An Investigation of Mental Rotation in Infancy Using Change Detection

Aaron G. Beckner[a], Annika T. Voss[c], Lindsey Phillips[b], Kathryn King[b], Marianella Casasola[b], and Lisa M. Oakes[b,c]

[a] School of Human Ecology, Cornell University

[b] Department of Psychology, University of California, Davis

[c] Center for Mind and Brain, University of California, Davis

Funding: This research and preparation of this manuscript were made possible by grant BCS 1823489 awarded to MC and LMO from the National Science Foundation.

Acknowledgements: We thank the students and staff in the Infant Cognition Laboratory at the University of California, Davis, and in particular Anthony Easter, Ananya Das, Austin Nguyen, and Amanda Wilheim, for their help with data collection and coding.

CRediT author statement:

Aaron G Beckner: Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Visualization, Software, Writing - Original Draft, Project Administration
Annika T. Voss: Formal Analysis, Data Curation, Visualization, Writing - Original Draft
Lindsey Phillips and Kathryn King: Methodology, Investigation, Writing – Review & Editing
Marianella Casasola: Conceptualization, Writing - Review & Editing, Funding Acquisition
Lisa M. Oakes: Conceptualization, Writing - Review & Editing, Supervision, Funding Acquisition

Research Highlights

- The robustness of infant mental rotation was assessed in two pre-registered replication studies
- Six-to-twelve-month-old infants were tested in a mental rotation change detection procedure
- Infants failed to display evidence of mental rotation in either an exact or conceptual replication
- These results suggest that infant mental rotation is fragile and difficult to isolate in the change detection procedure

Abstract

Two experiments were conducted to examine mental rotation in 6- to 12-month-old infants (N = 166) using a change detection task. These experiments were replications of Lauer and Lourenco (Lauer et al., 2015; Lauer & Lourenco, 2016), using identical stimuli and variations of their procedure, including an exact replication conducted in a laboratory setting (Experiment 1), and an online assessment using Lookit (Scott et al.,2017; Scott & Schulz, 2017) (Experiment 2). Both experiments failed to replicate the results of the original study; in neither experiment did infants' behavior provide evidence that they mentally rotated the object. Results are discussed in terms of the robustness of mental rotation in infancy and about limits in our experimental procedures for uncovering perceptual and cognitive abilities in infants.

An Investigation of Mental Rotation in Infancy Using Change Detection

1. Introduction

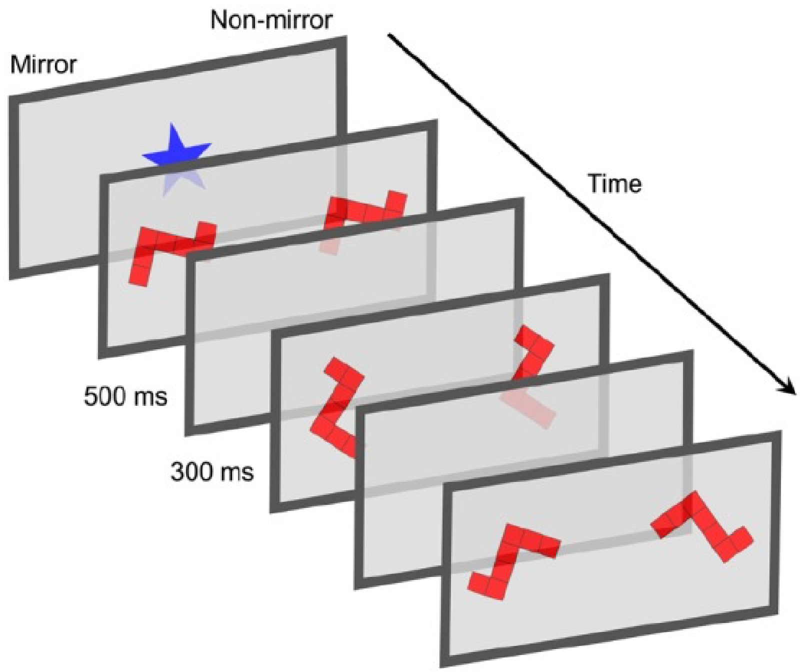
Humans recognize objects from different viewpoints, adjust behaviors to grasp and manipulate specific items, and predict the effects of their actions on objects to effectively interact with their environment. All these abilities require *mental rotation*, a spatial ability that allows us to mentally manipulate representations of objects (Shepard & Metzler, 1971; Vandenberg & Kuse, 1978). Mental rotation has been reported in infancy (for a review see Moore & Johnson, 2020), and individual differences in infants' and young children's mental rotation predict later mathematical competence (Cheng & Mix, 2014; Lauer & Lourenco, 2016; Mix et al., 2016; Newcombe et al., 2019).

However, studies of mental rotation in infancy have yielded conflicting results about when and how this spatial ability emerges. Some studies have reported findings to show that mental rotation emerges between 3 and 5 months of age (Moore & Johnson, 2008, 2011; Quinn & Liben, 2008), whereas the results of other studies suggest that the emergence of mental rotation depends on motor achievements at 6 months or later (Frick & Möhring, 2013; Möhring & Frick, 2013). In addition, some studies have observed that the differences in mental rotation abilities seen in adult men and women research participants (Hyde, 2007; Linn & Petersen, 1985; Voyer et al., 1995) are also evident in infants, with more advanced mental rotation reported in male than female infants (Lauer et al., 2015; Moore & Johnson, 2008, 2011; Quinn & Liben, 2008). However, other studies of infants' mental rotation find no difference between infant boys and girls (Erdmann et al., 2018; Hespos & Rochat, 1996; Slone et al., 2018). Thus, taken as a whole, the findings of studies on infant mental rotation present a conflicted depiction of when this ability emerges and raises questions about the reliability of the results.

One reason it is difficult to draw firm conclusions from the extant literature on infant mental rotation is that researchers have used a wide variety of methods, stimuli, and procedures. Studies have used static letters or numbers (Quinn & Liben, 2008, 2014) or dynamic, moving complex objects (Hespos & Rochat, 1996; Moore & Johnson, 2008, 2011). Of these studies, some have used 2-dimensional stimuli (Hespos & Rochat, 1996; Lauer et al., 2015; Quinn & Liben, 2008, 2014), 3-dimensional objects (Frick & Möhring, 2013; Möhring & Frick, 2013), or 2-dimensional representations of 3-dimensional objects presented on a screen (Christodoulou et al., 2016; Moore & Johnson, 2008, 2011; Slone et al., 2018). Many studies have used procedures engaging long-term memory, in which infants first are familiarized with an object and then tested with events that show impossible rotations or mirror images of the familiar object (Christodoulou et al., 2016; Constantinescu et al., 2018; Erdmann et al., 2018; Hespos & Rochat, 1996; Moore & Johnson, 2008, 2011; Quinn & Liben, 2008, 2014; Slone et al., 2018).

