Received: 24 July 2022 | Revised: 10 February 2023 | Accepted: 12 February 2023

DOI: 10.1111/desc.13381

Developmental Science WILEY

RESEARCH ARTICLE

A meta-analysis of mental rotation in the first years of life

Alexander Enge[1,2] | Shreya Kapoor[1] | Anne-Sophie Kieslinger[1] | Michael A. Skeide[1]

[1]Research Group Learning in Early Childhood, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[2]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany

Correspondence
Alexander Enge, Research Group Learning in Early Childhood, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany.
Email: enge@cbs.mpg.de

Funding information
German Research Foundation, Grant/Award Numbers: 433715509, 433758790; European Research Council, Grant/Award Number: 850752; Jacobs Foundation Research Scholarship; Studienstiftung des deutschen Volkes

Abstract

Mental rotation, the cognitive process of moving an object in mind to predict how it looks in a new orientation, is coupled to intelligence, learning, and educational achievement. On average, adolescent and adult males solve mental rotation tasks slightly better (i.e., faster and/or more accurate) than females. When such behavioral differences emerge during development, however, remains poorly understood. Here we analyzed effect sizes derived from 62 experiments conducted in 1705 infants aged 3–16 months. We found that male infants recognized rotated objects slightly more reliably than female infants. This difference survives correction for small degrees of publication bias. These findings indicate that gender differences in mental rotation are small and not robustly detectable in the first months of postnatal life.

KEYWORDS
gender, infants, mental rotation, meta-analysis, spatial cognition

Research Highlights

- We analyzed effect sizes of 62 mental rotation experiments including 1705 infants.
- Looking time reveals that 3–16-months-old infants are able to perform mental rotation.
- Mental rotation is slightly more reliable in male infants compared to female infants.
- Gender difference in mental rotation is robust to small degrees of publication bias.

1 | A META-ANALYSIS OF MENTAL ROTATION IN THE FIRST YEARS OF LIFE

The cognitive ability to move visual object representations in mind for recognition across different orientations, known as mental rotation, emerges gradually in the course of cognitive development (Johnson & Moore, 2020; Moore & Johnson, 2020). Mental rotation is a key component of intelligence and a predictor of learning outcome and educational achievement (Hegarty & Kozhevnikov, 1999; Johnson & Bouchard Jr., 2005; Shepard & Metzler, 1971). Moreover, mental rotation has been the subject of long and widespread research attention because it has consistently yielded one of the largest and most reliable gender differences of any cognitive task (Lauer et al., 2019). Successful mental rotation performance during childhood is associated with later math and science achievement (e.g., Casey et al., 2015; Cheung et al., 2020; Frick, 2019; Geer et al., 2019; Gilligan et al., 2017; Gunderson et al., 2012; van Tetering et al., 2019), while reduced mental rotation performance has been associated with several developmental disorders (e.g., dyslexia: Rusiak et al., 2007; Rüsseler et al., 2005; Williams syndrome: Stinton et al., 2008). Recent adaptations of mental rotation tasks for infants have provided preliminary evidence that mental rotation performance at 6 months of age predicts spatial skills and symbolic math during childhood (4 years of age; Lauer & Lourenco, 2016). However, it is unclear if these effects persist into adolescence or adulthood and if there are gender differences with respect to this longitudinal relationship. In fact, individual studies with infants and toddlers (0–36 months of age) have typically relied on small sample sizes (often $n \leq 20$) and thus provided conflicting answers to the question of whether or not participants in this age range are able to perform mental rotation above chance level, or to the question at what age mental rotation abilities first emerge (Johnson & Moore, 2020; Moore & Johnson, 2011).

Previous meta-analyses revealed that in adolescence and adulthood, males solve mental rotation tasks better (i.e., faster and/or more accurate) than females on average (Linn & Petersen, 1985; Maeda & Yoon, 2013; Voyer et al., 1995; Voyer, 2011). Effect sizes of this difference, however, are heterogeneous and often only medium in size (mean weighted $g = 0.37$–0.73). Interestingly, a recent meta-analysis in 3–17-year-old children and adolescents suggests only a small-to-medium difference (mean weighted $g = 0.39$; Lauer et al., 2019) which increases at a rate of 0.04 units per year. Whether gender differences in mental rotation behavior already emerge during infancy remains unknown. Individual studies have provided conflicting evidence (e.g., Moore & Johnson, 2008 vs. Erdmann et al., 2018) which might be explained by methodological challenges of examining mental rotation at such an early age. Specifically, infant research relies on less difficult tasks that might make gender differences more difficult to detect (Lauer et al., 2019). Another explanation might be the reliance on looking time rather than speed or accuracy as a measure of mental rotation proficiency (Moore & Johnson, 2020). Moreover, the field has just begun to examine the underlying factors that could contribute to such differences, including genetic influences (for adults, see Shakeshaft et al., 2016), different levels of sex hormones (Constantinescu et al., 2018; Erdmann et al., 2019; Toivainen et al., 2018), parental attitudes about gender (Constantinescu et al., 2018), or aspects of motor development such as manual experience with objects and crawling ability (Möhring & Frick, 2013; Schwarzer et al. 2013).

In the present study, we meta-analyzed 62 effect sizes derived from looking times in mental rotation tasks performed by 1705 infants (47.8% female) aged 3 to 16 months (mean age of 7 months 11 days; Table 1) (Antrilli & Wang, 2016; Christodoulou et al., 2016; Constantinescu et al., 2018; Erdmann et al., 2018; Frick & Möhring, 2013; Frick & Wang, 2014; Gerhard & Schwarzer, 2018; Gerhard-Samunda et al., 2021; Hespos & Rochat, 1997; Kaaz & Heil, 2020; Kelch et al., 2021; Lauer et al., 2015; Möhring & Frick, 2013; Moore & Johnson, 2008, 2011; Quinn & Liben, 2008, 2014; Rochat & Hespos, 1996; Schwarzer, Freitag, Buckel, et al., 2013; Schwarzer, Freitag, & Schum, 2013; Slone et al., 2018). All tasks were embedded either into habituation experiments (44 experiments) or violation of expectation experiments (18 experiments) (Baillargeon et al., 1985; Fantz, 1964). These experiments comprised real world stimuli (e.g., toy objects; Antrilli & Wang, 2016), three-dimensional digital stimuli (e.g., cube figures; Moore & Johnson, 2008), or two-dimensional digital stimuli (e.g., digits; Quinn & Liben, 2008). In habituation experiments, infants repeatedly saw an object (e.g., a three-dimensional cube figure) until their looking times declined and were then presented with a mirror image of the object or with the familiar object at a new angle. Longer looking times at the mirror image were taken as evidence that an infant still recognized the familiar object after rotation. In violation of expectation experiments, infants were viewing an object (e.g., the two-dimensional letter “P”) that was rotating on a circular trajectory for a certain amount of time before it disappeared behind an occluder on that trajectory. A bit later, the occluder was removed and revealed either the same object as before (but rotated further, as expected) or an inverted (unexpected) version of the object. As in habituation experiments, longer looking times at the inverted object as compared to the expected object were taken as evidence for infants’ mental rotation capability.

Although some of these studies also examined individual differences in mental rotation performance and their genetic, hormonal, and/or social-environmental correlates (e.g., Constantinescu et al., 2018; Frick & Möhring, 2013; Slone et al., 2018), we here focused on cross-sectional, group level effects. We hypothesized (a) that infants in the current age range would be able to perform mental rotation at above-chance level, as indicated by longer looking times for novel as compared to familiar rotated objects, and (b) that this mental rotation performance would be slightly greater in male infants as compared to female infants. Both of these effect sizes were assumed to be small, given (a) that mental rotation skill might not yet be fully developed at the age of interest and (b) that the gender difference in mental rotation performance has been shown to increase during childhood (Lauer et al., 2019). We reasoned that a small effect size would corroborate the theoretically important notion that increases in gender difference over the lifespan are better explained by an accumulating influence of social-environmental factors rather than biological-genetic factors.

Our analytic plan was to test (a) whether children are already capable of performing mental rotation in the first 3 years of life and whether early mental rotation performance (b) reveals gender differences and (c) differs as a function of the age group studied and the task used. To this end, we ran Bayesian and frequentist three-level meta-analytic and meta-regression models complemented by funnel plots and Egger regression tests to examine potential publication bias.

