t(3116) = 2.57, p = .01, OR = 1.226, 95% CI [1.049, 1.432]); there was no difference in growth rates for females in the imagine and enact conditions, p = .606; see Figure 2). The significant difference between the females' performance in the gesture versus enact conditions seems, at first blush, to conflict with the findings from Study 1 (recall that there was no significant difference in performance between the gesture and enact conditions for females after training in Study 1). Note, however, that Study 2 includes an additional follow-up session after 4-weeks. If we consider growth only to the first follow-up in Study 2 (i.e., at 1-week, analogous to Study 1), we again find no significant difference in growth for females in the gesture vs. enact conditions, p = .178. The impact of gesture on females' performance is thus slow acting and takes time to appear (cf. Levine et al., 2018).

These results indicate that, for females, both components of gesture—not seeing the outcome of a movement *and* incorporating the motor system in the training experience—facilitate performance gains over a 4-week time period. In contrast to females, for males, there was no difference in growth rates between the imagine and gesture conditions, p = .836, both of which

[6] As in Study 1, we compared males and females within each condition, and found no significant difference between males and females in Study 2 in the enact or gesture conditions, the conditions that were also used in Study 1. However, there was a significant difference in the imagine condition—males grew more than females (t(3090) = 4.346, p < .001, OR = 1.399, 95% CI [1.202, 1.628]), underscoring the fact that females showed little growth when asked to imagine.

were more effective than the enact condition (gesture vs. enact: ($t(3833) = 5.90$, $p < .001$, OR = 1.554, 95% CI [1.342, 2.799]); imagine vs. enact: ($t(3445) = 5.386$, $p < .001$, OR = 0.658, 95% CI [0.566, 0.767]; see Figure 2). These results suggest that for males, the motor component in gesture training was not likely to have contributed to performance gains above and beyond the benefits of not seeing the outcome of object rotation during training. Importantly, moving the object *per se* (as in the enact condition) was not sufficient to lead to learning in males or females, and was not optimal for females' learning; having to envision the outcome of the movement seems to be a key contributor to learning in all children. Moreover, for both males and females, the gesture training condition is an effective way to support gains in mental rotation skill.

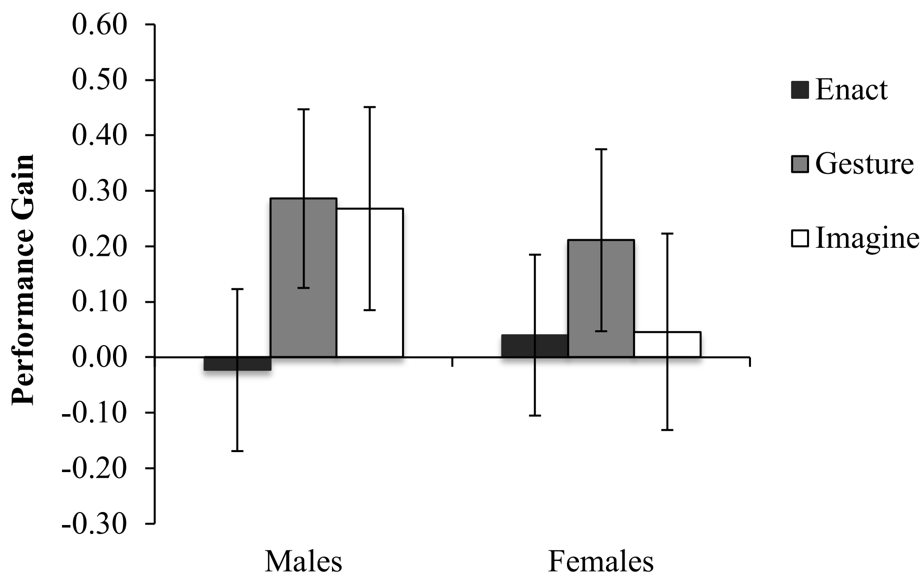


Figure 2. Model estimates of performance gain from pretest to second follow-up for each condition, separated by sex. Error bars represent 95% Confidence Intervals.

Discussion

Study 2 was designed to address whether gesture training is effective because it involves imagining the outcome of a movement, or because it involves making a relevant movement while imagining the outcome of the movement. Children were trained in one of three conditions, a gesture condition (– visible outcome; + motor movement), an imagine condition (– visible outcome; – motor movement) or an enact condition (+ visible outcome; + motor movement). We also included a second follow-up test in the design 4-weeks after training to gain a better understanding of the long-term effects of gesture training (cf. Levine et al., 2018).