Lauer et al. (2016; 2015) adapted the c*hange detection procedure*, originally developed by Ross-Sheehy et al. (2003) to study infants' visual short-term memory (VSTM), to evaluate infants' mental rotation. This task assesses infants' mental rotation of objects briefly stored in VSTM. Infants view stimulus streams in which an object repeatedly appears, briefly disappears, and then reappears in a different orientation (see Figure 1). On some streams, the object simply rotates; thus, the stream involves multiple images of the same object in different orientations. On other streams, sometimes the object that appears is not only rotated but is also the mirror image of the original object. Therefore, these mirror-change streams involve images of two different objects in different orientations. If infants recognize the invariance of the object representation in VSTM across changes in orientation, the mirror-change streams should be more interesting to infants. In this task, these two streams are presented side-by-side, and infants' mental rotation is inferred from their preference for the mirror-change stream stimulus over the non-mirror-change stream.

Figure 1. Schematic illustration of the mental rotation change detection procedure.



We conducted two pre-registered replications of the study conducted by Lauer et al. (2015), using the same basic simultaneous stream change detection task, the same stimulus objects, and testing infants in the same age range. Our two experiments differed from one another in the length and number of trials and whether the session was conducted in a laboratory or at home using Lookit, an online platform (Scott et al., 2017; Scott & Schulz, 2017). In addition, Experiment 2 included a within-subject manipulation to help us understand why this task may be difficult for infants. These experiments will reveal how robust infants' mental rotation is when assessed in the change detection task and will allow us to systematically examine how task variation influences infants' mental rotation in a single experimental procedure. In addition, we collected data about infants' motor development, allowing us to address whether mental rotation abilities are associated with emerging motor abilities.

2. Experiment 1

Our first pre-registered experiment (https://osf.io/vfa4q/?view_only=8f897e9f224e47e5a44a24c4c36826e5) was an in-lab experiment, designed to be a close replication of the original Lauer et al. (2015).

2.1. Method

2.1.1. Participants. Our final sample included 59 healthy, full-term infants (30 girls) ages 5- to 13-months-old (*M* = 295 days, *SD* = 53.88 days, range = 179 - 379 days), from the [blinded for review], tested between 11/22/2017 and 11/27/2018. The infants in our final sample had no familial risk of colorblindness. As specified in our pre-registration, we determined a target sample size of 64 infants based on a power analysis using G*Power and the results reported by Lauer et al. (2015). We continued testing infants until we had tested 84 infants, assuming that we would have an ultimate sample size of approximately 64 infants. After coding and all exclusions were applied, we had a sample of 59 infants, short of our target but comparable to Lauer et al.'s (2015) sample of 56 infants.

We excluded the data from 25 of the 84 infants tested because they became too fussy to complete all four trials of the study (e.g., crying, arching back, turning toward the caregiver; n = 18), equipment error (n = 4), or experimenter error (n = 2). In addition, 1 infant tested was later determined to be ineligible (e.g., was born premature), and their data were discarded.

The infants in our final sample were racially diverse: caregivers reported that 36 infants were White (8 were Hispanic), 1 was Black or African American and Hispanic, 5 were Asian, 1 was Native Hawaiian or other Pacific Islander, 15 were multiracial (2 were Hispanic), and 1 did not have race reported and was Hispanic. Of the families who reported maternal education (*n* = 57), all mothers had completed some college or a 2-year degree and 44 mothers had earned at least a 4-year degree.

Infant names were obtained from the [blinded for review] and local caregivers were sent informational letters about our research, including information about volunteering to participate. When infants approached the appropriate age for the present study, we contacted caregivers who had expressed interest in participating. If caregivers agreed and their infant was eligible (e.g., born full-term, no vision problems), informed consent was obtained during a single visit to our lab.

2.1.2. Stimuli. The experimental stimuli were created to be identical to those used by Lauer et al. (2015). Each stimulus stream contained a single red (RGB: 219, 67, 52), two-dimensional, Tetris-like block (approximately 10.16 cm wide x 10.16 cm high, or 5.82˚ by 5.82˚ visual angle at a viewing distance of 100 cm). Note that these visual angles differ from those reported by Lauer et al. (2015), who reported that their stimuli were 7.5° x 6°. Visual inspection of the Tetris shapes we used and those presented in the original paper, however, shows that both Tetris objects were square (i.e., four squares tall and 4 squares across). The discrepancies in aspect ratios and visual angle, therefore, likely reflect differences in how visual angle was calculated and reported. Our visual angles were calculated for unrotated objects; calculating visual angle for the object rotated at 202 degrees at a viewing distance of 100 cm yields visual angles of 7.53° x 5.99°, which is nearly identical to that reported in Lauer et al. (2015).

Objects were presented in varying degrees of rotation in an on-off-on cycle throughout the trial in the following sequence: the object appeared for 500 ms on the screen, followed by a 300 ms blank white screen, and then the object reappeared for 500 ms at a different orientation (see Figure 1). Across our 60-s trials, infants saw the object reappear 75 times. The stimuli for each trial consisted of a *non-mirror-change* stream, in which the only change on each reappearance was the new orientation of the Tetris-like block, and a *mirror-change* stream, in which on every third appearance a mirror image of the object was presented. The images in the two streams generated for a trial were identical except when the mirror image appeared on the mirror stream. Thus, for each reappearance of the object, the degree of rotation was identical for the items in both streams. The two streams generated for each trial were presented on the left and right of a 90 cm wide screen (approximately 66 cm apart) against a gray background.

As in Lauer et al. (2016; 2015), we used four sets of orientation ranges (0 to 180 degrees, 90 to 270 degrees, 180 to 360 degrees, and 270 to 90 degrees), and each infant received one trial with each of the four ranges, order determined randomly for each infant. The Tetris-like object on each trial "rotated" in a pseudo-random sequence of orientations within the range (e.g., 0 to 180 degrees) with the constraints that each rotation was greater than 14 degrees and less than 90 degrees from the previous orientation. For example, on a trial with orientation ranges between 90 to 270 degrees, the object might have the following sequence of orientations: 90 degrees, 118 degrees, 146 degrees, and so on. Sequences were generated separately for each infant, yielding different unique sequences for each infant.

