Development of Mental Rotation in 3- to 5-Year-Old Children

Andrea Frick[a][1], Melissa A. Hansen[b], and Nora S. Newcombe[c]

Temple University

[a]Department of Psychology, Temple University, Weiss Hall 318, 1701 North 13th Street, Philadelphia, PA 19122-6085, United States. E-mail: frick@psy.unibe.ch

[b]Department of Psychology, Temple University, Weiss Hall 318, 1701 North 13th Street, Philadelphia, PA 19122-6085, United States. Email: melissaannehansen@gmail.com

[c]Department of Psychology, Temple University, Weiss Hall 318, 1701 North 13th Street, Philadelphia, PA 19122-6085, United States. Email: Newcombe@temple.edu

[1]Andrea Frick is now at the Department of Psychology, University of Bern, Muesmattstr. 45, CH-3000 Bern 9, Switzerland.

Correspondence concerning this article should be addressed to Andrea Frick, Department of Psychology, University of Bern, Muesmattstrasse 45, CH-3000 Bern 9, Switzerland. E-mail: frick@psy.unibe.ch

Abstract

We assessed 3- to 5-year-olds' mental rotation abilities using a new puzzle paradigm. Children saw pairs of asymmetrical ghost figures in seven orientations. One of the ghosts would fit into a hole if rotated right-side up – the other ghost was its mirror image and would not fit. Children were asked to turn the ghosts in their heads and pick the one that would fit into the hole. The number of children who picked the correct ghost above chance increased dramatically from 10% of 3-year-olds to 95% of 5-year-olds; the average accuracy also increased significantly, from 54% to 83%. These results indicate considerable development in mental rotation between 3 and 5 years. A paper version of the task yielded similar results. This paradigm allows for assessing mental rotation abilities in children younger than 5 years, using a task comparable to tests and paradigms used with older children and adults.

Keywords: mental rotation, mental imagery, spatial cognition, children, development

Development of Mental Rotation in 3- to 5-Year-Old Children

Mental rotation is the ability to imagine how an object would look in a different orientation – in other words, to turn something in one's mind. Mental rotation tests often are used as index measures for spatial visualization abilities and mental imagery processes in general. Factor analytic research has shown that visualization is a well-defined component skill within general intelligence in adults (Carroll, 1993), and spatial visualization abilities, as measured by *Mental Rotation* and *Surface Development* tasks, have been shown to play an important role in achieving advanced degrees in academic disciplines such as science, technology, engineering, and mathematics (Wai, Lubinski, & Benbow, 2009). A number of previous studies have suggested that children can perform mental rotations by the age of 5 years, although at a slower speed than adults (e.g., Frick, Daum, Walser & Mast, 2009; Funk, Brugger, & Wilkening, 2005; Kosslyn, Margolis, Barrett, Goldknopf, & Daly, 1990; Marmor, 1975).

Research on the development of mental rotation in children younger than 5 years is challenging, however, because classic mental rotation paradigms used with adults and older children present high cognitive demands that may overburden young children's capacities. For example, in work by Shepard and Metzler (1971) and Marmor (1975), participants saw two objects that were either exactly the same or mirror images of each other – one oriented upright and one rotated. The participants' task was to pull one of two response levers to indicate whether they thought the objects were the same or different. Thus, to succeed in this task, children have to understand what constitutes a "same" or "different" object, remember which lever stands for which response, generate a mental image of the object, and maintain this image while performing a mental transformation on it. In fact, even for discriminating non-rotated mirror images, there are significant demands made on kindergartners by a same-different task (Cronin, 1967).

Given the cognitive complexity of same-different judgments, it is not surprising how rarely a mental rotation paradigm using this response mode has been given successfully to children younger than 5 years, and how inconsistent the findings have been. Marmor (1977) presented two pictures that differed in angular orientation and asked 4- and 5-year-olds to press a lever on the left when they saw matching-image pairs of stimuli (e.g., bears) or a lever on the right when they saw mirror-image pairs. Marmor argued that her data showed that 4-year-olds were already able to perform mental rotation. However, a follow-up study (Dean & Harvey, 1979) that employed the same procedure with slightly different stimuli failed to replicate Marmor's results and showed that 4- to 6-year-olds performed at chance levels. Marmor also found that training children to use a mental rotation strategy did not have a significant effect, suggesting that they already showed robust skill. However, a later replication study (Platt & Cohen, 1981) showed that twice as many 5-year-olds produced response patterns indicative of mental rotation with training compared to without training. Indeed, even though widely accepted at the time, there has always been some controversy about the conclusions from Marmor's studies (Newcombe, 2002). For instance, research that focused on individual children's response patterns revealed that only a small proportion of 4-year-olds showed a pattern consistent with mental rotation (Estes, 1998). Taken together, these mental rotation studies have yielded inconsistent results that cast some doubt upon the idea of robust mental rotation in preschool children. Moreover, some of these studies suggest that many children still perform poorly on mental rotation tasks at 4 to 5 years of age, and indicate that there are important individual differences in mental rotation abilities at this age.

On the other hand, mental rotation has recently been studied in infants and toddlers using paradigms not involving explicit judgment. Using looking-time paradigms, infants seem able to distinguish between objects and their rotated mirror objects (Frick & Möhring, 2013; Möhring & Frick, 2013; Moore & Johnson, 2008, 2011; Quinn & Liben, 2008;

Schwarzer, Freitag, Buckel, & Lofruthe, 2012) or between probable and physically

improbable rotation events (Frick & Wang, 2013; Hespos & Rochat, 1997; Rochat & Hespos,

1996). Moreover, in a task using physically-available objects and a clear goal, 22-month-olds

can rotate objects and successfully fit them through holes (Örnkloo & von Hofsten, 2007),

although this ability improves considerably across the age range from 15 to 30 months

(Shutts, Örnkloo, von Hofsten, Keen, & Spelke, 2009). These indications of early

understanding of rotation events bring into question whether the reports of relatively poor

performance among preschool children discussed above reflect their actual mental rotation

competence or are merely a result of high task demands.