Results from Study 2 answered our questions. Male participants showed significantly greater gains in both the gesture and imagine conditions than in the enact condition, but showed no difference in gains between the gesture and imagine conditions. Because both the gesture and imagine conditions involved *not* seeing the outcome of rotation, but the imagine condition did not *directly* involve overt motor movements in the training experience, this finding suggests that it was the “not seeing” component of gesture that boosted the males’ mental rotation ability in Study 1, rather than motor system involvement. In other words, encouraging males to visualize the rotation of an object facilitates gains in mental rotation ability.

Interestingly, females showed a different pattern: Females in the gesture condition showed significantly greater gains than those in the enact and imagine conditions, and there was no difference in gains between the enact and imagine conditions. This finding is notable for two reasons. First, it suggests that females benefit most from a training experience with both factors—training that involves performing a movement *and* imagining the outcome of that movement. Gesture thus appears to be an ideal tool for teaching young females mental rotation skills. Second, when the finding is considered in relation to Study 1, the results of Study 2 suggest that the effects of gesture training unfold gradually for females. In both studies, there were no significant differences in gains between females assigned to the gesture condition, compared to the other conditions (including enact), after a 1-week delay; however, a significant difference emerged between the gesture and enact conditions in Study 2 when performance was assessed after a 4-

week delay. Indeed, gains from gesture training might have surpassed gains from action, trace, and tap training in Study 1 had we assessed performance after 4 weeks. Next, we consider theoretical and practical implications of these results for mental rotation training in children.

General Discussion

We know that the actions we do influence how we think about the world. In the case of mental rotation, researchers have shown that the experience of physically rotating objects, or gesturing about rotating objects, can enhance mental rotation skill (e.g., Frick et al., 2013a; Goldin-Meadow et al., 2012; Platt & Cohen, 1981). Here, we asked whether training experiences that combine different components of action experience—directly engaging the motor system (whether through minimal engagement, as in the *tap* condition, or by producing more relevant, rotational actions, as in the *act*, *enact*, and *trace* conditions)—differentially impact performance gains on a mental rotation assessment. We also considered two other training conditions—(1) *gesture*, which involves many action components (directly engaging the motor system, tracing the path of rotation, producing a grasping handshape), but does not cause object rotation to occur, and (2) *imagine*, which is similar to gesture in that there is no visible object rotation, but differs from gesture in that no action components are involved in the experience.

Taken together, results from Studies 1 and 2 confirm previous work, showing that mental rotation ability is malleable for both female and male children (e.g., Frick et al., 2013a; Goldin-Meadow et al., 2012; Platt & Cohen, 1981), and extend the literature by showing that gesture appears to be an ideal tool for *all* children, but that the reason for this effect differs by sex. Not seeing the outcome of rotation – a hallmark characteristic of gesture – emerged as a crucial training component for both groups. Interestingly, not seeing the outcome turned out to be the *only* necessary component for males, who benefited equally from gesture training and simply *imagining* object rotation. Females, in contrast, only showed increased benefit from this experience when it also involved producing a relevant motor movement, and only after an extended delay. Given the arguments that have been made about the importance of action-experience and action-knowledge in building mental rotation skills, why did having to imagine the outcome of object rotation (a component of gesture, but not action) emerge as a crucial training component for all of the children in our study? And why was it necessary to combine imagining with making a relevant gesture for females to show gains that were comparable to those of males? We consider these two questions in turn.

***Why does having to imagine the outcome of the rotation promote learning*?** Early in development, the frequency of producing rotational movements during play and observing the resulting visual changes in object orientation may be crucial for building a rudimentary ability to mentally rotate objects (Nazareth, Herrera, & Pruden, 2013; Newcombe, Bandura, & Taylor, 1983). Such experiences may then set children up to profit from training that contains components of action (our tap, trace, enact, and act conditions). However, our findings suggest that, at least at a certain point in developing mental rotation skills, a different type of training – training that involves gesture – may be even better. Gesture may have improved children's performance on subsequent assessments because the gesture training experience in our study was similar to the testing they later received (e.g., Franks, Bilbrey, Lien, & McNamara, 2000; Halamish & Bjork, 2011; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006). But gesture also engages the motor system and, in our study, required children to simulate the movements they would have done to rotate an object. Because the movements were a simulation, they required children to visualize the outcome of the movement. Gesture thus engages the motor system as does action, but removes the 'training wheels' of action. Without a visible outcome, children have to do more 'heavy lifting' during training – they are forced to practice mentally rotating objects, aided by the visual cues provided by their hands. Having to visualize an outcome may present children just learning to mentally rotate objects with a desirable difficulty—a phenomenon in which the experience of difficulty during the learning process leads to deeper learning (e.g., Bjork, 1994; Vlach, Ankowski, & Sandhofer, 2012) — which then leads to improvement in mental rotation skills, an effect found in both males (Studies 1 and 2) and females (Study 2) after gesture experience.