2.1.3. Apparatus. We used an Apple iMac computer to control the experiment and present the stimuli using Adobe Director (version 11). Stimuli were presented on a monitor (Sony Bravia 40" W600B, 90 cm wide by 51 cm high), positioned on a table, with a Panasonic wv-cp240ex video camera, connected to a Mac Mini using Elgato video capture software (version 1.1.2), positioned beneath the monitor, that captured the infants' head and torso. A black curtain divided the room, with the stimulus monitor (and camera lens) on one side, and all the other equipment out of the infants' view on the other side.

We recorded the assessment of infants' motor abilities using two Sony HDR-CX440 video cameras, which both fed to a Dell OptiPlex3040 computer and combined using VMIX software (version 20).

2.1.4. Procedure. Protocols for all Experiments were reviewed and approved by the Institutional Review Board (IRB) at the [*blinded for review*]. The mental rotation change detection task was administered in a dim, sound attenuated room. Infants were seated either on a caregiver's lap or in a highchair (strapped in using the straps installed by the manufacturer) approximately 100 cm from the stimulus monitor. Note that the infants tested by Lauer et al. (2015) were seated 70 cm from the monitor. However, as described earlier, the stimuli were the same visual angles, so infants should not have any more difficulty seeing the items in our experiment than in the original experiment. In addition, we used an approximate center-to-center distance between the objects in each stream of 68.35 cm, which corresponds to a 37.74° visual angle at 100 cm viewing distance, which closely matched the 61 cm, or 41° visual angle at 70 cm viewing distance reported by Lauer et al. Despite this, it is possible that this difference in viewing distance contributed to any difference in results observed.

To prevent bias, caregivers wore felt-covered glasses or were instructed to close their eyes to obstruct their view of the stimuli and were asked to refrain from speaking during the session unless their infant became fussy. Infants received four 60-s trials in which two stimulus streams, one to the right and one to the left, were presented. On each trial, the side on which the mirror stream occurred was randomly selected with the constraint that across the four trials infants received two trials in which the mirror stream was on the right and two trials in which the mirror stream was on the left. In-between trials, an attention getter (looming star with sound effect) was presented in the center of the stimulus monitor. An experimenter, seated behind the curtain and viewing the infant via video feed, initiated the trial by pressing a computer key when they judged that the infant's eyes were oriented towards the attention getter.

After the mental rotation task, infants' gross motor development was assessed in a separate room using the Alberta Infant Motor Scale (AIMS, Piper et al., 1992). The infant was placed on a surface covered with interlocking foam mats, with the caregiver seated nearby on a couch or on the floor, filling out questionnaires. The session was recorded using a video camera mounted on a tripod and positioned at a low level near the floor and a second handheld camera operated by a trained experimenter, who moved around to best capture the infants' position.

The experimenter, with the help of the caregiver as needed, positioned the infant into supine, prone, sitting, and standing postures, and recorded on the AIMS scoresheet the behaviors observed. The experimenter and caregiver flexibly engaged with the infant to assess motor abilities, such as rolling over or pulling to a standing position. Experimenters could use the video recording of the session to confirm their scoring of the session. Infants were then assigned a score based on the highest level attained in each posture.

2.1.5. Coding. We used Datavyu (https://datavyu.org/) to code infants' looking to the left and right stream on each trial. First, the onset of each trial was marked (using the offset of the attention-getter sound as an indicator). Trained coders, unaware of the stimulus orientation or location of the mirror stream, then marked the start and end of each look to the left or right during each trial. A second trained coder re-coded the video sessions for 20 infants (34% of the sample). Agreement between coders, calculated by dividing the number of frames in which the two coders agreed by the total number of frames in that trial, averaged across trials per participant, was high, *M* = 95%, *SD* = 1.6%, range = 88% - 99%.

2.1.6. Data processing. We exported the amount of left and right looking on each trial from Datavyu and combined it with output from the Director application to incorporate stimulus information on each trial. For each infant, we calculated a mirror preference score on each trial by dividing the amount of time the infant looked at the mirror stream by their total looking at both streams. We calculated the infants' overall mirror preference score by averaging the mirror preference score for the four completed trials.

All of the 59 eligible infants who completed the experiment met the inclusion criteria specified in our pre-registration (https://osf.io/vfa4q/?view_only=8f897e9f224e47e5a44a24c4c36826e5): (1) they had accumulated at least 5 s of looking on all four trials, (2) they did not have a side bias (e.g., greater than 90% of looking to either the left or right across all four trials), and (3) they did not have a mirror preference score that was more than 3 standard deviations from the group mean.

Infants' AIMS scores were calculated for supine, prone, sitting, and standing positions as defined by the AIMS assessment; our analysis was based on scores totaled across all four positions. Five infants failed to contribute usable data for at least one of the positions, so their data were excluded from the analysis examining the relation between motor development and mental rotation.

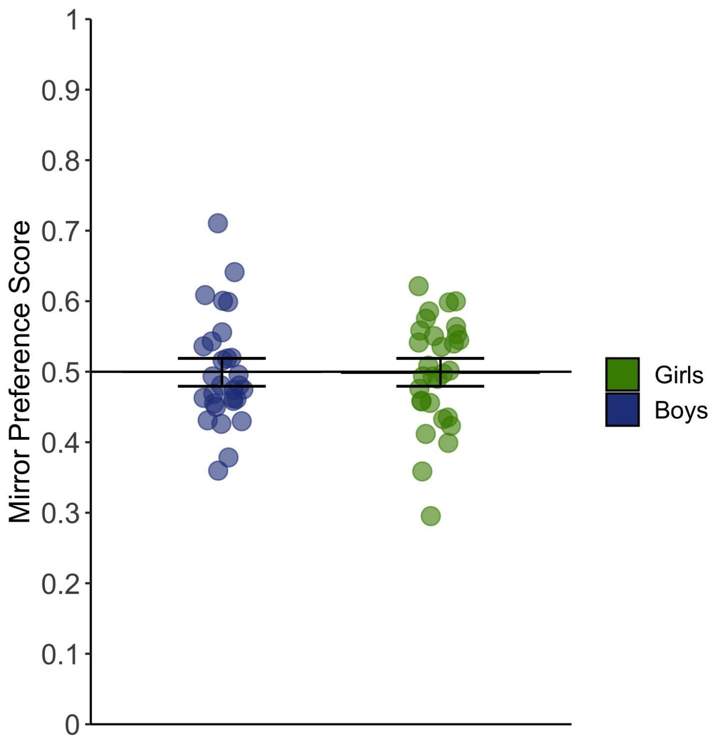
2.1.7. Statistical Approach. All statistical analyses were conducted using R version 4.1.2 (Winder et al., 2019), and the data and R scripts are available at OSF (https://osf.io/fvnxq/?view_only=3c2e875292b745f0bd42cc2f7d8e0b05). Our primary statistical approach was a series of t-tests and correlations using infants' mean mirror preference scores, consistent with previously published work (Lauer et al., 2015). We also conducted linear mixed-effects models (LMMs) on the trial-level data to accommodate both fixed and random effects, thus accounting for multiple sources of variation. However, these models did not yield any insight beyond the t-tests and are not reported here (see supplemental materials at OSF for full details and reporting of those results).