Previous mental rotation studies have also yielded inconsistent results with regards to

the question of sex differences. Sex differences are frequently reported in studies on mental

rotation in adults, as shown by two meta-analyses (Linn & Petersen, 1985; Voyer, Voyer, &

Bryden, 1995). However, Linn and Petersen's meta-analysis did not include children younger

than 10 years old. The more recent meta-analysis by Voyer et al. listed four studies of mental

rotation with children below the age of 10, three of which found no significant sex effects

(Caldwell & Hall, 1970; Jahoda, 1979; Kaess, 1971). Interestingly, Voyer and colleagues

found a positive relation between chronological age and effect size, suggesting that sex

differences increase with age. If this is the case, an important question is when sex

differences begin to appear. Levine, Huttenlocher, Taylor, and Langrock (1999) found sex

differences in children older than 4.5 years, but not in 4- to 4.5-year-olds, on a mental

transformation task that included items requiring some rotation. In contrast, other studies with

children aged 4 years and older did not find sex differences (Estes, 1998; Frick et al., 2009;

Kosslyn et al., 1990; Platt & Cohen, 1981), or found higher error rates in boys (Krüger &

Krist, 2009). The infant studies have also yielded heterogeneous results. The question of

whether sex differences in mental rotation – or other individual differences – exist in early

preschool years or develop during childhood is relevant in light of findings that proficiency in such spatial skills is correlated with later career choices and academic success (Wai et al., 2009). By creating tasks and measures to assess early mental rotation ability, we can gain a better understanding of the origins of individual differences and the developmental trajectories of this skill, which is instrumental for designing and evaluating early interventions and providing equal opportunities for children to develop their spatial skills.

In the present study, 3- to 5-year-olds' mental rotation abilities were investigated using a paradigm that minimized task demands while still presenting a cognitive task that was comparable to the ones typically used with older children and adults. Similar to those tasks, the present task required a forced choice between two rotated stimuli that were mirror images of each other and whose shapes varied from trial to trial. The age range for this study covered the age over which success in mental rotation tasks has been (inconsistently) reported in previous literature (age 4 and 5); additionally, we tested 3-year-olds, to further specify development during early childhood.

The task was presented in the form of a "puzzle game", in which children saw pairs of asymmetrical puzzle pieces in seven different orientations. One of the pieces would fit into a cut-out on a board if rotated right-side up – the other piece was its mirror version and would not fit (see Figure 1). Children were asked to turn the pieces in their heads and to pick the one that would fit into the hole. To minimize strategies that concentrated on distinctive stimulus features, the following measures were taken. First, and similar to classic paradigms (cf. Shepard & Metzler, 1971), we presented stimulus pairs consisting of asymmetrical (chiral) mirror images; however, to make the task more appealing for young children, the stimuli were drawings of ghosts, rather than abstract cube-shaped objects. Second, in contrast to the majority of previous studies with children, but in line with Shepard and Metzler's original

design, the shapes of the stimuli varied from trial to trial. Third, the reference shapes (holes) only showed the outline of one of the ghosts, but did not have any other features (e.g., eyes).

Children's performance was assessed on the basis of the number of correct choices. To lower task demands, children were allowed to pick the ghosts directly, rather than having to press a response button to indicate their choices. Thus, in terms of the dependent variable, the present paradigm differs from chronometric studies and is more comparable to paper-and-pencil assessments (e.g, *Mental Rotation Test,* Vandenberg & Kuse, 1978) in which adults' mental rotation abilities are assessed on the basis of multiple-choice responses. The youngest age group in which an adapted paper-pencil test has been used previously was 5 years (*Picture Rotation Test*, Quaiser-Pohl, 2003). In the present paradigm, only two choice alternatives were presented (as opposed to four in the *Mental Rotation Test* and three in the *Picture Rotation Test*), to lower task complexity.

In Experiment 1, the stimuli were presented as three-dimensional cut-outs, and children were allowed to manually pick up a ghost and place it into the hole. Thus children received feedback about the correctness of their choices from whether or not the piece fit the hole. In Experiment 2, the same ghost-pairs and outlines were presented on paper. In this two-dimensional version of the task, children were instructed to point to the correct ghost and did not receive feedback about the correctness of their choices. This manipulation allowed us to explore the role of feedback in this mental rotation task, and also to obtain information on whether a more portable and easy-to-administer paper version of the task would yield comparable results.

1. Experiment 1

1.1 Method

1.1.1 Participants

Sixty children participated, with 20 children in each of three age groups: 3-year-olds (mean age = 42 months, range = 41 – 44 months), 4-year-olds (mean age = 54 months, range = 52– 56 months), and 5-year-olds (mean age = 66 months, range = 65 – 68 months). There were 10 boys and 10 girls in each age group, and the mean age difference between boys and girls was no more than 6 days in any age group. Four additional children were tested but excluded from analyses: two 3-year-olds perseverated and chose the ghost in the same location on all trials, one 3-year-old did not want to finish the experiment, and one 3-year-old had to be excluded due to experimenter error. The sample was predominantly Caucasian, middle class, and was recruited in suburban areas of a large US city. All children spoke English and were tested in English.

