Early Sex Differences in Spatial Skill

Susan C. Levine, Janellen Huttenlocher, Amy Taylor, and Adela Langrock
University of Chicago

This study investigated sex differences in young children's spatial skill. The authors developed a spatial transformation task, which showed a substantial male advantage by age 4 years 6 months. The size of this advantage was no more robust for rotation items than for translation items. This finding contrasts with studies of older children and adults, which report that sex differences are largest on mental rotation tasks. Comparable performance of boys and girls on a vocabulary task indicated that the male advantage on the spatial task was not attributable to an overall intellectual advantage of boys in the sample.

The purpose of the present study was to investigate the emergence of sex differences in spatial skill during early childhood. The ability to represent and transform spatial information is a vital component of human intellectual competence. It is important in everyday activities, such as navigating in a new city or finding a car in a parking lot when approaching from a different direction. Even greater demands on spatial skill are made by various technical tasks pervasive in a complex society, such as interpretation of graphs, maps, architectural drawings, and X-rays. Such tasks often require the ability to mentally transform images and reconstruct 3-D forms from two-dimensional (2-D) images.

The age at which sex differences in spatial skill emerge has been a matter of debate. An influential book by Maccoby and Jacklin (1974) claimed that sex differences in spatial skill emerge with the onset of adolescence, leading to the widespread belief that such differences are a relatively late developmental phenomenon (e.g., Christiansen & Knussman, 1987; Hier & Crowley, 1982; McGee, 1979; Nyborg, 1983, 1984; Waber, 1977). In contrast to this view, several studies have reported sex differences on spatial tasks as early as the preschool years. Boys as young as 4 years of age performed better than girls on a task that involved replicating spatiotemporal patterns tapped out by the experimenter on a set of blocks, and the size of this sex difference remained constant across the 4- to 10-year age range (Grossi, Orsini, Monetti, & De Michele, 1979; Orsini, Schiappa, & Grossi, 1981). Furthermore, preschool boys have been found to perform at a higher average level than preschool girls on the Mazes subtest of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Fairweather & Butterworth, 1977; Wechsler, 1967; Wilson, 1975). Four- and 5-year-old boys also copied a 3-D Lego model faster than same-age girls but did not differ from girls on a 2-D puzzle task (McGuinness & Morley, 1991). Kindergarten boys more accurately constructed a 3-D model of their classroom than did kindergarten girls (Siegel & Schadler, 1977).

Several additional studies have reported sex differences in young children on spatial tasks involving mental rotation. Kindergarten and first-grade boys performed better than girls in discriminating mirror reversals of triangles from identical triangles (Cronin, 1967), a task that may involve mental rotation. Four and 5-year-old boys performed better than girls in discriminating a particular 2-D rotation of a stimulus with salient external features from foils, which included a mirror reversal of the correct choice and other 2-D rotations (Rosser, Ensing, Gilder, & Lane, 1984). Utal, Gregg, and Chamberlain (1999) found that 5-year-old boys were better at interpreting a map of a space than 5-year-old girls, particularly when the map was rotated with respect to the space it represented. Although 3-year-old girls had higher average performance than 3-year-old boys on the nonrotated map task—the only condition administered to this age group—Utal et al. reported that the 3-year-old boys did not appear to be as engaged in the task as 3-year-old girls.

In adults and older children, meta-analyses indicate that the most robust sex differences are found on spatial tasks involving mental rotation (Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995). The classic mental rotation task, developed by Shepard and Metzler (1971), is a reaction time task in which participants judge whether pairs of 2-D projections represent different 3-D forms or the same 3-D form in different orientations. The reaction time to make a decision is typically found to be a linear function of the number of degrees of separation between the two forms (e.g., Shepard, 1975; Shepard & Metzler, 1971). A paper-and-pencil version of the Shepard and Metzler task showed a marked sex difference (Vandenberg & Kuse, 1978), on the order of 0.7 standard deviations, compared with a sex difference on the order of 0.3 on other types of spatial tasks, such as spatial perception tasks that involved determining spatial relations with respect to body orientation (Linn & Petersen, 1985; Voyer et al., 1995).