2 | METHOD

2.1 | Protocol

In the present meta-analysis, we followed the established PRISMA 2020 guidelines (Preferred Reporting Items for Systematic reviews and Meta-Analyses; Page et al., 2021). Figure 1 displays the PRISMA flowchart while the PRISMA checklist is provided in Table S1. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

2.2 | Eligibility criteria

Articles needed to fulfill six criteria for being included in this meta-analysis: (1) The article is written in English or German; (2) The article includes results from a group study with human samples (thus excluding review articles, meta-analyses, case studies, and animal stud-

TABLE 1 Experiments included in the main analysis.

Article	Experiment	Sample size	Females	Age ($M \pm SD$)[a]	Task	Stimulus type	Stimulus dimensions
Rochat and Hespos (1996)	Experiment 1	30	11	6m 15d ± n/a[b]	VoE[c]	Real	3D
Hespos and Rochat (1997)	Experiment 2	21	9	5m 9d ± n/a	VoE	Real	3D
	Experiment 3	19	8	5m 16d ± n/a	VoE	Real	3D
	Experiment 4, 4-months-old infants	10	n/a	4m 13d ± n/a	VoE	Real	3D
	Experiment 4, 6-months-old infants	9	n/a	6m 2d ± n/a	VoE	Real	3D
	Experiment 5, 6-months-old infants	10	n/a	6m 16d ± n/a	VoE	Real	3D
	Experiment 6, 6-months-old infants	10	n/a	6m 16d ± n/a	VoE	Real	3D
Moore and Johnson (2008)	Females	20	20	5m 1d ± 10d	Habituation	Digital	3D
	Males	20	0	5m 1d ± 10d	Habituation	Digital	3D
Quinn and Liben (2008)	Females	12	12	3m 20d ± 10d	Habituation	Real	2D
	Males	12	0	3m 15d ± 12d	Habituation	Real	2D
Moore and Johnson (2011)	Females	20	20	3m 5d ± 12d	Habituation	Digital	3D
	Males	20	0	3m 5d ± 12d	Habituation	Digital	3D
Frick and Möhring (2013)	10-months-old infants	20	10	10m 21d ± 20d	VoE	Digital	2D
	8-months-old infants	20	10	8m 1d ± 8d	VoE	Digital	2D
Möhring and Frick (2013)	Exploration group	20	10	6m 2d ± 9d	VoE	Digital	2D
	Observation group	20	10	5m 26d ± 8d	VoE	Digital	2D
Schwarzer et al. (2013)	Crawling infants	24	11	9m 3d ± n/a	Habituation	Digital	3D
	Non-crawling infants	24	11	9m 3d ± n/a	Habituation	Digital	3D
Schwarzer et al. (2013)	Crawling infants	24	12	9m 5d ± n/a	Habituation	Digital	3D
	Non-crawling infants	24	10	9m 5d ± n/a	Habituation	Digital	3D
Frick and Wang (2014)	Experiment 1, 14-months-old infants	14	6	13m 23d ± n/a	VoE	Real	3D
	Experiment 1, 16-months-old infants	14	6	15m 23d ± n/a	VoE	Real	3D
	Experiment 2, 14-months-old infants	14	4	14m 8d ± n/a	VoE	Real	3D
	Experiment 2, 16-months-old infants	14	7	15m 20d ± n/a	VoE	Real	3D
	Experiment 3, other-turning group	14	6	13m 25d ± n/a	VoE	Real	3D
	Experiment 3, self-turning group	14	6	14m 9d ± n/a	VoE	Real	3D
Quinn and Liben (2014)	Experiment 2, 6-months-old females	12	12	6m 11d ± 17d	Habituation	Real	2D
	Experiment 2, 6-months-old males	12	0	6m 3d ± 13d	Habituation	Real	2D
	Experiment 2, 9-months-old females	12	12	9m 6d ± 13d	Habituation	Real	2D
	Experiment 2, 9-months-old males	12	0	9m 3d ± 9d	Habituation	Real	2D

(Continues)

TABLE 1 (Continued)

Article	Experiment	Sample size	Females	Age (*M* ± *SD*)[a]	Task	Stimulus type	Stimulus dimensions
Lauer et al. (2015)	Females	28	28	9m 21d ± 52d	Habituation	Digital	2D
	Males	28	0	10m 16d ± 57d	Habituation	Digital	2D
Antrilli and Wang (2016)	Vertical-stripes group	16	9	14m 4d ± n/a	VoE	Real	3D
Christodoulou et al. (2016)	Females	24	24	5m 0d ± 7d	Habituation	Digital	3D
	Males	24	0	5m 0d ± 7d	Habituation	Digital	3D
Constantinescu et al. (2018)	Females	28	28	5m 14d ± 10d	Habituation	Digital	3D
	Males	26	0	5m 14d ± 10d	Habituation	Digital	3D
Erdmann et al. (2018)	5-months-old females	104	104	5m 13d ± 10d	Habituation	Digital	3D
	5-months-old males	104	0	5m 12d ± 9d	Habituation	Digital	3D
	9-months-old females	84	0	9m 11d ± 10d	Habituation	Digital	3D
	9-months-old males	84	84	9m 11d ± 12d	Habituation	Digital	3D
Gerhard and Schwarzer (2018)	Crawling infants, 0° group	19	9	9m 13d ± 8d	Habituation	Digital	3D
	Crawling infants, 54° group	20	9	9m 13d ± 8d	Habituation	Digital	3D
	Non-crawling infants, 0° group	18	8	9m 13d ± 8d	Habituation	Digital	3D
	Non-crawling infants, 54° group	19	9	9m 13d ± 8d	Habituation	Digital	3D
Slone et al. (2018)	Mittens-first group, females	20	20	4m 16d ± 10d	Habituation	Digital	3D
	Mittens-first group, males	20	0	4m 16d ± 10d	Habituation	Digital	3D
	Mittens-second group, females	20	20	4m 16d ± 10d	Habituation	Digital	3D
	Mittens-second group, males	20	0	4m 16d ± 10d	Habituation	Digital	3D
Kaaz and Heil (2020)	Experiment 1, females	144	144	6m 9d ± 7d	Habituation	Digital	2D
	Experiment 1, males	144	0	6m 10d ± 8d	Habituation	Digital	2D
	Experiment 2, females	48	48	6m 14d ± 8d	Habituation	Digital	2D
	Experiment 2, males	48	0	6m 12d ± 8d	Habituation	Digital	2D
Gerhard-Samunda et al. (2021)	Horizontal-stripes group, crawling infants	11	3	9m 5d ± 9d	Habituation	Digital	3D
	Horizontal-stripes group, non-crawling infants	10	5	9m 3d ± 8d	Habituation	Digital	3D
	Vertical-stripes group, crawling infants	11	2	9m 4d ± 9d	Habituation	Digital	3D
	Vertical-stripes group, non-crawling infants	11	6	9m 3d ± 6d	Habituation	Digital	3D
Kelch et al. (2021)	Experiment 1, crawling infants	11	10	9m 17d ± 8d	Habituation	Real	3D

[a]Mean ± standard deviation
[b]Not available
[c]Violation of expectation.