1.1.2 Stimulus material

The stimuli consisted of three boards made out of foam material. One board was green and used in the instruction trials; two boards were black and used in the test trials. Asymmetrical pieces in the shapes of a key, a hammer, a sock, a sail boat, and an ice skate were cut out of the green foam board; asymmetrical pieces in the shape of various ghosts were cut out of the black foam boards (see Figure 1). One black foam board, subsequently referred to as Board A, had holes for 10 pieces; Board B had holes for 11 pieces. Two small boards were used rather than one large to avoid having to present the pieces too far away from the holes.

For each piece, a mirror-image duplicate was created. White paper with the outlines of the artifacts and ghosts were glued onto all pieces and into the holes, to make their shape more salient. The pieces also showed some details, such as the eyes of the ghosts, to make the stimuli more entertaining for the children; however, these details were not present in the holes. The matching pieces and their mirror-image duplicates were presented in pairs on white cardboard trays (ca. 7 by 13 cm) placed alongside the boards. Placing both pieces on

one tray ensured that they were presented at exactly the same time and in a specific orientation.

1.1.3 Procedure

Participants were tested in a laboratory room, seated at a table opposite from the experimenter. Children first received six instruction trials with five artifacts. On the first trial, two pieces (keys) were presented upright, and the children were told that these pieces looked similar, but only one of them would fit into the hole on the board. The children were asked to pick the one they thought would fit into the hole. The children could then try to put the piece into the hole. If they had picked the wrong piece and failed to place it into the hole the experimenter encouraged them to try the other piece. Next, the key piece was removed from the hole and presented again, but this time the two key pieces were first rotated 180° in front of the children. Children were told that the pieces were now upside down, so they would have to turn them in their heads to figure out which one would fit into the hole. Four more warm-up trials followed, in which pairs of artifacts were presented at 0°, 45°, 135° and 180° angular disparities to their respective holes, but they were not rotated in front of the children anymore. To avoid trial-and-error strategies, children were instructed to "take a good look first and only pick the one that fits" and to "not even touch the wrong one!" For all except for the 0° trials, the children were reminded to turn the pieces in their heads before they picked one. If a child tried to flip a piece, the experimenter explained that flipping was not allowed in this game and demonstrated that the back side of the piece was not the same color.

After the instruction trials, children were told that they would next play the same game with ghosts. On each test trial, the experimenter placed the tray with the two ghost pieces next to the board, so that the pieces were equidistant to the hole. Simultaneously, the experimenter pointed to the corresponding hole, and left her index finger next to the hole until the children filled it. If children chose the wrong piece, they were subsequently allowed

to fill the hole with the correct piece. Thus, children received feedback about the correctness of their choices through the outcomes of their attempts to fit the pieces into the holes. Additionally, the experimenter would praise them for picking the correct ghost or comment that they had chosen the wrong one.

1.1.3 Design

The ghost pairs were presented in seven different disparities, which differed from their respective holes by 0° to 180°, in steps of 30°. To minimize effects of specific ghost stimuli, each disparity was presented three times using different kinds of ghost pairs, which resulted in 21 trials. Children were randomly assigned to one of four presentation orders, with the restriction of having an equal number of children per order and age group, and of having a roughly equal number of boys and girls per order. Each one of the four quasi-random orders consisted of three blocks, in which all of the seven disparities were presented once, and no disparity was ever presented twice in a row. Two orders presented the ghosts from Board A first, whereas the other two started with Board B. For one of each order (AB and BA), the trays were presented rotated by 180°. For example, in the trial shown in Figure 1 the correct ghost was in the top placement and had to be rotated 120° clockwise to fit into the hole. For half of the participants the tray was presented the other way round, so that the correct ghost was now in the bottom placement and had to be rotated 60° counterclockwise to fit into the hole (i.e., this was the shortest angular distance to the hole). The familiarization trials also were presented in four different quasi-random orders, with trays turned around for half of the participants, except for the keys, which were always presented first, once at 0° and then again at 180° angular distance to the hole.

1.2 Results

Children's responses were videotaped, and coded after the experiment. Responses were coded as incorrect if the incorrect ghost was picked up, moved towards the board to

compare with the hole, or rotated on the tray. Responses were coded as correct if children directly picked the correct ghost, if they touched the incorrect ghost without moving it and then selected the correct ghost, or if they lifted the incorrect ghost no more than one inch without rotating or translating it, but then set it back down to pick the correct ghost. This coding procedure allowed for self-correction (as long as the piece was not rotated or moved closer to compare with the hole), assuming that aborting an initiated action may indicate that children anticipated the result of their action and, thus, may be a sign for successful mental rotation. A second experimenter recoded 18 data sets (6 of each age group); scores of the two experimenters agreed on 99% of the trials.

1.2.1 Mean Accuracy

To investigate whether accuracy decreased with increasing angle of rotation, data for trials with the same angular disparities were collapsed. An ANOVA was calculated with disparity (7), age group (3), sex (2), and order (4) as independent variables and the proportion of correctly solved trials as dependent variable. The analysis showed no significant main effects of or interactions with order (all Fs < 1.48, all ps $> .10$); therefore this variable was not included in subsequent analyses.

A repeated measures ANOVA with disparity (7) as the within-subject variable, age group (3) and sex (2) as the between-subjects variables, and proportion of correctly solved trials as the dependent variable yielded a significant main effect of age group, $F(2, 54) = 22.69$, $p < .001$, $\eta^2 = .46$ (see Figure 2). Post hoc comparisons (Tukey HSD) showed that all pairs of age groups differed significantly (all ps $< .01$). Table 1 shows the number (and percentage) of correct choices per age group. Furthermore, as would be expected if mental rotation were being tapped, the analysis yielded a significant main effect of disparity, $F(6, 324) = 8.28$, $p < .001$, $\eta^2 = .13$, with a significant linear component, lin. $F(1, 54) = 35.84$, $p < .001$, $\eta^2 = .40$. There was a tendency to an interaction between disparity and age group, $F(12,

324) = 1.68, p = .07, η^2 = .06, and a significant three-way interaction involving disparity, age group, and sex, $F(12, 324)$ = 1.99, p = .02, η^2 = .07. No other significant effects were found (all Fs < 1.82, all ps > .09).