In view of the robustness of sex differences on mental rotation tasks in older children and adults, it would be reasonable to hypothesize that the earliest manifestation of sex differences in spatial skill would be on tasks that involve mental rotation. However, preschool children show a male advantage on a number of...
spatial tasks, including those that appear to involve mental rotation and those that do not, with little or no evidence that the advantage is more robust when mental rotation is involved. These data raise the possibility that sex differences on spatial tasks early in life are not more robust for tasks involving mental rotation in preschoolers, in contrast to adults. Sex differences on tasks involving mental rotation are more consistently reported beginning at about 8 years of age (e.g., Guay & McDaniel, 1977; Johnson & Mcade, 1987; Kerns & Berenbaum, 1991). Furthermore, in 8- to 9-year-olds, as in adults, mental rotation has been implicated as a major source of sex differences on spatial tasks. For example, several studies reported that in the middle childhood years, boys outperform girls on spatial tasks involving mental rotation but not on spatial tasks that do not involve mental rotation (Guay & McDaniel, 1977; Richmond, 1980).

What might explain the lack of evidence for a more robust sex difference in favor of boys on spatial tasks involving mental rotation in preschool children? One possibility is that mental rotation tasks that have been given to preschool children mask the existence of sex differences. Many of the mental rotation tasks given to older children and adults use response time as a measure, which may not be appropriate for young children because of the attentional demands of speeded response. Because such mental rotation tasks are difficult for preschool children, this skill is not frequently assessed in this age range. For example, the nature of the Spatial Relations Test on the Primary Mental Abilities Test involves identifying picture-plane rotations of forms for children in Grades 6–12 but involves identifying the part that will make a complete square for children in kindergarten to Grade 6 (Thurstone & Thurstone, 1962). Another possibility is that the more robust sex difference on spatial tasks involving mental rotation develops over time and may not be present early in life.

The present research addressed the question of whether preschool children show a sex difference on a nonspeeded spatial task involving various types of mental transformations, one example of which is mental rotation. It was first essential to develop a task that taps the ability to mentally transform spatial stimuli that is at an appropriate level of difficulty for preschool children. The task we developed requires the child to mentally integrate two separate parts of a shape to form a single complete shape and to indicate which of four shapes in a choice array matches the mental representation formed. The spatial transformations used were all in the picture plane, in contrast to the 3-D transformations that have typically been used with adults (e.g., Shepard & Metzler, 1971; Vandenberg & Kuse, 1978). The use of picture-plane transformations was guided by prior work that has shown that this type of task is easier than tasks involving mental rotation in depth (see Linn & Petersen, 1985). Moreover, studies with adult participants showed sex differences with both 2-D and 3-D mental rotation tasks (Shepard & Cooper, 1982), although a meta-analysis showed a larger sex difference for 3-D tasks (Linn & Petersen, 1985). The spatial transformation task included both problems that could be solved by mentally rotating shapes and problems that could be solved by mentally translating shapes. The finding of a comparable male advantage on both types of problems would support an early sex difference in spatial transformation skill. In contrast, the finding of a larger male advantage on rotation problems than on translation problems would support a more specific early sex difference in mental rotation skill.

In our first study, children ranging in age from 4 to 7 years were given the spatial transformation task and the Vocabulary and Mazes subtests from the WPPSI—Revised (WPPSI-R; Wechsler, 1989). The inclusion of the Mazes subtest allowed us to determine whether the performance of our sample of preschool boys exceeded that of our sample of preschool girls, mirroring the results of prior studies (Fairweather & Butterworth, 1977; Wilson, 1975). The Mazes subtest also provided another way to examine whether the sex difference in spatial skill during the preschool years differed for tasks involving mental rotation and other types of spatial tasks. The inclusion of the Vocabulary subtest allowed us to rule out the possibility that any male superiority on the spatial tasks was attributable to generally higher levels of intellectual skill of boys than of girls in our sample. We carried out a second study to examine whether the effect of test half found in Experiment 1 was attributable to practice effects.