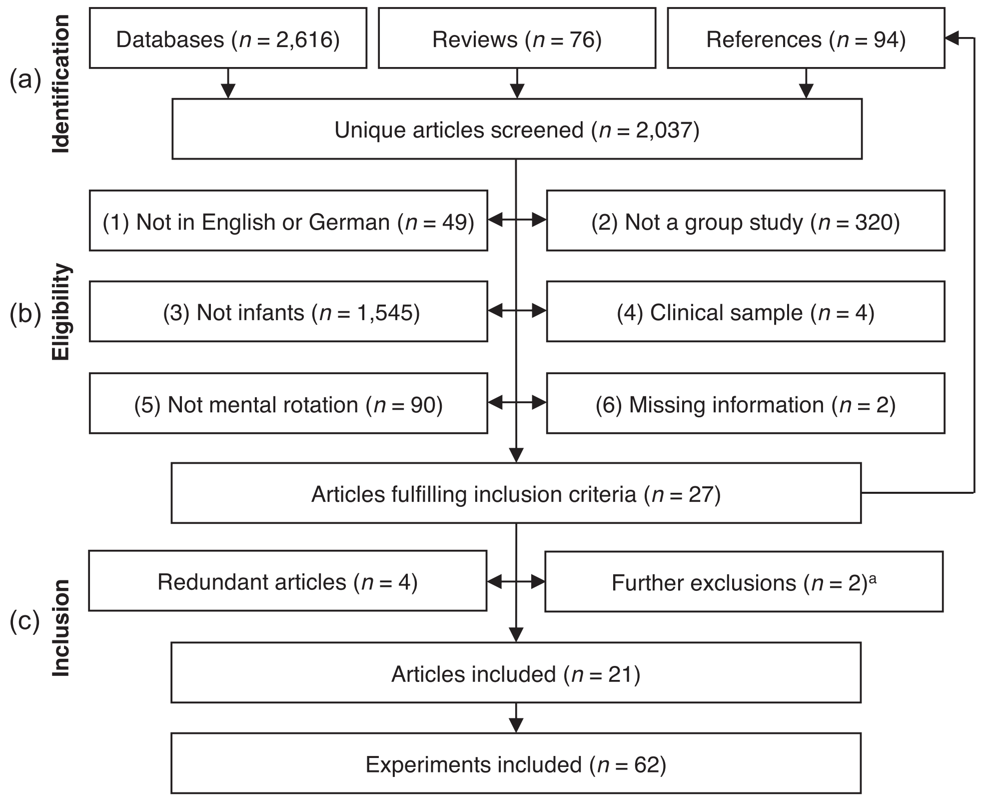

FIGURE 1 Literature search and selection process. (a) We searched four online databases as well as reviews and reference sections to identify articles in which mental rotation experiments in infants were reported. (b) Experiments were included in the meta-analysis if they fulfilled six pre-specified inclusion criteria. (c) Redundant articles comprising the same experiment(s) were excluded. [a]We decided to exclude two additional articles: one article because it was based on an uncommon experimental paradigm, differing substantially from all other articles, and another article because it was based on a sample of infants who were substantially older compared to all other articles.

ies); (3) These samples include at least one group of infants (mean age between 0 months and 36 months); (4) Infants were not born preterm and had no clinical diagnosis; (5) Infants performed a mental rotation task; (6) The article contains quantitative scores that can be converted into a standardized mean difference (see Data collection process and items below). We explicitly refrained from excluding works that were not peer reviewed (e.g., dissertations and preprints) to reduce the impact of publication bias. The age range (0–36 months; Criterion (2)) was chosen because this range had not been covered by previous meta-analyses of gender differences in mental rotation (Lauer et al., 2019; Linn & Petersen, 1985; Maeda & Yoon, 2013; Voyer et al., 1995; Voyer, 2011).

2.3 | Information sources and search strategy

We entered the search terms ("mental rotation" OR "mental transformation" OR "spatial rotation" OR "spatial transformation" OR "spatial ability" OR "spatial skills") AND ("infant" OR "infants" OR "infanthood" OR "toddler" OR "toddlers" OR "toddlerhood" OR "child" OR "children" OR "childhood" OR "month" OR "months") into four online databases (APA PsycINFO, PubMed/MEDLINE, Scopus, and ProQuest Dissertations & Theses Global). All database queries were completed on December 6, 2021. We configured the databases to check for article titles, abstracts, and keywords while applying no other filters or limits. This yielded 2616 articles in total, 1954 of which remained after removing duplicate records (Figure 1). We further identified 76 articles by screening the reference sections of previous reviews and meta-analyses on mental rotation and related skills (Frick et al., 2014; Johnson & Moore, 2020; Kubicek & Schwarzer, 2018; Lauer et al., 2019; Linn & Petersen, 1985; Moore & Johnson, 2020; Uttal et al., 2013; Voyer et al., 1995; Yang et al., 2020). Of these, 34 articles had not been covered by the database search. We also identified 94 articles by screening the reference sections of all publications that had been included after the first pass of the selection process. Of these, 49 articles had not been covered by the database search. Accordingly, we screened 2037 unique articles in total.

2.4 | Selection process

The final sample size of articles was determined by including all articles that fulfilled the inclusion criteria described above. Two independent raters read the abstract and, if necessary, relevant sections of the full text to check if an article fulfilled the inclusion criteria. Interrater agreement for the binary decision to include versus exclude an article was 98.5% (κ_w [Cohen's weighted kappa] = 0.67, 95% CI [0.55, 0.78]). Interrater agreement for the specific eligibility criteria was 88.2% (κ_w = 0.72, 95% CI [0.40, 1.00]). Cases where the two ratings diverged were resolved via discussion among all raters until a consensus was reached. One article (Mash et al., 2007) was excluded because the authors used a unique mental rotation paradigm that was not comparable to the paradigms used in the other articles. Another article (Pedrett et al., 2020) was excluded because the average age of the infants studied (30.7 months) was almost twice as high as the average age of the next article (15.8 months; $z = 7.20$ compared to all articles). To prevent any potential age-related effects from being distorted by this single outlier study, we decided to narrow our analysis from the first 3 years of life to the first 16 months of life. This procedure led to a total of 21 articles being included in the meta-analysis (Table 1).

Many of these articles consisted of multiple experiments, for example, using different variations of the mental rotation task or different subsamples of infants. We included all of these experiments in the meta-analysis and accounted for the dependencies between them by means of multilevel modeling with by-article random effects (see Bayesian meta-analysis below). When there was insufficient information in the original article to compute a standardized effect size (see Data collection and process items below), we contacted the corresponding author using a standardized email template plus up to two reminder emails. We excluded articles and/or experiments if after this procedure we still did not have sufficient information to compute an effect size.

Whenever an article reported separate effect sizes for males and females—or other subgroups like crawling and non-crawling infants—but also an effect size combining these groups, we included the most group-specific effect size. If this was not the case, we included either the group-specific effect size (i.e., separate effect sizes for all-male and all-female subgroups) or the combined effect size (i.e., across male and female infants), depending on which of the two was available. We refer to these groups using the term "experiment." Moreover, we disregarded effect sizes that were clearly based on the same data but reported in different articles. This procedure led to a total of 62 experiments being included in the meta-analysis of mental rotation performance and 30 experiments being included in the meta-analysis of gender differences in mental rotation performance.

2.5 | Data collection process and items

Outcome measures and other relevant variables were extracted from each article by one of three raters and verified by a second rater. For the meta-analysis of mental rotation performance, outcome measures were any summary statistic (Table S2) that could be used to determine the standardized mean difference between novel/unexpected rotation events and familiar/expected rotation events. Other extracted variables included, if available, the sample size, the number of males and females, the mean age and its standard deviation, the minimum and maximum age, the type of mental rotation task (habituation or violation of expectation), the modality of stimulus presentation (real objects or objects on a computer screen), and the dimensionality of the stimuli (2D or 3D; Table 1).

We also conducted a meta-analysis of the gender differences in mental rotation performance observed within the original articles. For this analysis, the outcome measures were any summary statistic (Table S3) that could be used to determine the standardized mean difference between male infants' mental rotation performance and female infants' mental rotation performance. Other extracted variables included the sample size, the mean age and its standard deviation, and the minimum and maximum age of each gender group.

2.6 | Effect measures

For the meta-analysis of mental rotation performance, one outcome measure per experiment was converted into a standardized mean difference with small sample correction (Hedges' g) using the formulas provided in Table S2 (Cumming, 2012; Goulet-Pelletier & Cousineau, 2018; Hedges & Olkin, 1985; Lakens, 2013; Rosenthal, 1991). The standard error of Hedges' g for each experiment was computed using the formula provided by Goulet-Pelletier and Cousineau (2018):

$$SE_{rotation} = \sqrt{\frac{df}{df-2}\frac{2(1-r)}{n}\left(1+g_{rotation}^2\frac{n}{2(1-r)}\right)-\frac{g_{rotation}^2}{J(df)^2}}$$

where n is the sample size of the experiment, df are the degrees of freedom (with $df = 2[n - 1]$), r is the correlation between the two dependent measures in the experiment, and $J(df)$ is the correction factor for small samples as described in Table S2. The correlation r was not reported in any of the original articles (see also Harrer et al., 2021). We therefore always assumed a correlation of $r = 0.50$ to make our analysis comparable to standard (between-group) meta-analyses (Cohen, 1988; Lakens, 2013; Morris & DeShon, 2002) and because we were able to infer an average correlation of $r \approx 0.50$ from a subsample of articles which provided sufficient information (Supplementary Methods 1). A sensitivity analysis indicated that changing the assumed correlation to values from $r = -0.90$ via $r = 0.00$ to $r = 0.90$ had no meaningful impact on the meta-analytic effect sizes (Tables S4 and S5).