To further investigate the three-way interaction, simple main effects analyses for each age group were conducted. There was no significant interaction between disparity and sex in 4-year-olds, F < 1, or 5-year-olds, $F(6, 324)$ = 1.89, p = .08. However, 3-year-olds showed a significant interaction between disparity and sex, $F(6, 324)$ = 3.01, p < .01. Post hoc comparisons (Sidak corrected) indicated that for 135° angular disparity, boys showed 28% lower accuracy than girls (p < .05), whereas for 180° angular disparity, girls showed 30% lower accuracy than boys (p < .05). For all other disparities sex differences were non-significant (all ps > .14), suggesting that sex differences in the youngest age group were inconsistent.

1.2.2 Individual Analyses

Finally, data were categorized on the basis of individual children's performance level. Children who solved more than 14 out of the total 21 test trials correctly (i.e., the level that exceeded chance at p < .05 according to the binomial distribution) were categorized as "above chance performers". The numbers of above chance performers per age group demonstrated a clear age effect: 2 (10%) of the 3-year-olds, 8 (40%) of the 4-year-olds, and 19 (95%) of the 5-year-olds performed above chance (Fisher's exact test: p < .001). Inspection of individual response patterns further suggested that more than half of the 3-year-olds (13) exhibited a response bias and preferentially (i.e., more than 14 times) selected the ghost in one location. In contrast, only few of the 4- and 5-year-olds (5 and 3, respectively) showed such a bias towards one location.

1.3 Discussion

Results showed clear improvements in task performance between 3 and 5 years of age. The average percentage of correctly solved trials increased considerably, from near-chance performance (54% correct) at 3 years of age, to 83% correct at 5 years of age. Even more strikingly, analyses of individual participants' performance showed that very few 3-year-olds and less than half of the 4-year-olds exceeded chance performance in this mental rotation task, but that by 5 years of age, all but one child performed above chance. In both individual analyses and analyses on the group level, performance differed significantly between all age groups. Taken together, these results suggest considerable improvement in mental rotation ability between 3 and 5 years of age.

Analyses further showed a main effect of disparity, which followed a linear trend, indicating that performance in general decreased with increasing angle of rotation. Figure 3 suggested that this was more pronounced for older children, but there was only a trend to an interaction between disparity and age. A ceiling effect in the oldest age group may possibly have dampened this interaction, as by 5 years of age, accuracy was almost at ceiling for small disparities. However, 5-year-olds were still not perfect for larger disparities, leaving room for improvement even in the oldest age group tested.

A significant three-way interaction involving sex turned out to be due to an inconsistent advantage at larger angles of rotation for 3-year-olds, once in favor of boys (at 180°) and once in favor of girls (at 135°). At age 4 and 5, comparisons of boys' and girls' performance were non-significant. These data suggest that there are no consistent sex differences in mental rotation abilities at this young age. In this respect, it should be noted that floor effects might have masked sex differences in our 3-year-old group. In other words, sex differences would not have been expected if children were responding randomly. However, two observations render this possibility unlikely. First, sex differences that were significant (albeit inconsistently so) in post hoc tests were found at larger angles, where floor-

effects could have expected to be strongest. Second, by age 5, children were clearly performing above chance and still no sex effects could be detected, making it unlikely that sex differences were present in younger children.

Our task appeared to be somewhat easier than previous tasks. For example, whereas in our study, children at a mean age of 54 and 66 months, respectively, solved 69% and 83% of the trials correctly, in a previous chronometric study (Estes, 1998) children responded correctly on 60% and 74% of the trials at a mean age of 56 and 66 months, respectively. The slightly higher percentages in our task could be a result of the specific feedback the children received. In Estes' study, correct responses were followed by brief tunes and incorrect responses by silence. Our study actually allowed children to perform a manual rotation after their choice and thus might have provided more informative feedback.

To further investigate the role of contingent feedback within our paradigm, in Experiment 2 children were prevented from physically verifying their choices and only allowed to point to the correct ghost. The same ghost stimuli as in Experiment 1 were presented on sheets of paper rather than clued on foam board pieces, and thus children were not able to pick them up or rotate them. In that respect the stimulus presentation was more comparable to previous studies that presented stimuli two-dimensionally on plywood panels (e.g., Marmor, 1975, 1977), paper (e.g., Vandenberg & Kuse, 1978; Quaiser-Pohl, 2003) or computer screens (e.g, Estes, 1998; Frick et al., 2009; Funk et al., 2005; Kosslyn et al., 1990). Because 3-year-olds' responses were near chance level in Experiment 1, and there was no reason to assume that they would perform better without feedback, 3-year-olds were not included in Experiment 2.

To date, there is no paper-and-pencil test of mental rotation that would allow for easy assessment of this important spatial ability at this young age. Comparing 4- and 5-year-olds'

results from a hands-on, three-dimensional version of this task to a paper version could

provide the basis for an easy-to-administer tool for assessing early mental rotation abilities.

2. Experiment 2

2.1 Method

2.1.1 Participants

Forty children participated, with 20 children in each of two age groups: 4-year-olds

(mean age = 55 months, range = 53 – 56 months), and 5-year-olds (mean age = 66 months,

range = 65 – 68 months). There were 10 boys and 10 girls in each age group, and the mean

age difference between boys and girls was 10 days. One additional 4-year-old was tested but

excluded from analyses because he perseverated and chose the ghost in the same location on

all trials. The sample was recruited and tested as in Experiment 1.