2.2. Results

Infants looked, on average, for 30.70 s (SD = 7.29) of the 60-s trials. They did not, however, show a preference for the mirror streams (i.e., their mirror preference scores were not greater than the chance level of .50, see Figure 2), and thus we did not find evidence of mental rotation in this sample. Although there was variability in individual infants' scores [1], as a group, infants did not show a significant preference for the mirror stream (M = .50, SD = .08), compared to chance (.50), $t(58) = -0.09$, p = .93, d = .01.

Figure 2. Infants' mean mirror preference scores. The height of the bar represents the mean score for all infants (note that the bars are barely visible because infants' mean mirror preference scores were at chance) and individual circles represent mean mirror preference scores for each infant. The line bisecting the y-axis represents chance performance (.50) and the error bars represent 95% confidence intervals. The data are disaggregated by infant sex (boys are represented by the blue dots and girls are represented by the green dots).

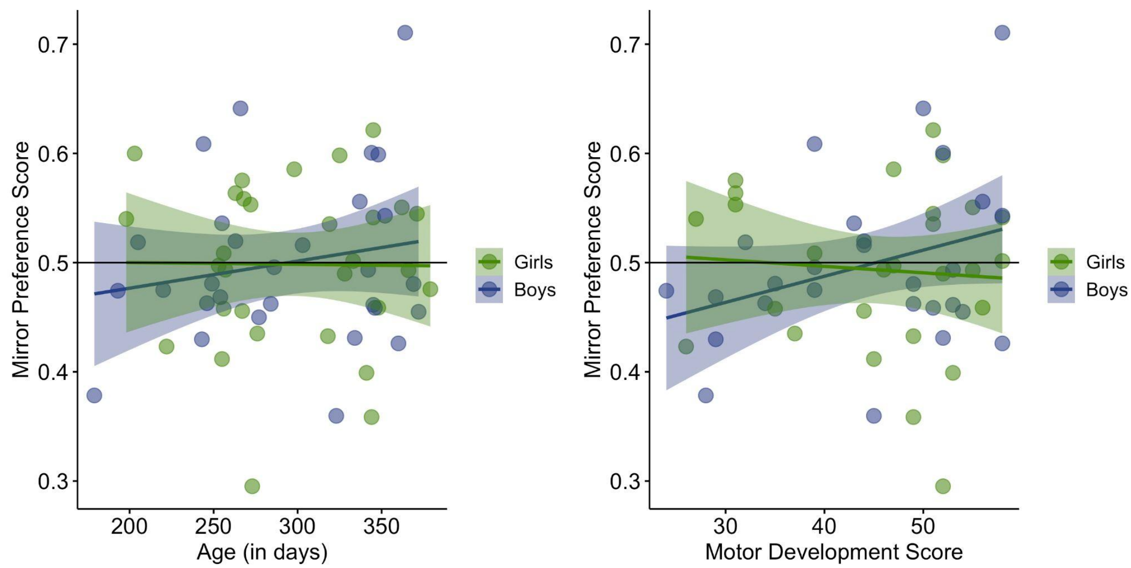
[1] It should be noted that linear mixed-effect modeling revealed similar within- and between-subject variation; thus, the observed variation in infants' scores likely reflects noise rather than individual differences in participants' preference scores.



To determine whether there was an effect of sex on mirror preference scores, as reported by Lauer et al. (2015), we next conducted an unpaired *t*-test (2-tailed) comparing boys and girls mean mirror preference scores. This analysis revealed no significant difference between boys (M = .50, SD = .08) and girls (M = .50, SD = .08), $t(57) = -0.06$, $p = .95$, $d = .02$. Not surprisingly, comparisons to chance did not reveal significant preferences for the mirrored stream for boys, $t(28) = -0.02$, $p = .98$, $d = .004$, or girls , $t(29) = -.11$, $p = .91$, $d = .02$. .

Next, we tested the effect of age on infants' mirror preference scores. As reported by Lauer et al. (2015), in our sample, mean mirror preference score was not significantly correlated with infant age, $r(57) = .09$, $p = .49$ (see Figure 3A). To connect with the broader literature, we also examined the correlations between infants' total AIMS scores and their mirror preference scores. As with age, this correlation revealed no significant relation between motor ability and mirror preference scores, $r(52) = .13$, $p = .36$ (see Figure 3B).

Figure 3. Infants' mean mirror preference as a function of age (left) and motor ability (right). Individual circles represent mean mirror preference scores for boys (blue) and girls (green). Shading around the lines represent 95 percent confidence intervals and regression lines indicate the correlation between each measure and mirror preference scores for boys and girls separately. The line bisecting the y-axis represents chance performance (.50).



We also fit a series of exploratory LMM on the mirror preference score to examine these relations. None of these models yielded any significant effects (see supplementary materials for details).

2.3. Discussion

Unlike the results reported by Lauer et al. (2015), our results failed to provide evidence of mental rotation ability in this task. Although we attempted a near exact replication of the

Lauer et al. procedure, there were slight differences in the viewing distances. We created our stimuli and presentation to control for those differences, but it is possible that the difference in distance made it harder for infants to demonstrate a change preference or for observers to accurately code infants' looks to the left and right. We think these alternatives are unlikely, however, given that previous studies have shown change preference at this viewing distance (Ross-Sheehy et al., 2003), and we obtained high reliability in our coding between independent coders. Nevertheless, it is possible that some aspect of our testing context or stimulus generation provided a less sensitive indicator of infants' ability to detect a change.