Why do females need to produce a gestured movement in order to benefit from training? It is well established that male children are more likely to engage in activities that incorporate object rotation, such as block play, than females (Berenbaum & Hines, 1992; Connor, Serbin, & Schackman, 1977; Nazareth et al., 2013; Tracy, 1987). Indeed, this difference is often cited as the reason males outperform females on spatial tasks (e.g., Nazareth et al., 2013). The 4- and 5-year-old males and females in our study are likely to have had these differential experiences, and these differences may have influenced how they responded to our training conditions. That is, having had fewer opportunities to visualize the rotation of objects in play such as block play, females may have needed the cues provided by the hand movements in a gesture in order to reap the benefits of visualizing the outcome – gesture shows children *how* an object will rotate, and involves producing a similar motor pattern as would be used in actual object rotation. In other words, even though there were no pretest differences in our sample (which might be due to the novelty of the task for both males and females), the play experiences of males and females may have made them differentially ready to learn from training that does (gesture) and does not (imagine) provide cues for *how* to mentally rotate an object.

A non-mutually exclusive explanation for this finding is that males and females differ in how easily they engage their motor systems during mental rotation tasks and other tasks involving visual imagery. Previous studies show that the development of mental imagery skill is protracted, with children displaying continued development of the skill into adolescence – well beyond the age of participants in our study (e.g., Guilbert, Jouen, & Molina, 2018; Molina, Tijus, & Jouen, 2008). Guilbert and colleagues (2018) recently showed that refinement of this skill may rely on children's ability to integrate visual information with proprioceptive motor input. Although these researchers did not consider sex differences, it is possible that differences in play experiences could provide young males with more support for this integration. If so, young females might be able to engage their motor systems more effectively when actually moving (in the gesture condition) and not when imagining movement. In contrast, young males might be able to engage their motor systems not only when moving (in the gesture condition), but also when envisioning movement (in the imagine condition).

A related explanation stems from the different strategies males and females may be using during mental rotation. If males approached the mental rotation problems using a more holistic strategy, both the gesture and imagine training might encourage them to practice this strategy. In contrast, if females approached the mental rotation problems using a more piece-meal strategy, they might need gesture training to prompt them to use the more successful holistic strategy, a prompt they would not receive in the imagine condition. This explanation could also account for the second interesting finding in our study – that, for females, gesture emerged as the most effective training tool only four weeks after training (Study 2), rather than one week after training (Studies 1 and 2). Females may have needed time to profit from the prompt to use a holistic strategy. This finding is consistent with results from Terlecki, Newcombe, and Little's training study (2008), who showed slower mental rotation growth in female than male adults. Our finding is also consistent with recent results from Levine and colleagues (2018), who show particularly slow gains when children learn through gesture. It is likely that the reason all children showed slower growth in their study, rather than just females, stems from the task children had to undertake. Levine and colleagues (2018) used the Child's Mental Transformation Task (CMTT; Levine, Huttenlocher, Taylor, & Langrock, 1999), which involves more complex mental rotation and transformations than the task in our study. Because of the increased task complexity, gesture training in the Levine et al. (2018) study would have required gesturing the rotation of two images and, arguably, would have involved a greater mental load, which might have resulted in slower learning in the gesture condition for both males and females.

One somewhat surprising finding in the current work was that the tap condition (Study 1) did not result in lower performance gains than the act, enact, and trace conditions for males, and than any of the training conditions for females. Whereas all of the other conditions included a relevant rotational movement (reflecting the path of rotation for the object), the tap condition simply involved touching the object and then watching it move around its axis—arguably a less relevant motor movement even though the motor system was engaged. In recent work on mental rotation training, Levine and colleagues (2018) found that children made greater gains when producing rotational gestures than pointing gestures, suggesting that the relevance of a motor movement matters during training (see also Brooks & Goldin-Meadow, 2016). One key difference between this previous work and the current study, however, is that even the tap condition *did* result in a visual change in object orientation (which does not happen with pointing). Being able to view this change may have been enough to boost learning in the short run. But based on the desirable difficulties literature, if we consider a longer retention period (perhaps four weeks, as in Study 2), we ought to expect the tap condition to plateau in growth, compared to training conditions involving more relevant motor movements.