2.1.2 Stimuli, procedure, and design

The ghost stimuli were printed on letter-size paper in landscape orientation (see

Figure 4) and were presented in document pockets inside a 3-ring binder. Two ghost figures

were printed on the right-hand side of the paper, in a vertical arrangement, similar to the

arrangement of the ghosts on the trays in Experiment 1. The bodies of the ghosts were light-

grey and the outlines were black. On the left-hand side of the paper and 2.5 cm to the left of

the ghost figures was a black circle, 10 cm in diameter, with a light-grey "hole" in the center

in the shape of one of the ghost figures. The green foam board with the asymmetrical artifact

pieces from Experiment 1 was used to instruct the task, followed by one trial with one of

these artifacts (hammer) printed on paper, to ease the transition from the manipulable

instruction trials to the test trials on paper. Except for this one additional trial, the orders and

design were identical to Experiment 1.

2.2 Results

2.2.1 Mean Accuracy

As in Experiment 1, data for trials with the same angular disparities were collapsed. Again, an ANOVA with age group (2), sex (2), order (4) and disparity (7) as independent variables and proportion of correctly solved trials as the dependent variable yielded no effects involving order (all Fs < 2.42, all ps > .09); therefore, this variable was not included in subsequent analyses.

A repeated measures ANOVA was performed with disparity (7) as the within-subject variable, age group (2) and sex (2) as the between-subjects variables, and proportion of correctly solved trials as the dependent variable. The analysis yielded a significant main effect of age group, $F(1, 36) = 6.47$, $p < .05$, $\eta^2 = .15$. Furthermore, there was a significant main effect of disparity, $F(6, 216) = 7.83$, $p < .001$, $\eta^2 = .18$, with a significant linear component, lin. $F(1, 36) = 30.44$, $p < .001$, $\eta^2 = .46$, but no interaction between disparity and age group, $F < 1$. All effects involving the variable of sex were non-significant (all Fs < 1.22, all ps > .30).

2.2.2 Individual Analyses

As in Experiment 1, children who solved more than 14 out of the 21 test trials correctly were categorized as "above chance performers". According to this criterion, 7 (35%) of the 4-year-olds, and 13 (65%) of the 5-year-olds performed above chance (Fisher's exact test: $p = .11$). Inspection of individual response patterns further suggested that four 4-year-olds and two 5-year-olds exhibited a response bias and preferentially (i.e., more than 14 times) selected the ghost in one location.

2.2.3 Comparison between Experiment 1 and Experiment 2

Finally, performance of the 4- and 5-year-olds in Experiments 1 and 2 were compared by means of an ANOVA with disparity (7) as the within-subject variable, experiment (2), age group (2), and sex (2) as between-subjects variables, and number of correct trials as the dependent variable. The analysis yielded significant main effects of age group, $F(1, 72) =$

17.04, $p < .001$, $\eta^2 = .19$, and disparity, $F(6, 432) = 16.52$, $p < .001$, $\eta^2 = .19$, but no

interaction between age group and disparity, $F < 1$. No significant effects involving the

variables of experiment or sex were found (all Fs < 2.51, all ps $> .10$).

On the individual level, there was no difference in the number of 4-year-olds

performing above chance in both experiments (Fisher's exact test: $p = 1.00$), although

significantly fewer 5-year-olds performed above chance in Experiment 2 than in Experiment

1 (Fisher's exact test: $p < .05$).

2.3 Discussion

In Experiment 2, stimuli were printed on paper and children were not able to move the

pieces into the holes. Thus, children did not receive feedback about the correctness of their

choices. Similar to Experiment 1, group analyses yielded a main effect of age, confirming

that 5-year-olds performed significantly better than 4-year-olds. These analyses also yielded a

main effect of disparity, showing that accuracy decreased with increasing angle of rotation,

but no interaction between age and disparity, even though a ceiling effect in 5-year-olds was

less pronounced in Experiment 2. This suggests that there was no important age difference

between 4- and 5-year-olds in their decrease in accuracy with increasing rotation angles, even

though 5-year-olds performed on a higher level overall.

On the individual level, fewer 5-year-olds performed above chance in the paper

version used in Experiment 2 as compared to Experiment 1. However, an ANOVA showed

that accuracy scores were not statistically different in the two Experiments on the group level.

Thus, results suggested that visual and haptic feedback on whether a piece would fit into a

hole after an executed rotation did not have a significant effect on 4- and 5-year-olds' average

group performance. Nevertheless, it is interesting to note that the slightly lower scores in

Experiment 2 were closer to the performance achieved in a previous chronometric study

discussed above (Estes, 1998). Thus, our results were comparable to previous results without substantial feedback.

In the present study, several instruction trials with cardboard stimuli were used to introduce the task. In these trials, children were repeatedly instructed to turn the pieces in their heads, in order to discourage trial-and-error strategies and to make the task more accessible to 3-year-olds in Experiment 1. The same instruction trials were used in Experiment 2, to rule out the possibility that differences in visual and haptic feedback would be confounded with differences in task instructions. However, in order to use the present task as an assessment tool, manipulable paper cutouts of the instruction stimuli would probably serve the same purpose, and fewer trials would likely suffice, especially if children are older than 3 years.