**Experiment 1**

**Method**

**Participants.** Two hundred eighty-eight children participated in the study. Participants were divided into six age groups of 48 children (24 boys and 24 girls) in each 6-month interval: (a) Age Group 1: 4 years to 4 years 5 months; (b) Age Group 2: 4 years 6 months to 4 years 11 months; (c) Age Group 3: 5 years to 5 years 5 months; (d) Age Group 4: 5 years 6 months to 5 years 11 months; (e) Age Group 5: 6 years to 6 years 5 months; and (f) Age Group 6: 6 years 6 months to 6 years 11 months. The children attended parochial schools in the Chicago metropolitan region. The mean age (in months) of each age group by sex is shown in Table 1.

**Materials.** There were 32 problems on the spatial transformation task. Each problem consisted of two target pieces as well as a 2 X 2 choice array that included the target shape that could be formed by the two target pieces and three foils. Sixteen target shapes were unilaterally symmetric around the vertical axis, and 16 were bilaterally symmetric around the horizontal and vertical axes. Target pieces were created by dividing each target shape in half along the vertical axis. The three foils for each target shape were constructed by adding or subtracting angles, curves, or lines and by adding features not included in the target shape. The target pieces, that is, the two pieces that made the target shape when mentally moved, were solid black and were displayed on 5½” X 8” (13.97 cm X 20.32 cm) white cards. The four choice shapes were also solid black and were displayed in 2 X 2 arrays on 9” X 10” (22.86 cm X 25.40 cm) white cards. Other materials consisted of the Mazes and Vocabulary subtests of the WPPSI-R (Wechsler, 1989).

**Design and procedure.** Participants were tested individually on two separate days. Each of the two test sessions lasted about 15 min. During the first test session, the 32-item spatial transformation task was administered. During the second test session, participants were administered the Mazes and Vocabulary subtests from the WPPSI-R (Wechsler, 1989). The Mazes

<table>
<thead>
<tr>
<th>Age group</th>
<th>Age range</th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 yr to 4 yr 5 mo</td>
<td>50.54</td>
<td>1.72</td>
</tr>
<tr>
<td>2</td>
<td>4 yr 6 mo to 4 yr 11 mo</td>
<td>56.29</td>
<td>1.81</td>
</tr>
<tr>
<td>3</td>
<td>5 yr to 5 yr 5 mo</td>
<td>63.04</td>
<td>1.76</td>
</tr>
<tr>
<td>4</td>
<td>5 yr 6 mo to 5 yr 11 mo</td>
<td>69.17</td>
<td>1.31</td>
</tr>
<tr>
<td>5</td>
<td>6 yr to 6 yr 5 mo</td>
<td>74.58</td>
<td>1.89</td>
</tr>
<tr>
<td>6</td>
<td>6 yr 6 mo to 6 yr 11 mo</td>
<td>80.79</td>
<td>1.72</td>
</tr>
</tbody>
</table>
subtest was administered first because it precedes the Vocabulary subtest on the WPPSI-R. These subtests were administered according to the directions in the test manual. On the Mazes subtest, the child is asked to solve pencil-and-paper mazes of increasing difficulty. There are 11 items on the test, but administration is stopped after the child fails 2 consecutive items. The easier mazes are linear in form, and the more difficult mazes are box mazes. Each maze has a time limit, and different scores are given depending on the number of errors the child makes. The Vocabulary subtest has two parts. Items on the first part involve naming a pictured object, and items on the second part involve providing verbal definitions for orally presented words (Wechsler, 1989). On the second part of the test, children's scores on each item depend on the quality of the definition. There are 3 picture-naming items and 22 verbal definition items. Testing is discontinued when the child misses 5 consecutive items on the definition portion of the test. Fifteen of the 288 participants (approximately 5%) did not take the WPPSI-R subtests because they were unavailable for a second session. Seven of these 15 were in the youngest age group, and the remainder were roughly evenly distributed across the remaining age groups.

The spatial transformation task consisted of 32 problems, each involving a different target shape. On each problem, the child was shown two halves of a shape that had been divided along the vertical axis. The child's task was to select the whole shape from among four choices in a 2 × 2 array that could be formed from the halves (see Figure 1 for sample item). The target shape appeared eight times at each of the four possible positions in the 2 × 2 array. The position of the target shape in the choice array was randomly varied across trials, with the constraint that it could not appear in the same position on more than two consecutive trials.