For the meta-analysis of gender differences in mental rotation performance, one outcome measure per contrast between male and female infants was converted into a standardized mean difference with small sample correction (Hedges' g) using the formulas provided in Table S3. In some cases where the authors of the original articles did not observe a statistically significant gender difference, they did not report a precise outcome measure. In these cases, we assumed an effect size of $g = 0.00$, thus rendering our meta-analytic effect size of the gender difference more conservative. The standard error of Hedges' g for each gender contrast was computed using the formula provided by Goulet-Pelletier and Cousineau (2018):

$$SE_{gender} = \sqrt{\frac{df}{df-2}\frac{2}{\tilde{n}}\left(1+g_{gender}^2\frac{\tilde{n}}{2}\right)-\frac{g_{gender}^2}{J(df)^2}}$$

where df are the degrees of freedom (with $df = n_{female} + n_{male} - 2$), $\tilde{n}$ is the harmonic mean of the group sizes (i.e., $\tilde{n} = 2/[n_{female}{}^{-1} + n_{male}{}^{-1}]$), and $J(df)$ is the correction factor for small samples as described in Table S3.

2.7 | Bayesian meta-analysis

We synthesized the effect sizes and their sampling variances using a Bayesian multilevel model. This model had three levels, with infant participants nested in experiments and experiments nested in articles (Van den Noortgate et al., 2013). We used a weakly-informative N (0, 1) (normal) prior for the meta-analytic effect size and a weakly-informative HC (0, 0.3) (half-Cauchy) prior for the two standard deviations (i.e., the random effects of experiments and articles; Williams et al., 2018). A prior sensitivity analysis indicated that making these priors either more informative or less informative did not have a strong influence on the meta-analytic results (Tables S4, S5, and S6). For the meta-analysis of mental rotation performance, the dependent variable was the standardized mean difference (Hedges' g) between the novel and familiar rotation condition, weighted by its standard error (see Effect measures above). For the meta-analysis of gender differences, the dependent variable was the standardized mean difference (Hedges' g) between male infants' mental rotation performance and female infants' mental rotation performance, weighted by its standard error. All Bayesian models were fitted using the brms package (Version 2.16.3; Bürkner, 2017, 2018) in R (Version 4.1.2; R Core Team, 2021) and the Stan language (Version 2.21.3; Stan Development Team, 2022). Markov Chain Monte Carlo (MCMC) sampling was used with four parallel chains, each sampling 20,000 draws (including 2000 warm-up draws) from the posterior distribution. To verify the convergence of the Markov chains, we examined rank plots as well as the $\hat{R}$ and N_{eff} statistics (Vehtari et al., 2021; Figure S1). For reporting, the credible interval (CrI) for each model parameter was computed as the 95% equal-tailed interval (ETI) of its posterior distribution, although replacing this with the 95% highest density interval (HDI) yielded highly similar results (Kruschke, 2015).

2.8 | Bayesian meta-regression

We examined the influence of three moderator variables on the mental rotation outcomes across experiments, namely (a) the gender of the sample of infants, (b) the age of the sample of infants, and (c) the type of mental rotation task. Gender was coded as a categorical predictor (male sample, female sample, mixed-gender sample) and contrast-coded using two successive difference contrasts so that we could compare male samples versus female samples and female samples versus mixed-gender samples (Schad et al., 2020). Age was coded as a continuous predictor in years and centered by subtracting the average across all experiments. Task type was coded as a categorical predictor (habituation task, violation of expectation task) and contrast-coded using a scaled sum contrast (Schad et al., 2020). We included task type as a predictor because this might reduce the error variance of the model or, in other terms, explain some of the heterogeneity in the observed effect sizes.

We then included these predictors for gender, age, and task type as well as two predictors for the interaction between gender and age (i.e., [male − female] × age, [female − mixed] × age) into a Bayesian meta-regression model. This model was based on the same random effects structure, sampling parameters, and prior specification as described above, but adding a weakly-informative N (0, 0.5) (normal) prior for all slope parameters.

2.9 | Frequentist meta-analysis

We verified the results obtained from our Bayesian analyses using classical frequentist meta-analysis and meta-regression. To this end, we used the metafor package (Version 3.0.2; Viechtbauer, 2010) in R to specify the same three-level models as described above but without the Bayesian priors (Table S7). These models were fitted using restricted maximum likelihood estimation (REML). To verify that this procedure converged on the most probable estimates, we examined profile likelihood plots (Raue et al., 2009; Viechtbauer, 2010) for the two variance components in the model (i.e., the random effects of experiments and articles; Figure S2).

2.10 | Publication bias assessment

Publication bias manifests itself in the form of published experiments reporting false positive effects and in the form of experiments not getting published when failing to obtain statistically significant effects. This bias was evaluated based on funnel plots and Egger regression tests (Egger et al., 1997; Sterne & Egger, 2005). Funnel plots visualize the relationship between standard errors and effect sizes. They were created by adapting code from the R package metaviz (Version 0.3.1; Kossmeier et al., 2020). The Egger regression test is a formal statistical test for this relationship between standard errors and effect sizes, probing if the weighted linear regression weight of the effect sizes on the standard errors is significantly different from zero. This test was performed using the metafor package and applying a two-sided false-positive error rate of $\alpha = 0.05$.

To scrutinize the robustness of the meta-analytic effect size against the influence of any individual experiment (which may or may not be a false positive), we conducted a jackknife (leave-one-out) analysis for the meta-analysis of mental rotation performance (Efron & Tibshirani, 1993). To this end, we refitted the Bayesian three-level model repeatedly while leaving out one of the original experiments on every iteration. We then checked if the meta-analytic effect size and heterogeneity remained constant or if it was sensitive against the influence of any individual experiment (Table S8).

We also performed trim-and-fill analyses and tested selection models to confirm the robustness of the results obtained from the frequentist meta-analyses against publication bias (Duval & Tweedie, 2000; Vevea & Hedges, 1995; Vevea & Woods, 2005; Supplementary Methods 2; Tables S9 and S10; Figure S3).

2.11 | Data and code availability

The data and code necessary to reproduce the analyses presented here are available at https://github.com/SkeideLab/meta_rotation.

3 | RESULTS

3.1 | Mental rotation performance

For our first meta-analytic model, effect sizes were quantified as the standardized mean difference in infants' looking times for novel and familiar rotated objects. Using this effect size index, we ran a Bayesian three-level random-effects model to test if there was evidence that infants did perform mental rotation. Across studies, infants looked longer at novel rotated objects than at familiar rotated objects, with a standardized mean difference of $g = 0.21$, 95% CrI [0.06, 0.37] (Figure 2). The probability for this effect being greater than zero was 99.6%. The heterogeneity of effect sizes was $\sigma^2_{\text{experiment}} = 0.15$, 95% CrI [0.07, 0.26], at the experiment level and $\sigma^2_{\text{article}} = 0.05$, 95% CrI [0.00, 0.16] at the article level. Therefore, 75.2% of the heterogeneity between effect sizes was attributable to differences between experiments *within* articles and 24.8% was attributable to differences *between* articles.

3.2 | Effects of gender, age, and task type

As a next step, we conducted a meta-regression analysis to test if the gender of the infants, their age, or the type of mental rotation task was related to mental rotation performance (Figure 3). Indeed, male groups of infants in the mental rotation studies revealed larger looking time differences than female groups, $b = 0.32$ (where $b = \Delta g$), 95% CrI [0.02, 0.62]. The probability for this effect being larger than zero was 98.2%. We found no difference between female groups and mixed-gender groups (i.e., experiments for which no separate male and female group statistics were available), $b = -0.02$, 95% CrI [−0.38, 0.36] (53.6% probability of an effect smaller than zero). Additionally, mean age was not reliably related to mental rotation performance, with a change per year of $b = 0.33$ per year (0.03 per month), 95% CrI [−0.24, 0.90]. We also did not detect an interaction between gender and age ([females – mixed] × age: $b = 0.15$, 95% CrI [−0.66, 0.96]; [males – females] × age: $b = 0.17$, 95% CrI [−0.66, 0.99]). Finally, there was some evidence that violation of expectation tasks yielded larger effects than habituation tasks, $b = 0.38$, 95% CrI [0.01, 0.73]. The probability for this effect being greater than zero was 97.8%. Note that this finding was not hypothesized a priori and might partly be driven by some of the habituation experiments obtaining familiarity preferences rather than novelty preferences (e.g., Moore & Johnson, 2011; see Discussion).