A second somewhat surprising finding was the lack of a main effect of age in both studies, as previous work has shown developmental changes in mental rotation ability (Estes, 1998; Frick et al., 2013a; Frick, Hansen, & Newcombe, 2013b). In particular, Estes (1998) assessed mental rotation ability in children within the same age range as used in the present study, and found that older children had higher accuracy on problems than younger children. The difference in findings between the current work and Estes' may be due to the different paradigms used. In Estes' studies, children were asked to compare two images (one of which was rotated) and decide whether they were the same or different as quickly as possible. The children were not specifically encouraged to engage in mental rotation. In contrast, in the current studies, children were encouraged to envision the rotation of animals or vehicles before making their responses, and were given unlimited time to make their choice. In addition, our framing of the problem (e.g., which party is the animal going to when on its feet) rather than making a same/different judgement, which is arguable more abstract, may have contributed to this difference. Thus, it may be the time pressure and lack of scaffolding (instruction to mentally rotate, abstractness of task) that led to the age differences documented by Estes, suggesting that 4-to-6-year-old children can perform similarly on mental rotation problems when given enough support and time to engage in the task.

Beyond beginning to unpack the aspects of action and gesture experience that are most beneficial in building children's mental rotation ability, our results have implications for parents and teachers, particularly as the use of tablets becomes more common. Tablets allow for screen-based learning, and move beyond the capabilities of traditional computers by allowing what is often thought of as ‘physical’ interactions with objects. Children can drag and drop objects from one place to another, enlarge or shrink objects, and, as in the present study, rotate objects around an axis. However, the present work suggests that the experience of doing a movement that has an effect on an object is not always best for learners – at times it may be better to do a movement and then visualize the effect that it would have had on the object. Children in our study did make some learning gains in all conditions, suggesting that touchpad based learning can be effective. But parents and teachers may also want to transition children to more abstract learning experiences, like gesture, in order to secure the biggest boosts in performance.

In conclusion, our studies were designed to evaluate which components of action and gesture experiences are particularly effective in improving mental rotation ability. Our results suggest that gesture training, which requires the child to make a relevant motor movement and to visualize the outcome of the movement, results in the best learning for both males and females. Interestingly, males were able to profit from training that involved visualizing an outcome even if they did not perform a movement, whereas females were unable to do so. Future work is, of course, needed to further solidify the mechanism driving these effects, and to understand how children’s prior experiences, strategies, and levels of ability impact the kinds of training that are most effective in improving their spatial thinking. This question is particularly important for supporting spatial learning in children from lower SES backgrounds, who are currently under-represented in STEM fields. However, our findings strongly suggest that, if only one tool is going to be used in an educational setting, *gesture* ought to be that tool since it alone has the power to help *both* males and females improve their mental rotation ability.

References

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*, 327-344. doi:10.1007/BF00287729

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science, 3*, 203-206. doi:10.1111/j.1467-9280.1992.tb00028.x

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Brooks, N., & Goldin-Meadow, S. (2016). Moving to learn: How guiding the hands can set the stage for learning. *Cognitive Science, online first*, 1-19. doi:10.1111/cogs.12292

Casey, B. J., Cohen, J. D., Jezzard, P., Turner, R., Noll, D. C., & Trainor, R. J. (1995). Activation of prefrontal cortex in children during a nonspatial working memory task with functional MRI. *Neuroimage, 2*, 221-229. doi:10.1006/nimg.1995.1029

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., & Giedd, J. N. (1997). A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. *Journal of Cognitive Neuroscience, 9*, 835-847. doi:10.1162/jocn.1997.9.6.835

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., & Anderson, A. K. (1996). Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain, 119*, 89–100. doi:10.1093/brain/119.1.89

Connor, J. M., Serbin, L. A., & Schackman, M. (1977). Sex differences in children's response to training on a visual-spatial test. *Developmental Psychology, 13*, 293-294. doi:10.1037/0012-1649.13.3.293

Estes, D. (1998). Young children's awareness of their mental activity: The case of mental rotation. *Child Development, 69*, 1345-1360. doi:10.1111/j.1467-8624.1998.tb06216.x

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory and Cognition, 28*, 1140-1151. doi:10.3758/BF03211815

Frick, A., Ferrara, K., & Newcombe, N. (2013a). Using a touch screen paradigm to assess the development of mental rotation between 3½ and 5½ years of age. *Cognitive Process, 14*, 117-127. doi:10.1007/s10339-012-0534-0