In Experiment 2, we further examined the robustness of this task by conducting a conceptual replication of the original study. In this new experiment, we used the same stimulus objects (i.e., the Tetris-like objects) and presented stimulus streams side-by-side. However, this experiment included several additional changes. Because this experiment was conducted during to the COVID-19 pandemic, we conducted it online using a web-based system. Although this introduced more variability in aspects of the procedure, it also meant that infants were tested in a context in which they were comfortable. In addition, the trials were shorter and we included a within-subject manipulation to address the possibility that our failure to replicate was due to differences in the difficulty of our stimuli.

3. Experiment 2

Experiment 2 was a pre-registered (https://osf.io/bmt53/?view_only=084cff7ab0b2439a9642d0a262721bc4) experiment conducted using *Lookit*, a platform for collecting data online. To determine whether infants' failure to prefer the mirror stream in Experiment 1 reflected, at least in part, the difficulty of the mental rotation task, in Experiment 2 we included trials in which the rotation cycles were sequential (i.e., the object always rotated in the same direction in 14 degree increments) as well as trials in which the rotation cycles were non-sequential as in Experiment 1.

3.1. Method

3.1.1. Participants. Our target sample size was 100 infants. Between 8/31/2020 and 1/20/2021, 150 infants between the ages of 6 and 12 months of age participated in our online study. Our final sample included 107 infants (*M*age = 256.50 days, *SD* = 57.38, range 180-363; 48 girls). Note that the mean age of infants in Experiment 2 was higher than those reported by Lauer et al. (2015) (8.49 months for girls and 8.60 months for boys in this study vs 9.71 months for girls and 10.54 months for boys in the original study). A sensitivity analysis confirmed that the effects reported below were not due to the slight difference in age of this sample (see results section for additional information).

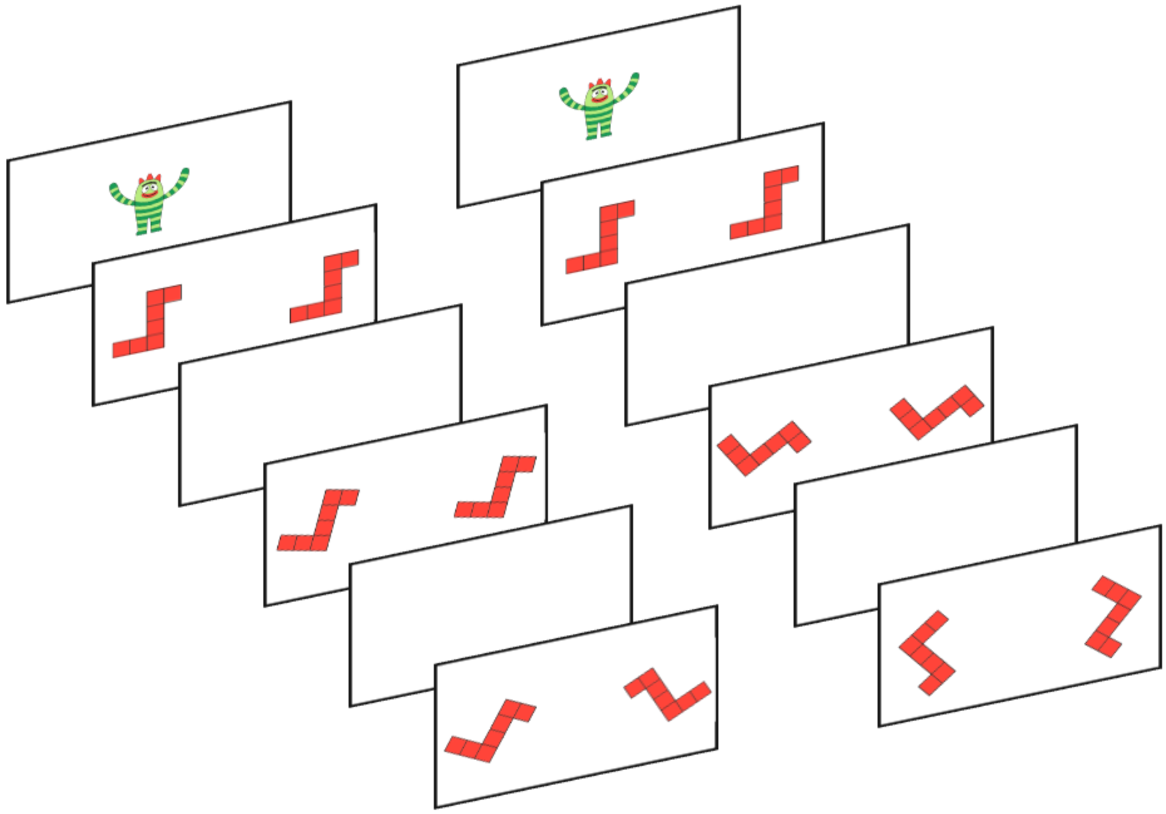
The infants in our sample resided across the US, representing 26 different states. Typical for studies conducted on Lookit, our infants were racially diverse (69 were White, 10 were Asian, 3 were Black or African American, 24 were multiracial, and race was not reported for 1 infant; of these infants, 6 White, 3 Black or African American, 3 multiracial, and 1 infant without racial information reported were also reported to be Hispanic or Latino) and highly educated (102 of the mothers having earned at least a 4-year degree, 4 having an associate's degree or some college, 1 had a high school diploma).

Of the 43 infants who participated but were not included in the final analyses, 17 were ineligible due to prematurity (e.g., born earlier than 37 weeks before their due date) or were not in the appropriate age range at the time of participation, 11 failed to provide usable data due to equipment error, 7 infants' caregiver or sibling interfered, 2 infants were fussy or their eyes were not visible, 2 infants were excluded because their data could not be reliably coded, and 4 infants failed to contribute at least 1 usable trial for each trial type. One caregiver consented but did not record their infant watching any of the test trials.

Participants were recruited from the Lookit recruiter, ads placed on social media posts, and informational emails we sent to eligible families identified as described in Experiment 1. Caregivers received a $5 Amazon gift certificate for participating.

3.1.2. Stimuli and Materials. As in Experiment 1, the stimulus streams used a red Tetris-shaped block (RGB: 215, 21, 33) that appeared in the same on-off-on cycle. However, our trials were only 20 s in the on-line version, motivated by the fact that infants in Experiment 1 looked only about 30 s on average during each trial and informed by our experience with collecting data online. Half of the stimuli were *sequential*, in which the Tetris stimuli rotated in 14 degree increments for 168 degrees in one direction before reversing in the other direction (e.g., if the object started at 0 degrees, it rotated in 14 degree increments clockwise until it reached 168 degrees, then rotated counterclockwise back). The other stimuli were *non-sequential*, in which the objects rotated through a 168-degree range, but each step was randomly chosen in terms of the degree and direction of rotation, with the constraint that each sequence of possible rotations had to be at least 42 degrees apart on subsequent rotation (see Figure 4).