To confirm the gender difference between males and females, we set up another Bayesian meta-analysis, this time focusing on the looking-time contrasts between male and female infants reported *within* each experiment by the original authors. Our additional analysis revealed a meta-analytic effect size of $g = 0.14$, 95% CrI [−0.01, 0.30] (Figure S4) and a probability of this effect being greater than zero of 96.5%. The heterogeneity of effect sizes was $\sigma^2_{\text{experiment}} = 0.02$, 95% CrI [0.00, 0.10], at the experiment level and $\sigma^2_{\text{article}} = 0.04$, 95% CrI [0.00, 0.17] at the article level. Therefore, 43.3% of the heterogeneity between effect sizes was attributable to differences between experiments *within* articles and 56.7% was attributable to differences *between* articles.

The results obtained from our Bayesian analyses were reproduced by classical frequentist three-level meta-analysis and meta-regression models (Table S7).

3.3 | Publication bias assessment

We inspected funnel plots and performed Egger regression tests to examine the possibility of publication bias in the literature included here. For the meta-analysis of mental rotation performance (Figure 4a), we observed a slight asymmetry in the funnel plot, indicating a small publication bias. This was confirmed by an Egger regression test that showed a reliable association between effect sizes and their corresponding standard errors, $b = 1.93$, t (60) = 3.40, $p = 0.001$, 95% confidence interval (CI) [0.80, 3.07] (two-sided test). Nevertheless, a jackknife (leave-one-out) analysis confirmed that the current results are robust to the effects of individual outlier experiments (Table S7). For the meta-analysis of gender differences in mental rotation performance (Figure 4b), the asymmetry in the funnel plot was less pronounced and the slope of the Egger regression test was not statistically significant, $b = 0.32$, t (28) = 0.65, $p = 0.522$, 95% CI [−0.69, 1.32] (two-sided test). Selection models revealed that these results are only robust to small degrees of publication bias (Tables S9 and S10).

4 | DISCUSSION

We analyzed looking times during mental rotation in 1705 infants ranging from 3–16 months of age. To this end, we scrutinized the robustness of 62 experimental effect sizes. We found that on average, male infants looked slightly longer at novel rotated objects compared to female infants. This effect was small, only partially robust, and unrelated to age in the current range.

4.1 | Meta-analytic effect size

We interpret the meta-regression-based estimate of the gender difference ($b = 0.32$, where $b = \Delta g$) as an upper bound of the true effect size since this estimate is based on experiments that reported separate effect sizes for males and females. The gender effect of these

Article	Experiment	Effect size	2.5%	97.5%
Rochat & Hespos 1996	Experiment 1	0.43	0.06	0.81
Hespos & Rochat 1997	Experiment 2	0.80	0.33	1.28
	Experiment 3	0.58	0.09	1.06
	Experiment 4, 4−months−old infants	0.21	−0.45	0.88
	Experiment 4, 6−months−old infants	0.83	0.06	1.60
	Experiment 5, 6−months−old infants	0.82	0.09	1.54
	Experiment 6, 6−months−old infants	0.90	0.16	1.63
Moore & Johnson 2008	Females	−0.06	−0.51	0.39
	Males	0.60	0.13	1.07
Quinn & Liben 2008	Females	0.02	−0.58	0.61
	Males	2.01	1.13	2.89
Moore & Johnson 2011	Females	0.18	−0.28	0.63
	Males	−0.62	−1.10	−0.15
Frick & Möhring 2013	10−months−old infants	0.98	0.47	1.49
	8−months−old infants	0.09	−0.37	0.54
Möhring & Frick 2013	Exploration group	0.44	−0.03	0.90
	Observation group	0.03	−0.42	0.48
Schwarzer et al. 2013a	Crawling infants	0.49	0.06	0.91
	Non−crawling infants	−0.25	−0.66	0.16
Schwarzer et al. 2013b	Crawling infants	0.70	0.26	1.13
	Non−crawling infants	0.10	−0.31	0.51
Frick & Wang 2014	Experiment 1, 14−months−old infants	−0.47	−1.03	0.10
	Experiment 1, 16−months−old infants	0.73	0.14	1.31
	Experiment 2, 14−months−old infants	0.17	−0.37	0.72
	Experiment 2, 16−months−old infants	0.70	0.12	1.28
	Experiment 3, other−turning group	−0.13	−0.67	0.42
	Experiment 3, self−turning group	0.77	0.18	1.36
Quinn & Liben 2014	Experiment 2, 6−months−old females	−0.19	−0.79	0.41
	Experiment 2, 6−months−old males	1.28	0.56	2.00
	Experiment 2, 9−months−old females	−0.06	−0.66	0.53
	Experiment 2, 9−months−old males	0.99	0.32	1.66
Lauer et al. 2015	Females	0.58	0.19	0.98
	Males	1.13	0.70	1.57
Antrilli & Wang 2016	Vertical−stripes group	1.02	0.45	1.60
Christodoulou et al. 2016	Females	−0.26	−0.68	0.15
	Males	−0.80	−1.25	−0.36
Constantinescu et al. 2018	Females	−0.20	−0.58	0.18
	Males	0.84	0.41	1.26
Erdmann et al. 2018	5−months−old females	−0.17	−0.36	0.02
	5−months−old males	−0.19	−0.39	0.00
	9−months−old females	0.00	−0.22	0.21
	9−months−old males	0.02	−0.20	0.23
Gerhard & Schwarzer 2018	Crawling infants, 0° group	0.49	0.01	0.97
	Crawling infants, 54° group	−0.54	−1.01	−0.07
	Non−crawling infants, 0° group	−0.32	−0.80	0.16
	Non−crawling infants, 54° group	0.16	−0.30	0.62
Slone et al. 2018	Mittens−first group, females	0.11	−0.34	0.56
	Mittens−first group, males	0.04	−0.41	0.49
	Mittens−second group, females	−0.43	−0.89	0.04
	Mittens−second group, males	0.07	−0.38	0.52
Kaaz & Heil 2020	Experiment 1, females	−0.09	−0.25	0.08
	Experiment 1, males	−0.06	−0.23	0.10
	Experiment 2, females	−0.08	−0.37	0.20
	Experiment 2, males	0.35	0.06	0.64
Gerhard−Samunda et al. 2021	Horizontal−stripes group, crawling infants	−0.70	−1.37	−0.03
	Horizontal−stripes group, non−crawling infants	0.17	−0.49	0.83
	Vertical−stripes group, crawling infants	0.63	−0.02	1.29
	Vertical−stripes group, non−crawling infants	0.05	−0.57	0.67
Kelch et al. 2021	Experiment 1, crawling infants	−0.88	−1.57	−0.18
	Experiment 1, non−crawling infants	0.07	−0.50	0.64
	Experiment 2, crawling infants	−0.22	−0.80	0.35
	Experiment 2, non−crawling infants	0.22	−0.33	0.77
Three−level model		**0.21**	**0.06**	**0.37**