Frick, A., Hansen, M. A., & Newcombe, N. S. (2013b). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development, 28*, 386-399. doi:10.1016/j.cogdev.2013.06.002

Ganley, C. M., Vasilyeva, M., & Dulaney, A. (2014). Spatial ability mediates the gender difference in middle school students' science performance. *Child Development, 85*, 1419-1432. doi:10.1111/cdev.12230

Geiser, C., Lehmann, W., Corth, M., & Eid, M. (2008). Quantitative and qualitative change in children's mental rotation performance. *Learning and Individual Differences, 18*, 419-429. doi:10.1016/j.lindif.2007.09.001

Gohm, C. L., Humphreys, L. G., & Yao, G. (1998). Underachievement among spatially gifted students. *American Educational Research Journal, 35*, 515-531. doi:10.2307/1163447

Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science, 15*, 876-884. doi:10.1111/j.1467-7687.2012.01185.x

Guilbert, J., Jouen, F., & Molina, M. (2018). Motor imagery development and proprioceptive integration: Which sensory reweighting during childhood? *Journal of Experimental Child Psychology, 166*, 621-634. doi:10.1016/j.jecp.2017.09.023

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 801–812. doi:10.1037/a0023219

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology, 61*, 683-689. doi:10.1080/17470210701822967

Janssen, A. B., & Geiser, C. (2010). On the relationship between solution strategies in two mental rotation tasks. *Learning and Individual Differences, 20*, 473-478. doi:10.1016/j.lindif.2010.03.002

Kosslyn, S. M., Thompson, W. L., Wraga, M., & Alpert, N. M. (2001). Imagining rotation by endogenous versus exogenous forces: distinct neural mechanisms. *NeuroReport, 12*, 2520-2525. doi:10.1097/00001756-200108080-00046

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S. B., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Cognitive Science, 7*, 127-155. doi:10.1002/wcs.1380

Levine, S. C., Goldin-Meadow, S., Carlson, M., & Hemani-Lopez, N. (2018). Mental transformation skill in young children: the role of concrete and abstract motor training. *Cognitive Science*, 1-22. doi:10.1111/cogs.12603

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex difference in spatial skill. *Developmental Psychobiology, 35*, 940-949. doi:10.1037/0012-1649.35.4.940

Mathot, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav Res Methods*.

Molina, M., Tijus, C., & Jouen, F. (2008). The emergence of motor imagery in children. *Journal of Experimental Child Psychology, 99*, 196-209. doi:10.1016/j.jecp.2007.10.001

Nazareth, A., Herrera, A., & Pruden, S. M. (2013). Explaining sex differences in mental rotation: role of spatial activity experience. *Cogn Process, 14*, 201-204. doi:10.1007/s10339-013-0542-8

Nazareth, A., Killick, R., Dick, A. S., & Pruden, S. M. (2018). Strategy selection versus flexibility: Using eye-trackers to investigate strategy use during mental rotation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, online first*. doi:10.1037/xlm0000574

Newcombe, N., Bandura, M. M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. *Sex Roles, 9*. doi:10.1007/BF00289672

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology, 74*, 18-22. doi:10.1037/0022-0663.74.1.18

Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., . . . Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature, 375*, 54-58. doi:10.1038/375054a0

Pezaris, E., & Casey, M. B. (1991). Girls who use “masculine” problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition, 17*, 1-22. doi:10.1016/0278-2626(91)90062-D

Platt, J. E., & Cohen, S. (1981). Mental rotation task performance as a function of age and training. *Journal of Psychology: Interdisciplinary and Applied, 108*, 173-178. doi:10.1080/00223980.1981.9915260

Roediger, H. L., & Karpicke, J. D. (2006). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 173*, 249-255. doi:10.1111/j.1467-9280.2006.01693.x

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception, 33*, 217-236. doi:10.1068/p5117

Terlecki, M. S., Newcombe, N., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22*, 996-1013. doi:10.1002/acp.1420

Tracy, D. M. (1987). Toy-playing habits, spatial ability, science and mathematics achievement: are they related? . *Sex Roles, 17*, 115-138. doi:10.1007/BF00287620

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*, 352-402. doi:10.1037/a0028446

Vingerhoets, G., de Lange, F. P., Vandemaele, P., Deblaere, K., & Achten, E. (2002). Motor imagery in mental rotation: an fMRI study. *Neuroimage, 17*. doi:10.1006/nimg.2002.1290

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 246–254. doi:10.1037/a0025260

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835. doi:10.1037/a0016127