3. General Discussion

Results showed a clear developmental trend with a considerable increase in mental rotation abilities between 3 and 5 years of age. In Experiment 1, children were presented with mirror-image pairs of ghosts in different orientations and asked to pick the one that fits into a hole. The percentage of correctly-solved trials increased from 54% to 83%, and the number of individual children who performed above chance increased from 10% to 95% between 3 and 5 years of age. In Experiment 2, a paper version of the task was presented to 4- and 5-year-olds, and children did not receive feedback about the correctness of their responses, which resulted in a slightly lower overall accuracy and fewer children performing above chance level. However, an ANOVA comparing the results from both experiments yielded no significant main effect of experiment or interaction with disparity, suggesting that online feedback did not have a reliable effect on 4- and 5-year-olds' mental rotation performance.

By 5 years of age, the majority of children exceeded chance performance, and mean accuracy was almost at ceiling for small disparities. However, there was still room for

improvement for large disparities, in line with previous studies suggesting developmental progression in mental rotation abilities even in older children. Mental rotation has been shown to continuously strengthen through early childhood (Estes, 1998; Levine et al., 1999), and a study by Kail, Pellegrino, and Carter (1980) suggested that between Grade 4 and college, the developing speed of mental rotation may become an increasingly important source of individual differences. Furthermore, in a study that involved the mental rotation of hands, (Krüger & Krist, 2009) even at 6 years of age 40% of the children performed below chance level. Our results suggest that if task demands are lowered and children can take their time to choose an alternative, most 5-year-olds are able to perform above chance level.

At age 4, performance was significantly lower than at age 5, and less than half of the 4-year-olds performed above chance level. These results challenge Marmor's (1977) general conclusion that 4-year-olds are able to spontaneously perform mental rotations. Our results suggest that there are considerable individual differences in mental rotation abilities at age 4, in line with previous studies that failed to replicate Marmor's results (e.g., Dean & Harvey, 1979) or found evidence for mental rotation in only a subgroup of 4-year-olds (Estes, 1998).

For 3-year-olds this task proved even more challenging, although they were not pressured to respond quickly and did not have to remember any response buttons. Group means suggested that 3-year-olds performed near chance level, and on the individual level, only two out of twenty children performed significantly above chance. Moreover, a large number of 3-year-olds chose the ghost in one location significantly more often, suggesting that they fell into a perseverative response pattern because they lacked a successful cognitive strategy. In fact, for 3-year-olds, this task proved challenging regardless of how much the ghosts had to be rotated to match the holes. Even if the ghosts were presented in angular alignment with their holes (cf. 0°-condition in Figure 3), accuracy was close to chance level.

This result suggests that 3-year-olds were challenged in mentally matching the pieces to the holes, even if no rotation was necessary.

It seems unlikely that this problem was due to difficulties in comparing positive shapes (pieces) and negative shapes (holes), in light of previous findings that by 30 months of age, the ability to fit objects into holes is fairly well established (Shutts et al., 2009). A more likely explanation rests on the fact that even though on aligned trials the pieces did not have to be rotated, they still had to be mentally translated (horizontally) to compare them with their respective holes. A previous study investigated 4- to 6-year-olds' ability to mentally combine two halves of a shape (by translation or rotation) and compare them to four response alternatives showing whole shapes. Findings showed that even though translational items were solved significantly more often than rotational items, scores on translational items were still far from perfect (4.41 out of 8 items correct for horizontal translations; Levine et al., 1999). Even though this task was probably cognitively more demanding because shapes had to be combined, consistent with our results Levine et al. found that 5-year-olds performed significantly better than 4-year-olds, suggesting considerable developmental progression in rotational as well as translational mental transformation abilities at this age.

The finding that most 3-year-olds performed near chance is puzzling in light of recent research showing that even infants are able to distinguish between objects and their rotated mirror image versions (Frick & Möhring, 2013; Möhring & Frick, 2013; Moore & Johnson, 2008, 2011; Quinn & Liben, 2008; Schwarzer, Freitag, Buckel, & Lofruthe, 2012) or discriminate rotation events that are physically probable or improbable (Frick & Wang, 2013; Hespos & Rochat, 1997; Rochat & Hespos, 1996). Similar paradoxical discrepancies have been shown in other domains (e.g., Keen, 2003; Krist, 2010) and further research is needed to determine which factors affect task performance early in life.

In both experiments no significant interactions between age and disparity were found, suggesting that 4- and 5-year-olds both showed a similar decrease in accuracy with increasing rotation angles, even though 5-year-olds performed on a higher level. This result seems to be surprising given Figure 3, showing that 5-year-olds performed nearly perfectly for small angles, whereas 3-year-olds performed near chance level throughout. Thus, it is conceivable that the two-way interaction failed to reach statistical significance due to power issues. However, two results render this rather unlikely: First, the three-way interaction between those two variables and sex was significant in Experiment 1. Second, in the combined analysis of Experiments 1 and 2 with twice the numbers of 4- and 5-year-olds, disparity still did not interact with age. Furthermore, the possibility that a ceiling effect could have attenuated this interaction in Experiment 1, as discussed above, seems rather unlikely, given that in Experiment 2 such a ceiling effect was less pronounced and yet no interaction of age and disparity was found. Instead, it is possible that large individual differences in mental rotation performance accounted for this result, and indeed, the range of children's performance scores was quite wide in all age groups (see Table 1), ranging from chance performance to nearly maximum scores.