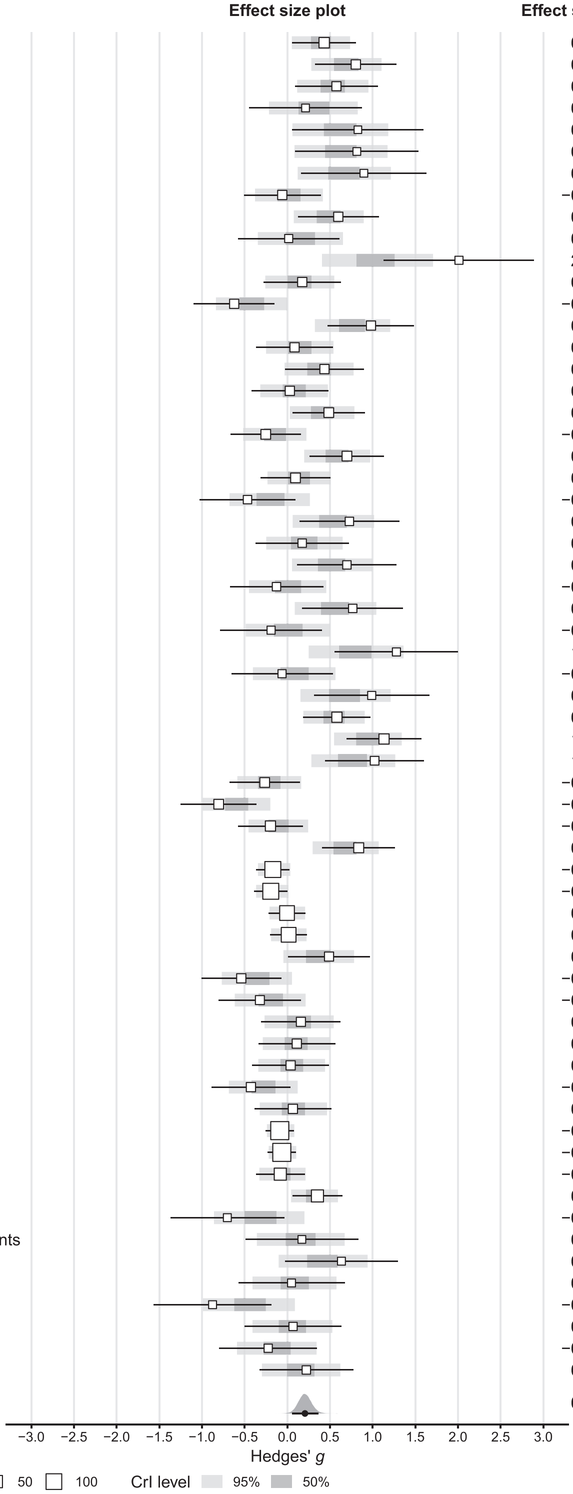

FIGURE 2 Mental rotation performance. A Bayesian three-level meta-analysis provided evidence for mental rotation ability in infants. White squares depict the effect sizes (Hedges' *g*) for infants' mental rotation performance in all individual experiments and black lines depict their 95% confidence interval. Gray bars indicate the 50% and 95% CrI from the Bayesian model, which takes into account that experiments with smaller sample sizes or more extreme effect sizes provide less reliable information. The last line shows the meta-analytic effect size (black dot) together with its 95% CrI (black line) and its posterior distribution (gray curve). CrI, credible interval.

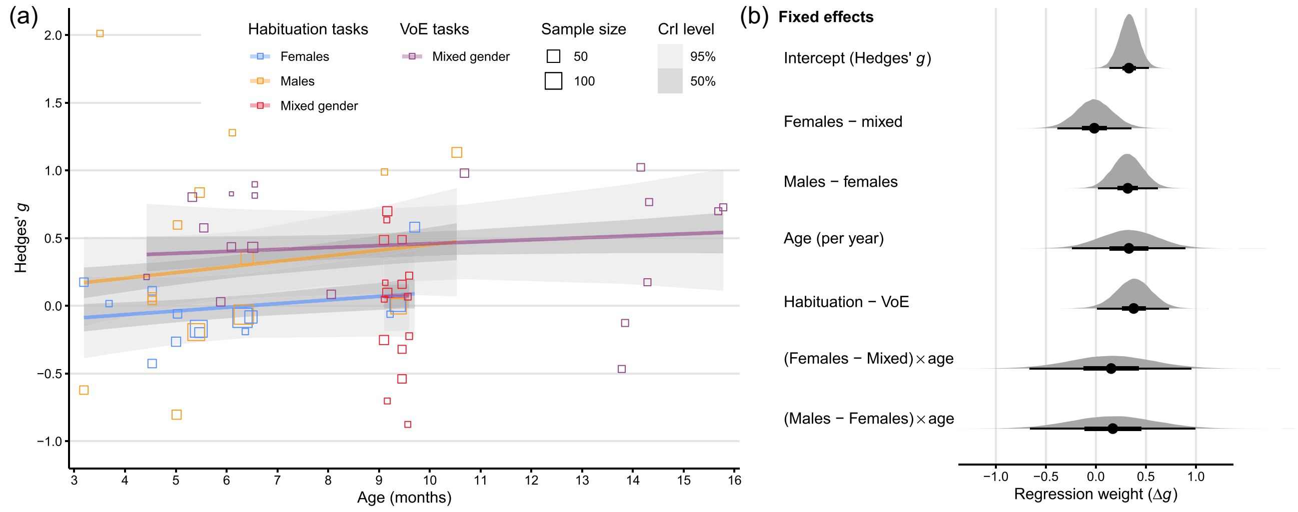

FIGURE 3 Effects of gender, age, and task type. (a) Squares show the effect size (Hedges' g) for infants' mental rotation performance in each of the 62 individual experiments. Squares are color-coded according to the type of mental rotation task and the gender of the infants (blue = habituation task, all-female sample; yellow = habituation task, all-male sample; red = habituation task, mixed-gender sample; purple = VoE task, mixed-gender sample). Lines indicate the best-fit regression estimates according to a Bayesian three-level meta-regression model and gray ribbons indicate their corresponding 50% and 95% CrI. (b) Fixed effect estimates obtained from the Bayesian three-level meta-regression model are depicted as black dots together with their 50% CrI (thick black lines) and 95% CrI (thin black lines). Gray curves indicate the posterior distribution for each effect. CrI, credible interval; VoE, violation of expectation.

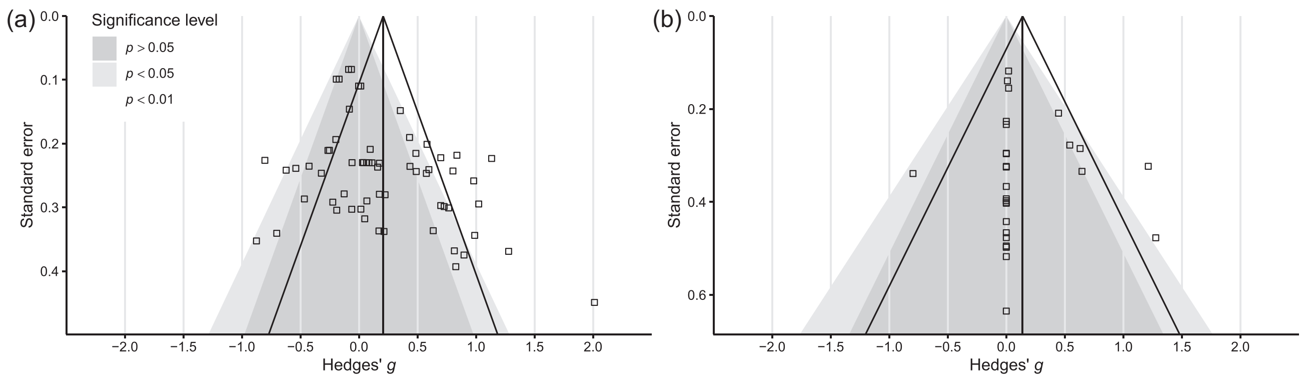

FIGURE 4 Evaluation of publication bias. Funnel plots for the meta-analysis of mental rotation performance (a) and for the meta-analysis of gender differences in mental rotation performance (b). The plots show the standard error and effect size for each of the individual experiments as a black square. The funnel contours (diagonal black lines) depict a 95% pseudo-confidence interval around the meta-analytic effect size (vertical black line). Gray shades indicate a 95% pseudo-confidence interval (dark gray) and a 99% pseudo-confidence interval (light gray) under the null hypothesis. These shades thus illustrate which of the original experiments observed a significant effect. For the meta-analysis of mental rotation performance, a slight asymmetry induced by the underrepresentation of experiments with high standard errors and small effect sizes suggests a small publication bias.

experiments can be considered as positively biased when assuming that authors who observe a statistically significant gender difference would be more likely to report separate effect sizes for males and females. In contrast, the meta-analytic estimate derived from the observed gender differences within each experiment ($g = 0.14$) can be viewed as a lower bound of the true effect size. This view is plausible because the unknown effect sizes of experiments without significant gender differences were set to zero although non-zero differences were likely also observed but just not reported because of missing power to render these effects statistically significant.

To facilitate the interpretability of our models, the effect size of the looking time difference between novel and familiar rotation events was coded in a linear fashion. Accordingly, a positive looking time difference (i.e., a novelty preference) was taken as stronger evidence for mental rotation than a looking time difference of zero (i.e., no systematic preference). A zero difference in turn was taken as stronger evidence for mental rotation than a negative looking time difference (i.e., a familiarity preference). This approach can be corroborated by the established view that novelty preference is generally considered as the paradigmatic mental rotation behavior (for related discussion, see Black & Bergmann, 2017; Cristia, 2018; Houston-Price & Nakai, 2004). Nevertheless, some authors have argued that familiarity preference should also be taken as evidence for mental rotation ability (Erdmann et al., 2018; Johnson & Moore, 2020; Moore & Johnson, 2020). Interpreting novelty preference *and* familiarity preference as successful mental rotation leads to an increase in the meta-analytic effect sizes for infants' mental rotation performance in general (Figure S5) as well as for the gender difference in mental rotation performance (Figure S6).

4.2 | Influences on mental rotation performance

The gender differences observed in this study remain to be explained by interacting genetic and environmental factors that are largely unknown. To our knowledge, there are currently no genetic association or gene-environment interaction studies with a focus on mental rotation. Nevertheless, it is documented that genetic contributions to behavioral variance in mental rotation are substantially smaller than unique non-shared environmental contributions both in male and female adults (Shakeshaft et al., 2016). Whether this observation also applies to infants remains to be explored.