Figure 4. Schematic illustration of the mental rotation change detection procedure used in Experiment 2. Sequential trials are displayed on the left and non-sequential trials are displayed on the right.



In addition, we had several 3-s attention getters, which were segments of children's videos (e.g., Cookie Monster saying "Oh boy! Oh boy! Happy!") that appeared in the center of the screen. We also created two 6-second calibration videos in which Elmo appeared for 3-s on the left and right (in one calibration Elmo appeared on the left first and in the other he appeared on the right first). Including this calibration video provided coders with an indication of how infants looked when there was a stimulus only on the left or only on the right of the screen.

3.1.3. Procedure. At a time of their choosing, caregivers logged on to the Lookit website, selected our experiment, and (because of restrictions from our IRB and funding source) confirmed that they were in the United States (if they indicated that they were not in the US, they exited the experiment). Caregivers then viewed a video introduction to the study, read the consent document, and recorded a video providing consent for their infant to participate. They also answered four questions about their child's gross motor development: whether, in the past week, their infant 1) sat, 2) crawled, 3) pulled to stand, and 4) walked. Finally, caregivers viewed a video providing information on how they should position themselves and their infant during the experiment and instructing them to close their eyes and refrain from interacting with their infant during the experiment.

When ready, caregivers pressed a button on the screen to initiate the study, which launched the experiment in full screen mode. The experiment had the following sequence: calibration, four test trials, calibration, and four more test trials. Each test trial started with an attention getter (e.g., the 3-s clip from a children's video) to orient the infants' attention to the screen. Immediately after the attention-getter, mirror-change and non-mirror-change stimulus streams were presented, one on the right and one on the left, for 20-s. There were four trials in each block (i.e., between calibration videos). Each block contained two sequential trials and two non-sequential trials, for a total of four trials of each type. The side of the mirror stream as well as trial type alternated within each block, such that infants saw the mirror stream on each side twice for each trial type (e.g., within a block, each infant saw one sequential trial and one non-sequential trial in which the mirror stream was on the right and the same was true for the left).

Approximately half (n = 58) of the infants were tested in an order that began with a sequential trial on the first trial, and half (n = 49) the infants were tested in an order that began with a non-sequential on the first trial. Regardless of order, the mirror-change stream was always on the right on the first trial. Subsequent trials alternated both the trial type and the side of the mirror stream (e.g., For the infants who saw a sequential trial with the mirror stream on the right on the first trial, they then saw on the second trial a non-sequential trial where the mirror stream was on the left, a sequential trial where the mirror stream was on the right, and finally a non-sequential trial where the mirror stream was on the left).

When the session ended, an audio recording informed caregivers that the study was over and that they could open their eyes. Caregivers were then given a brief description of the study and shown an image of all the stimuli presented.

3.1.4. Coding. Coding was conducted as described in the previous experiment. However, to deal with the variation in lighting and positioning of infants tested in their homes, coders were additionally instructed to identify periods in which infants' gaze position was relatively stable within a single region of the monitor for three consecutive frames before classifying the instance as a new bout of looking. A second trained coder re-coded a randomly selected 25% of the trials for each infant in each block (e.g., 1 trial in each block). The reliability for two infants was low (less than 80%) and their data were dropped from further analysis. For the remaining 107 infants, agreement between the two coders was high (M = 94.40%, SD = 4.0%, range = 82.90% - 99.25%).

3.1.5. Data processing. Looking data were exported from Datavyu and demographic information and information about the timing of specific events (e.g., onset of the experimental stimuli, onset of the webcam recording) for each session were downloaded from Lookit. This data was processed and combined using custom R scripts and look durations and mirror preference were calculated as in Experiment 1.

Data inclusion was applied on a trial-by-trial basis. We discarded trials that were not at least 20 s in duration (3 s of the attention getter followed by 17 s of the experimental trial); the length of the recordings varied somewhat due to variations in computers and internet speed. Seventy-three videos were determined to be too short; the average length of the final set of videos was 21.15 s, ranging from 20.001 s to 24.27 s, or 17 to 21 s of the experimental trial. We also excluded trials with caregiver/sibling interference ($n = 20$), infant fussiness ($n = 19$), or data that could not be coded due to obstruction of infant face in recording ($n = 7$). Our primary analyses were conducted on the remaining 743 videos.

3.2. Results and Discussion

Infants contributed, on average, 6.94 trials to these analyses, and looked overall on average 8.10 s (*SD* = 3.56 s) of the 17 to 20 s change detection trial. The average mirror preference scores are presented in Figure 5. The average score on sequential trials was .49 (*SD* = .13), which was not different from chance (.50), $t(106) = -.37$, $p = .71$ $d = .04$, and on non-sequential trials was 0.52 (*SD* = .12), which was also not different from chance, $t(106) = 1.16$, $p = .25$, $d = .11$. Examining the preferences separately by sex showed that neither boys ($M = .51$, $SD = .07$) nor girls ($M = .51$, $SD = .08$) in this sample had preferences that differed from chance, $ps > .57$, $d < .09$. Excluding boys and girls that were greater than 1 standard deviation below the mean ages reported by Lauer et al. (2015) for each group similarly indicated that neither boys, $t(27) = 0.502$, $p = .87$, $d = .03$, nor girls, $t(26) = .504$, $p = .81$, $d = .05$, displayed a significant preference for the mirror stream.

The correlation between age and mirror preference scores was not significant, $r(105) = .002$, $p = .98$ (Figure 6A). A linear regression model with mirror preference score as the outcome and motor ability (categorical: pre-locomotor, crawling, sitting, or standing) as the predictor (pre-locomotor was set as the reference level) revealed no significant impact of crawling, $\beta = 0.03$, $SE = 0.02$, $t = 1.34$, , $p = .18$, standing, $\beta = 0.02$, $SE = 0.02$, $t = 0.90$, $p = .37$, or walking, $\beta = -0.02$, $SE = 0.03$, $t = -0.59$, $p = .56$, on infants' mirror preference scores (see Figure 6B).