In contrast to a number of studies that found sex differences in mental rotation performance of older participants (for a meta-analysis, see Linn & Petersen, 1985), there were no consistent effects of sex in our task. Our results are in line with findings from several studies in infants (Frick & Wang, 2013; Hespos & Rochat, 1997; Möhring & Frick, 2013; Rochat & Hespos, 1996; Schwarzer et al., 2012) and young children (Estes, 1998; Frick et al., 2009; Kosslyn et al., 1990; Platt & Cohen, 1981) reporting no reliable sex differences in mental rotation performance. In a more recent meta-analysis, Voyer and colleagues (1995) found a positive relation between chronological age and effect size, and the majority of studies with an age range below 10 years reported no sex differences, suggesting that sex

differences in mental rotation abilities increase with age. Our results provide converging evidence that sex differences observed frequently in adults may develop later than early childhood (cf., Johnson & Meade, 1987; Titze, Jansen, & Heil, 2010). A possible reason for this may be an increased awareness of gender stereotypes, which has been found to affect adults' performance on paper-and-pencil mental rotation tasks (Moè & Pazzaglia, 2006; Heil, Jansen, Quaiser-Pohl, & Neuburger, 2012). Alternatively, hormonal affects (Hausmann, Slabbekoorn, Van Goozen, Cohen-Kettenis, & Güntürkün, 2000) or differential experience and practice (Terlecki & Newcombe, 2005; Voyer, Nolan, & Voyer, 2000), also in connection with socio-economic status (Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005), may account for sex differences later in development. Further research on the role of these variables in children and adolescents are needed to understand the source of gender differences in adults' mental rotation.

The absence of consistent sex differences could also be due to the fact that there was no time pressure involved in our task. Previous researchers have argued that women may be slower and more cautious than men when solving mental rotation tasks (Goldstein, Haldane, & Mitchell, 1990). Indeed, their data showed that sex differences were no longer significant when the time to complete a paper-pencil mental rotation tests was not limited. This task factor was further investigated in a meta-analysis (Voyer, 2011), confirming that eliminating the time pressure served to attenuate (although not eliminate) sex differences regardless of age (in an age range from 8 to 32 years). Our results are in line with this notion. However, in the present study, time pressure was not manipulated, and further research in young children is needed to find out at what age and under which conditions sex differences in mental rotation emerge.

The finding that the paper version in Experiment 2 led to fewer 5-year-olds performing above chance on an individual level is not very surprising, considering that in

Experiment 1, children received visual and haptic feedback about the correctness of their choices and thus learning effects could have been expected. However, the analysis comparing the accuracy scores from both experiments yielded no significant effects, suggesting that children did not perform significantly better in Experiment 1 on the group level. One possible explanation for these diverging results on the individual and group level is that children who already had some basic mental rotation skills improved through feedback, whereas children who did not know how to solve the task were frustrated by the often negative feedback and thus performed worse. Hence on the individual level, some children may have profited from feedback and exceeded chance, whereas on the group level accuracy scores remained the same.

This result may indicate that some 5-year-olds were developmentally ready (cf. Vygotsky, 1978) to acquiring mental rotation skills and were able to process feedback information, whereas others (and especially 4-year-olds) were not cognitively ready to profit from feedback yet. In line with these results, a recent study (Frick, Ferrara, & Newcombe, 2013) showed that observational and manual experience improved mental rotation performance of 5-year-olds who were performing above chance, but did not improve the performance of 4-year-olds. Individual differences in cognitive readiness at age 5 also could explain the incongruent results regarding the trainability of mental rotation in previous studies (e.g., Marmor, 1977; Platt & Cohen, 1981).

However, the present findings should not be interpreted to imply that younger children would not profit from more intense practice or other kinds of feedback. Previous work (Levine, Ratliff, Huttenlocher, & Cannon, 2012) showed that puzzle play frequency, observed over the course of two years, was positively correlated to children's mental transformation performance at 4.5 years of age. More specifically, the quality of puzzle play (i.e., difficulty, level of engagement, and spatial language used in parent-child interactions) predicted girls' performance. Thus, with the right support (e.g., a parent commenting the

visual feedback children receive as they do a puzzle), younger children may very well profit from feedback through practice. This finding also suggests that increased experience and practice with spatial toys may be a driving force behind the age differences observed between 3 and 5 years in the present study.

The paper version of our task proved useful for assessing children's mental rotation abilities starting at age 4 and provides an easy-to-administer assessment tool that measures a behaviorally relevant ability at an early age. Furthermore, the paper version can assess mental rotation abilities using a task and dependent variable similar to those used in adult mental rotation research, thus allowing for comparison with older age groups. To date, paper-pencil assessments have only been successfully used in children as young as 5 years (Quaiser-Pohl, 2003), at which age children performed at an above chance level. The present task allows for assessing mental rotation in younger children, providing a research tool that can help narrow the gap between infant research and research with older children and adults. Further research using comparable methodology will allow us to obtain a more cohesive picture of the developmental trajectory of mental rotation abilities. A deeper understanding of the early development of mental rotation is indispensable for developing training programs and for promoting this basic spatial skill at a young age.

Taken together, our results suggest a developmental progression of mental rotation abilities between 3 and 5 years of age, with 3-year-olds showing near-chance performance, but about half of the 4-year-olds and almost all of the 5-year-olds performing at an above-chance level. Furthermore, the present task yielded no consistent performance differences in favor of boys or girls. Finally, a paper version of the task yielded similar results, providing us with a mental rotation assessment tool suitable for early childhood. This may be instrumental in future research and in promoting higher-level cognitive functioning and academic achievement.

4. Acknowledgements

5. References

Caldwell, E. C., & Hall, V .C. (1970). Concept learning in discrimination tasks. *Developmental Psychology, 2*, 41-48.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY, US, Cambridge University Press.

Cronin, V. (1967). Mirror-image reversal discrimination in kindergarten and first grade children. *Journal of Experimental Child Psychology, 5*, 577-585.

Dean, A. L., & Harvey, W. O. (1979). An information-processing analysis of a Piagetian imagery task. *Developmental Psychology, 15*, 474-475.

Estes, D. (1998). Young children's awareness of their mental activity: The case of mental rotation. *Child Development, 69*, 1345-1360.