One recent study on 5–6-month-old female infants provided preliminary evidence for possible social-environmental effects related to parental attitudes towards gender which might partly explain the results of our present work (Constantinescu et al., 2018). As far as we know, potentially mediating and moderating factors that could already be operational in infancy, however, are not yet empirically established. In a similar vein, while mental rotation training has small-to-medium post-test effects in children, it is unclear whether it can remove gender differences and be adapted to infants (Uttal et al., 2013).

Sex hormone concentration in male infants, especially postnatal testosterone in the first 6 months of life, could also contribute to gender differences in mental rotation performance (Constantinescu et al., 2018; Erdmann et al., 2019; Toivainen et al., 2018). However, possible biological developmental pathways, bridging the gap from hormonal to behavioral differences, are currently far from understood.

A number of additional factors have been associated with individual differences in infant mental rotation performance. For example, mental rotation is related to previous relevant experience with the particular objects used in the specific task (Möhring & Frick, 2013; Schwarzer, Freitag, Buckel, et al., 2013; Slone et al., 2018). This relation also applies to previous experience with manually rotating toys (Schwarzer, Freitag, & Schum, 2013). While these preliminary results require replication, they are in line with the longstanding notion that prior knowledge is the strongest predictor of learning outcomes in a range of cognitive domains (e.g., Ausubel, 1968; Bradley & Bryant, 1983; Halberda et al., 2008). Furthermore, there is yet to be confirmed preliminary evidence for possible links between mental rotation performance and several sensory-motor skills including fine and gross motor skills, oculomotor control, and crawling skills (Schwarzer, Freitag, Buckel, et al., 2013; Schwarzer, Freitag, & Schum, 2013).

4.3 | Future research directions

More research is needed to explore and confirm the underlying genetic, hormonal, and environmental factors (as well as their interactions) that contribute to individual differences in mental rotation performance, including the small gender difference observed in the present meta-analysis. We were not able to meta-analyze these factors given that each of them had only been investigated in one or two individual studies at most. Larger sample sizes, independent replication, and, ultimately, future meta-analyses will help to provide a clearer picture for why these differences emerge.

Furthermore, our meta-analysis could only cover the age range of 3–16 months of age although we had initially planned to include studies with a mean age of up to 3 years of age (based on the fact that studies with older children have already been covered by Lauer et al., 2019). However, we identified only one article that tested infants between 1.5 and 3 years of age (Pedrett et al., 2020; mean age = 30.7 months). Future studies and meta-analyses should therefore aim to close this gap by focusing explicitly on mental rotation performance and gender differences in the second half of the second year as well as the third year of life.

Another relevant perspective for future work is to account for stimulus characteristics like stimulus dimensionality (two-dimensional vs. three-dimensional objects) and stimulus type (real or animate objects vs. geometric figures). It is plausible to assume that gender differences could be smaller in tasks using two-dimensional objects and real or animate figures. However, compared to what is known in the adult literature (e.g., Voyer et al., 1995), infant researchers have not yet systematically probed the effect of different tasks or stimuli on mental rotation performance and on the size of gender differences. In this context, it is important to note that the present study differs from previous meta-analyses in adults in that it pools across different types and implementations of mental rotation tasks, some of which have only been used in one or a few articles. While this heterogeneity in terms of tasks and stimuli could speak to the generalizability of our meta-analytic findings across different tasks, future meta-analyses based on a larger number of original studies could apply stricter inclusion criteria or model different types of tasks and stimuli in a more fine-grained fashion.

5 | CONCLUSION

The present study revealed that male infants look slightly longer at novel rotated objects than female infants. Accordingly, on average, males show slightly more reliable mental rotation behavior already in the first months of postnatal life. However, this difference is small and only partially robust to publication bias.

ACKNOWLEDGMENTS

We would like to thank Stella F. Lourenco and David S. Moore for providing unpublished effect size data as well as Scott P. Johnson and three anonymous reviewers for comments on earlier versions of the paper. Michael A. Skeide was supported by the German Research Foundation (DFG Grants 433715509 and 433758790), the European Research Council (ERC Starting Grant 850752), and a Jacobs Foundation Research Scholarship. Alexander Enge was supported by a PhD scholarship from the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes).

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data and code necessary to reproduce the analyses presented here are available at https://github.com/SkeideLab/meta_rotation.

ORCID

Alexander Enge https://orcid.org/0000-0003-0100-2297

REFERENCES

*Antrilli, N. K., & Wang, S. (2016). Visual cues generated during action facilitate 14-month-old infants' mental rotation. *Journal of Cognition and Development*, *17*(3), 418–429. https://doi.org/10.1080/15248372.2015.1058262

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208. https://doi.org/10.1016/0010-0277(85)90008-3

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, (pp. 124–129). Cognitive Science Society.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read a causal connection. *Nature*, *301*(5899), 419–421. https://doi.org/10.1038/301419a0

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Casey, B. M., Pezaris, E., Fineman, B., Pollock, A., Demers, L., & Dearing, E. (2015). A longitudinal analysis of early spatial skills compared to arithmetic and verbal skills as predictors of fifth-grade girls' math reasoning. *Learning and Individual Differences*, *40*, 90–100. https://doi.org/10.1016/j.lindif.2015.03.028

Cheung, C. N., Sung, J. Y., & Lourenco, S. F. (2020). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research*, *84*(7), 2000–2017. https://doi.org/10.1007/s00426-019-01202-5

*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*(Part A), 64–70. https://doi.org/10.1016/j.infbeh.2016.09.005

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

*Constantinescu, M., Moore, D. S., Johnson, S. P., & Hines, M. (2018). Early contributions to infants' mental rotation abilities. *Developmental Science*, *21*(4), e12613. https://doi.org/10.1111/desc.12613

Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, *170*, 312–327. https://doi.org/10.1016/j.cognition.2017.09.016

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Efron, B., & Tibshirani, R. (1993). *An Introduction to the bootstrap*. Chapman & Hall.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

*Erdmann, K., Kavšek, M., & Heil, M. (2018). Infants' looking times in a dynamic mental rotation task: Clarifying inconsistent results. *Cognitive Development*, *48*, 279–285. https://doi.org/10.1016/j.cogdev.2018.10.002

Erdmann, K., Schaal, N. K., Meinlschmidt, G., Tegethoff, M., Fröhlich, S., Kozlowski, P., Rivet, N., Jamey, C., Reix, N., Kintz, P., Raul, J. S., & Heil, M. (2019). Sex specific relationships between infants' mental rotation ability and amiotic sex hormones. *Neuroscience Letters*, *707*, 134298. https://doi.org/10.1016/j.neulet.2019.134298

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670. https://doi.org/10.1126/science.146.3644.668

*Frick, A., & Möhring, W. (2013). Mental object rotation and motor development in 8- and 10-month-old infants. *Journal of Experimental Child Psychology*, *115*(4), 708–720. https://doi.org/10.1016/j.jecp.2013.04.001

Frick, A., Möhring, W., & Newcombe, N. S. (2014). Development of mental transformation abilities. *Trends in Cognitive Sciences*, *18*(10), 536–542. https://doi.org/10.1016/j.tics.2014.05.011

*Frick, A., & Wang, S. H. (2014). Mental spatial transformations in 14- and 16-month-old infants: Effects of action and observational experience. *Child Development*, *85*(1), 278–293. https://doi.org/10.1111/cdev.12116

Frick, A. (2019). Spatial transformation abilities and their relation to later mathematics performance. *Psychological Research*, *83*(7), 1465–1484. https://doi.org/10.1007/s00426-018-1008-5

Geer, E. A., Quinn, J. M., & Ganley, C. M. (2019). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology*, *55*(3), 637–652. https://doi.org/10.1037/dev0000649

*Gerhard, T. M., & Schwarzer, G. (2018). Impact of rotation angle on crawling and non-crawling 9-month-old infants' mental rotation ability. *Journal of Experimental Child Psychology*, *170*, 45–56. https://doi.org/10.1016/j.jecp.2018.01.001

*Gerhard-Samunda, T. M., Jovanovic, B., & Schwarzer, G. (2021). Role of manually-generated visual cues in crawling and non-crawling 9-month-old infants' mental rotation. *Cognitive Development*, *59*, 101053. https://doi.org/10.1016/j.cogdev.2021.101053