Finally, as in Experiment 1, we fit infants' mirror preference scores to a series of LMM. None of these models yielded any significant results (see supplemental materials for details).

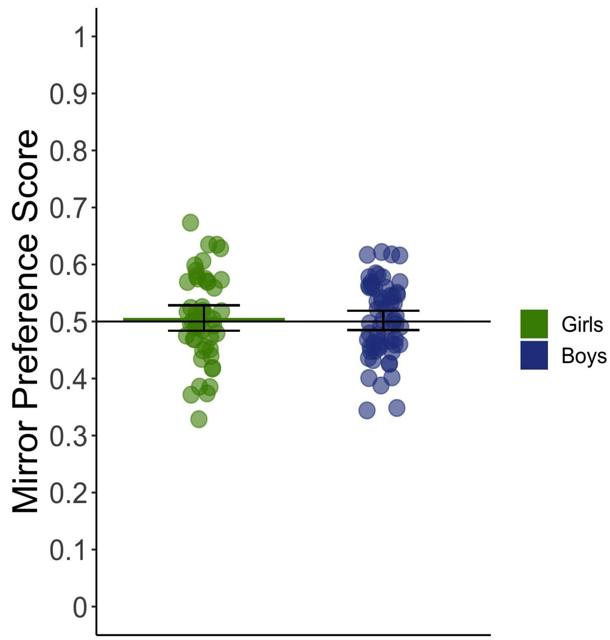
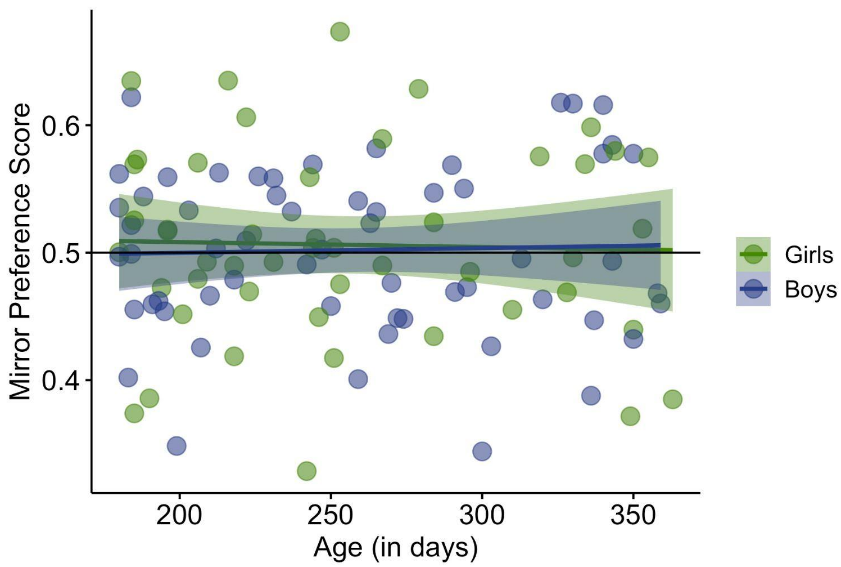


Figure 5. Infants' mean mirror preference scores. The height of the bar represents the mean score for all infants and individual circles represent mean mirror preference scores for each infant. The line bisecting the y-axis represents chance performance (.50) and the error bars represent 95% confidence intervals. Boys are represented by the blue dots and girls are represented by the green dots.



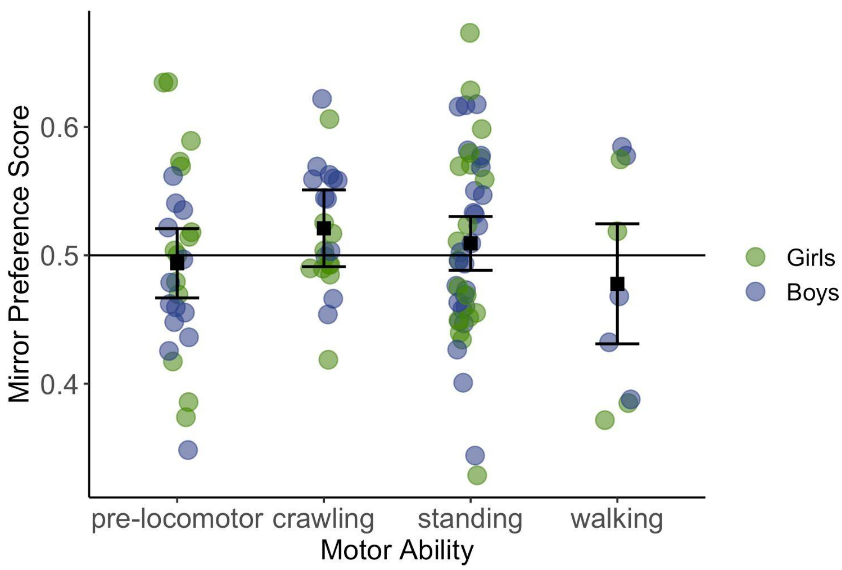


Figure 6. Infants' mean mirror preference as a function of age (top) and motor ability (bottom). Individual circles represent mean mirror preference scores for boys (blue) and girls (green). Error bars represent 95 percent confidence intervals and the black boxes (bottom) represent the mean score for all infants. Regression lines (top) indicate the correlation between age in days and mirror preference scores for boys and girls and the lines bisecting the y-axis represents chance performance (.50).

4. General Discussion

The goal of this study was to examine the robustness of infant mental rotation across variations in the change detection task. Neither experiment yielded evidence of mental rotation in this task. There were no effects of infant sex, age, or motor ability. The lack of an effect of age is consistent with previous findings using the same procedure (Lauer et al., 2015), but overall, we failed to replicate the finding that infants display mental rotation in this task.

It is important to point out that we failed to replicate both in a close replication of the original study (Experiment 1) and in a conceptual replication using online testing, a large sample size, and a condition designed to facilitate infants' object recognition across changes in orientation (Experiment 2). Thus, our failure to replicate is robust across variations in the change detection procedure. Lauer et al. (2015) do not provide demographic details of their sample, so it is possible that our samples differed in some important way from the original sample. Both of our samples were highly educated as well as racially and ethnically diverse: one sample was restricted to infants from [blinded for review] and the other included infants from across the United States. Further research would be needed to understand whether demographic differences contributed to the differences in results.