Frick, A., Daum, M. M., Walser, S., & Mast, F. W. (2009). Motor processes in children's mental rotation. *Journal of Cognition and Development, 10*, 18-40.

Frick, A., Ferrara, K., & Newcombe, N. S. (2013). Using a touch screen paradigm to assess the development of mental rotation between 3½ and 5½ years of age. *Cognitive Processing, 14,* 117-127.

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

Frick, A., & Wang, S. (2013). Mental spatial transformations in 14- and 16-month-old infants: Effects of action and observational experience. *Child Development.* DOI: 10.1111/cdev.12116.

Funk, M., Brugger, P., & Wilkening, F. (2005). Motor processes in children's imagery: The case of mental rotation of hands. *Developmental Science, 8*, 402-408.

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition, 18*, 546-550.

Hausmann, M., Slabbekoorn, D., Van Goozen, S. H., Cohen-Kettenis, P. T., & Gunturkun, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral neuroscience, 114*, 1245-1250.

Heil, M., Jansen, P., Quaiser-Pohl, C., & Neuburger, S. (2012). Gender-specific effects of artificially induced gender beliefs in mental rotation. *Learning and Individual Differences, 22,* 350–353.

Hespos, S. J., & Rochat, P. (1997). Dynamic mental representation in infancy. *Cognition, 64*, 153-188.

Jahoda, G. (1979). On the nature of difficulties in spatial-perceptual tasks: Ethnic and sex differences. *British Journal of Psychology, 70*, 351-363.

Johnson, E. S., & Meade, A. C. (1987). Developmental patterns of spatial ability: An early sex difference. *Child Development, 58*, 725-740.

Kaess, D. W. (1971). Measures of form constancy: Developmental trends. *Developmental Psychology, 4*, 296.

Kail, R., Pellegrino, J., & Carter, P. (1980). Developmental changes in mental rotation. *Journal of Experimental Child Psychology, 29*, 102-116.

Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science, 12*, 79-83.

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age differences in imagery ability. *Child Development, 61*, 995-1010.

Krist, H., (2010). Development of intuitions about support beyond infancy. *Developmental Psychology, 46*, 266–278.

Krüger, M. & Krist, H. (2009). Imagery and motor processes - when are they connected? The mental transformation of body parts in development. *Journal of Cognition and Development, 10*, 239-261.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in

spatial skills. *Developmental Psychology, 35*, 940-949.

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A

predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48*,

530-542.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in

spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498.

Marmor, G. S. (1975). Development of kinetic images: When does the child first represent

movement in mental images? *Cognitive Psychology, 7*, 548-559.

Marmor, G. S. (1977). Mental rotation and number conservation: Are they related?

Developmental Psychology, 13, 320-325.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in

mental rotation. *Learning and Individual Differences, 16*, 369–377.

Möhring, W., & Frick, A. (2013). Touching up mental rotation: Effects of manual experience

on 6-month-old infants' mental object rotation. *Child Development*. DOI:

10.1111/cdev.12065.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference.

Psychological Science, 19, 1063-1066.

Moore, D. S., & Johnson, S. P. (2011). Mental rotation of dynamic three-dimensional stimuli

by 3-month-old infants. *Infancy, 16*, 435-445.

Newcombe, N. S. (2002). The nativist-empiricist controversy in the context of recent research

on spatial and quantitative development. *Psychological Science*, *13*, 395-401.

Örnkloo, H., & von Hofsten, C. (2007). Fitting objects into holes: On the development of

spatial cognition skills. *Developmental Psychology*, *43*, 404-416.

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

Quaiser-Pohl, C. (2003). The mental-cutting test "Schnitte" and the Picture-Rotation-Test: Two new measures to assess spatial ability. *International Journal of Testing, 3*, 219–231.

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

Rochat, P., & Hespos, S. J. (1996). Tracking and anticipation of invisible spatial transformation by 4- to 8-month-old infants. *Cognitive Development, 11*, 3-17.

Schwarzer, G., Freitag, C., Buckel, R., & Lofruthe, A. (2012). Crawling is associated with mental rotation ability by 9-month-old infants. *Infancy, 18*, 432–441.

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

Shutts, K., Örnkloo, H., von Hofsten, C., Keen, R., Spelke, E. S. (2009). Young children's representations of spatial and functional relations between objects. *Child Development, 80*, 1612-1627.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles, 53*, 433-441.

Titze, C., Jansen, P., & Heil, M. (2010). Mental rotation performance and the effect of gender in fourth graders and adults. *European Journal of Developmental Psychology, 7*, 432-444.

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

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin & Review, 18*, 267–277.

Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles, 43*, 891-915.

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*, 250–270.

Vygotsky, L. (1991). Interaction between learning and development. In M. Gauvain & M. Cole (Eds.). *Readings on the development of children* (pp. 29-36). New York: W.H. Freeman and Company. (Reprinted from *Mind in Society*, 1978, Cambridge, MA: Harvard University Press, 79-91.)

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817-835.

Figure Captions

Figure 1. Examples of stimuli: two asymmetrical mirror-reversed ghost pieces (left) and

board with an empty hole for one of the ghosts (right).

Figure 2. Overall accuracy (number of correct responses out of 21 trials in %) per age group

in Experiments 1 and 2. Error bars represent standard errors of means.

Figure 3. Accuracy (correct responses in %) by disparity per age group in Experiment 1.

Error bars represent standard errors of means.

Figure 4. Example of two-dimensional stimulus, printed on letter-size paper in Experiment

2.

Figure 5. Accuracy (correct responses in %) by disparity per age group in Experiment 2.

Error bars represent standard errors of means.