Gilligan, K. A., Flouri, E., & Farran, E. K. (2017). The contribution of spatial ability to mathematics achievement in middle childhood. *Journal of Experimental Child Psychology*, *163*, 107–125. https://doi.org/10.1016/j.jecp.2017.04.016

Goulet-Pelletier, J. C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, part I: The Cohen's *d* family. *The Quantitative Methods for Psychology*, *14*(4), 242–265. https://doi.org/10.20982/tqmp.14.4.p242

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, *48*(5), 1229–1241. https://doi.org/10.1037/a0027433

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. https://doi.org/10.1038/nature07246

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Chapman & Hall/CRC Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689. https://doi.org/10.1037/0022-0663.91.4.684

*Hespos, S. J., & Rochat, P. (1997). Dynamic mental representation in infancy. *Cognition*, *64*(2), 153–188. https://doi.org/10.1016/s0010-0277(97)00029-2

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*(4), 341–348. https://doi.org/10.1002/icd.364

Johnson, S. P., & Moore, D. S. (2020). Spatial thinking in infancy: Origins and development of mental rotation between 3 and 10 months of age. *Cognitive Research: Principles and Implications*, *5*(1), 10. https://doi.org/10.1186/s41235-020-00212-x

Johnson, W., & Bouchard, T. J. Jr. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393–416. https://doi.org/10.1016/j.intell.2004.12.002

*Kaaz, T., & Heil, M. (2020). Infants' looking times in a 2-D mental rotation task. *Infant and Child Development*, *29*(2), e2167. https://doi.org/10.1002/icd.2167

*Kelch, A., Schwarzer, G., Gehb, G., & Jovanovic, B. (2021). How 9-month-old crawling infants profit from visual-manual rotations in a mental rotation task. *Infant Behavior & Development*, *65*, 101642. https://doi.org/10.1016/j.infbeh.2021.101642

Kossmeier, M., Tran, U. S., & Voracek, M. (2020). metaviz: Forrest plots, funnel plots, and visual funnel plot inference for meta-analysis. https://github.com/Mkossmeier/metaviz

Kruschke, J. K. (2015). Bayesian approaches to testing a point ("null") hypothesis. *Doing Bayesian data analysis* (2nd ed., pp. 335–358). Academic Press.

Kubicek, C., & Schwarzer, G. (2018). On the relation between infants' spatial object processing and their motor skills. *Journal of Motor Learning and Development*, *6*, S6–S23. https://doi.org/10.1123/jmld.2016-0062

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

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. https://doi.org/10.1177/0956797616655977

*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*, 558. https://doi.org/10.3389/fpsyg.2015.00558

Lauer, J. E., Yhang, E., & Lourenco, S. F. (2019). The development of gender differences in spatial reasoning: A meta-analytic review. *Psychological Bulletin*, *145*(6), 537–565. https://doi.org/10.1037/bul0000191

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. https://doi.org/10.2307/1130467

Maeda, Y., & Yoon, S. Y. (2013). A meta-analysis on gender differences in mental rotation ability measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R). *Educational Psychology Review*, *25*(1), 69–94. https://doi.org/10.1007/s10648-012-9215-x

Mash, C., Arterberry, M. E., & Bornstein, M. H. (2007). Mechanisms of visual object recognition in infancy: Five-month-olds generalize beyond the interpolation of familiar views. *Infancy*, *12*(1), 31–43. https://doi.org/10.1111/j.1532-7078.2007.tb00232.x

*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. https://doi.org/10.1111/cdev.12065

*Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*(11), 1063–1066. https://doi.org/10.1111/j.1467-9280.2008.02200.x

*Moore, D. S., & Johnson, S. P. (2011). Mental rotation of dynamic, three-dimensional stimuli by 3-month-old infants. *Infancy*, *16*(4), 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. https://doi.org/10.1016/bs.acdb.2020.01.001

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. https://doi.org/10.1037/1082-989X.7.1.105

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pedrett, S., Kaspar, L., & Frick, A. (2020). Understanding of object rotation between two and three years of age. *Developmental Psychology*, *56*(2), 261–274. https://doi.org/10.1037/dev0000871

*Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*(11), 1067–1070. https://doi.org/10.1111/j.1467-9280.2008.02201.x

*Quinn, P. C., & Liben, L. S. (2014). A sex difference in mental rotation in infants: Convergent evidence. *Infancy*, *19*(1), 103–116. https://doi.org/10.1111/infa.12033

R Core Team. (2021). R: A Language and Environment for Statistical Computing.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, *25*(15), 1923–1929. https://doi.org/10.1093/bioinformatics/btp358

*Rochat, P., & Hespos, S. J. (1996). Tracking and anticipation of invisible spatial transformations by 4- to 8-month-old infants. *Cognitive Development*, *11*(1), 3–17. https://doi.org/10.1016/S0885-2014(96)90025-8

Rosenthal, R. (1991). *Meta-Analytic procedures for social research*. SAGE Publications.

Rusiak, P., Lachmann, T., Jaskowski, P., & van Leeuwen, C. (2007). Mental rotation of letters and shapes in developmental dyslexia. *Perception*, *36*(4), 617–631. https://doi.org/10.1068/p5644

Rüsseler, J., Scholz, J., Jordan, K., & Quaiser-Pohl, C. (2005). Mental rotation of letters, pictures, and three-dimensional objects in German dyslexic children. *Child Neuropsychology*, *11*(6), 497–512. https://doi.org/10.1080/09297040490920168

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

*Schwarzer, G., Freitag, C., Buckel, R., & Lofruthe, A. (2013). Crawling is associated with mental rotation ability by 9-month-old infants. *Infancy*, *18*(3), 432–441. https://doi.org/10.1111/j.1532-7078.2012.00132.x

*Schwarzer, G., Freitag, C., & Schum, N. (2013). How crawling and manual object exploration are related to the mental rotation abilities of 9-month-old infants. *Frontiers in Psychology*, *4*, 97. https://doi.org/10.3389/fpsyg.2013.00097

Shakeshaft, N. G., Rimfeld, K., Schofield, K. L., Selzam, S., Malanchini, M., Rodic, M., Kovas, Y., & Plomin, R. (2016). Rotation is visualisation, 3D is 2D: Using a novel measure to investigate the genetics of spatial ability. *Scientific Reports*, *6*(1), 30545. https://doi.org/10.1038/srep30545

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

*Slone, L. K., Moore, D. S., & Johnson, S. P. (2018). Object exploration facilitates 4-month-olds' mental rotation performance. *Plos One*, *13*(8), e0200468. https://doi.org/10.1371/journal.pone.0200468

Stan Development Team. (2022). Stan Modeling Language Users Guide and Reference Manual.

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 99–110). John Wiley & Sons.

Stinton, C., Farran, E. K., & Courbois, Y. (2008). Mental rotation in Williams syndrome: An impaired ability. *Developmental Neuropsychology*, *33*(5), 565–583. https://doi.org/10.1080/87565640802254323

Toivainen, T., Pannini, G., Papageorgiou, K. A., Malanchini, M., Rimfeld, K., Shakeshaft, N., & Kovas, Y. (2018). Prenatal testosterone does not explain sex differences in spatial ability. *Scientific Reports*, *8*(1), 13653. https://doi.org/10.1038/s41598-018-31704-y

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

van Tetering, M., van der Donk, M., de Groot, R. H. M., & Jolles, J. (2019). Sex differences in the performance of 7–12 year olds on a mental rotation task and the relation with arithmetic performance. *Frontiers in Psychology*, *10*, 107. https://doi.org/10.3389/fpsyg.2019.00107

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2021). *Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC* [Preprint]. https://doi.org/10.48550/arXiv.1903.08008

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. https://doi.org/10.1007/BF02294384

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin and Review*, *18*(2), 267–277. https://doi.org/10.3758/s13423-010-0042-0

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. https://doi.org/10.1037/0033-2909.117.2.250

Williams, D. R., Rast, P., & Bürkner, P. C. (2018). *Bayesian meta-analysis with weakly informative prior distributions* [Preprint]. https://doi.org/10.31234/osf.io/7tbrm

Yang, W., Liu, H., Chen, N., Xu, P., & Lin, X. (2020). Is early spatial skills training effective? A meta-analysis. *Frontiers in Psychology*, *11*, 1938. https://doi.org/10.3389/fpsyg.2020.01938

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Enge, A., Kapoor, S., Kieslinger, A.-S., & Skeide, M. A. (2023). A meta-analysis of mental rotation in the first years of life. *Developmental Science*, e13381. https://doi.org/10.1111/desc.13381