It is impossible to conduct an exact replication, particularly in a different lab and location, and thus even in Experiment 1 there were differences that may have contributed to the differences in the results. As described in the method section for Experiment 1, our infants were seated further away from the monitor than were infants in the original experiment. Although it is not obvious that this difference would have a large effect, particularly because we controlled for stimulus size to equate visual angle with the original study, it is possible that our testing context made it more difficult for infants to detect the change. In addition, we had a higher attrition than was reported by Lauer et al. (2015). Specifically, in Experiment 1, 21% of the infants tested did not provide data due to fussiness or inattention. In contrast, Lauer et al. report that 11% of their sample was excluded for such reasons. Although this may indicate that some aspect of our procedure was more difficult or challenging than that used by Lauer et al., it is not clear what that difference would be. In addition, far fewer infants were excluded due to such factors in Experiment 2 (only 2% were excluded due to fussiness) and yet this experiment also failed to replicate the original findings. Thus, although we are unable to rule out the possibility that the differences in results between Experiment 1 and the original Lauer et al. (2015) paper, we believe that such differences cannot fully explain our failure to replicate the original findings.

Experiment 2 was a conceptual replication of the original study and therefore was different in more ways. In addition to being conducted online, the infants were slightly younger than the sample in the original study, the trial durations were shorter, there were more trials, and there were two types of trials. However, some of these differences were designed to make the task easier for the infants. We included a sequential rotation trial type that should have made it easier for infants to detect when the mirror image was presented. Nevertheless, even in this condition we failed to obtain evidence of change detection.

More consistent results have been observed in studies testing infants' rotation of items held in long-term memory, specifically studies using habituation or violation of expectation procedures (see Moore & Johnson, 2020 for review). In contrast, the change detection task used here assesses infants' mental rotation for object representations briefly held in VSTM. Thus, it is possible that infants can robustly rotate object representations held in long term memory, but that their ability to rotate object representations held in short term memory is more fragile.

In conclusion, our findings are consistent with the interpretation that mental rotation in infancy is fragile, especially when assessing infants' ability to rotate items held in VSTM. These results suggest that the change detection may not be ideal for studying infant mental rotation. Additional studies will be necessary to further elucidate what factors affect whether infants display mental rotation in the change detection procedure as well as how various decision points such as experimental procedure and stimulus selection impact infants' mental rotation performance.

6. References

Cheng, Y.-L., & Mix, K. S. (2014). Spatial Training Improves Children's Mathematics Ability. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society*, *15*(1), 2–11.

Christodoulou, J., Johnson, S. P., Moore, D. M., & Moore, D. S. (2016). Seeing double: 5-month-olds' mental rotation of dynamic, 3D block stimuli presented on dual monitors. *Infant Behavior & Development*, *45*, 64–70.

Constantinescu, M., Moore, D. S., Johnson, S. P., & Hines, M. (2018). Early contributions to infants' mental rotation abilities. *Developmental Science*, *21*(4), e12613.

Erdmann, K., Kavšek, M., & Heil, M. (2018). Infants' looking times in a dynamic mental rotation task: Clarifying inconsistent results. *Cognitive Development*, *48*, 279–285.

Frick, A., & Möhring, W. (2013). Journal of Experimental Child Mental object rotation and motor development in 8- and 10-month-old infants. *Journal of Experimental Child Psychology*, *115*(4), 708–720.

Hespos, S. J., & Rochat, P. (1996). Tracking invisible spatial transformations by 4- and 6-month-old infants. *Infant Behavior & Development*, *19*, 504.

Hyde, J. S. (2007). New Directions in the Study of Gender Similarities and Differences. *Current Directions in Psychological Science*, *16*(5), 259–263.

Lauer, J. E., & Lourenco, S. F. (2016). Spatial Processing in Infancy Predicts Both Spatial and Mathematical Aptitude in Childhood. *Psychological Science*, *27*(10), 1291–1298.

Lauer, J. E., Udelson, H. B., Jeon, S. O., & Lourenco, S. F. (2015). An early sex difference in the relation between mental rotation and object preference. *Frontiers in Psychology*, *6*(MAY). https://doi.org/10.3389/fpsyg.2015.00558

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Development*, *56*(6), 1479–1498.

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology. General*, *145*(9), 1206–1227.

Möhring, W., & Frick, A. (2013). Touching up mental rotation: effects of manual experience on 6-month-old infants' mental object rotation. *Child Development*, *84*(5), 1554–1565.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: a sex difference. *Psychological Science*, *19*(11), 1063–1066.

Moore, D. S., & Johnson, S. P. (2011). Mental Rotation of Dynamic, Three-Dimensional Stimuli by 3-Month-Old Infants. In *Infancy* (Vol. 16, Issue 4, pp. 435–445). https://doi.org/10.1111/j.1532-7078.2010.00058.x

Moore, D. S., & Johnson, S. P. (2020). The development of mental rotation ability across the first year after birth. *Advances in Child Development and Behavior*, *58*, 1–33.

Newcombe, N. S., Booth, J. L., & Gunderson, E. A. (2019). Spatial skills, reasoning, and mathematics. In *The Cambridge Handbook of Cognition and Education* (pp. 100–123). Cambridge University Press.

Piper, M. C., Pinnell, L. E., Darrah, J., Maguire, T., & Byrne, P. J. (1992). Construction and validation of the Alberta Infant Motor Scale (AIMS). *Canadian Journal of Public Health. Revue Canadienne de Sante Publique*, *83 Suppl 2*, S46-50.

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*(11), 1067–1070.

Quinn, P. C., & Liben, L. S. (2014). A sex difference in mental rotation in infants: Convergent evidence. *Infancy: The Official Journal of the International Society on Infant Studies*, *19*(1), 103–116.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The Development of Visual Short-Term Memory Capacity in Infants. *Child Development*, *74*(6), 1807–1822.

Scott, K., Chu, J., & Schulz, L. (2017). Lookit (part 2): Assessing the viability of online developmental research, results from three case studies. *Open Mind*, *1*(1), 15–29.

Scott, K., & Schulz, L. (2017). Lookit (Part 1): A New Online Platform for Developmental Research. *Open Mind*, *1*(1), 4–14.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science (New York, N.Y.)*, *171*(3972), 701–703.

Slone, L. K., Moore, D. S., & Johnson, S. P. (2018). Object exploration facilitates 4-month-olds' mental rotation performance. *PloS One*, *13*(8), e0200468.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270.

Winder, C., Dowlatabadi, Z., & Miller-Zarneke, T. (2019). The Core Team. In *Producing Animation* (pp. 59–83). https://doi.org/10.1201/9780429490521-4