Both the stimulus card (card with the target pieces) and the choice array (card with four whole shapes) were placed on a table in front of the child. The choice array was placed closest to the child, and the stimulus card with the target pieces was placed directly above it. On the first trial, the experimenter gestured to the target pieces and then to the array of four shapes and said, "Look at these pieces. Look at these pictures. If you put the pieces together, they will make one of the pictures. Point to the picture the pieces make." On subsequent trials, the experimenter said, "Point to the picture the pieces make." No feedback was given on any item. Pilot testing showed that there was no need to give practice items.

Each participant viewed four different types of problems, eight trials of each type. Figure 2 shows the four types of problems, which varied with respect to the relative positioning of the two target pieces as follows: In the horizontal translation (Figure 2A), the two target pieces were separated so that their closest points were about 2 cm apart on the horizontal axis; in the diagonal translation (Figure 2B), the two target pieces were separated so that their closest points were about 2 cm apart on both the vertical and horizontal axes; in the horizontal rotation (Figure 2C), each target piece was rotated 60° from the vertical axis, one clockwise and the other counterclockwise, and the closest points of the figures were separated by about 2 cm along the horizontal axis; in the diagonal rotation (Figure 2D), the target pieces were each rotated 60° from the vertical axis, and the closest points of the pieces were separated by about 2 cm along both the horizontal and vertical axes. For diagonal translation and diagonal rotation configurations, the target pieces were presented with the left piece higher than the right piece on half the trials (16 trials) and vice versa on the other half. The order of problem types was randomized across the 32 problems, with the constraints that the same problem type was not presented twice in a row and that each type occurred four times in the first half of the trials and four times in the second half.

Four different forms of the task were used in the study, and each form was given to one quarter of the participants (6 girls and 6 boys) in each age group. The forms varied in the positioning of the target pieces for a particular target shape but were identical in the order of the 32 target shapes. For example, Target Shape 1 was the same across the four forms, but in Form A the pieces were displayed in a horizontal translation configuration; in Form B they were displayed in the diagonal translation configuration; in Form C they were displayed in the horizontal rotation configuration; and in Form D they were displayed in the diagonal rotation configuration.

Results

Our first analyses focused on the spatial transformation task. The distribution of boys and girls performing at different levels on this task is shown in Figure 3. Although some girls performed at the level of the highest performing boys, more girls than boys performed at the low end of the score distribution and more boys than girls performed at the high end of the distribution. This overlapping pattern of skill levels is also reported in adults (e.g., Caplan, MacPherson, & Tobin, 1985).
We performed an analysis of variance (ANOVA) on the number of items each participant responded to correctly, with age group and sex as between-subjects variables and problem type (horizontal translation, diagonal translation, horizontal rotation, diagonal rotation), target shape (unilaterally symmetrical, bilaterally symmetrical), and test half as within-subjects variables. All 288 participants were included in this analysis. The mean overall scores and standard deviations for each age group and for the group as a whole are reported in Table 2.

A main effect of age reflected a general improvement in scores with increasing age, $F(5, 276) = 46.32$, $p < .0001$. Tukey's honestly significant difference (HSD) tests revealed that the performance level of the two youngest age groups did not differ from one another, but both were significantly lower than each of the older age groups ($p < .05$ in each case). Tukey's HSD tests also showed that Age Group 3, which consisted of children 5 years old to 5 years 5 months, performed significantly better than each of the younger age groups and significantly worse than each of the three oldest age groups ($p < .05$ in each case), whose performance levels did not significantly differ from each other.

The main effect of sex also was significant, $F(1, 276) = 7.89$, $p < .005$. Across the six age groups, the mean number of target shapes responded to correctly was 18.02 ($SD = 6.84$) for boys and 16.30 ($SD = 6.89$) for girls (out of total possible number correct of 32). A calculation of Cohen's $d$ yielded an effect size of .25 for the sex difference across age (Cohen, 1977). Although the Sex $\times$ Age Group interaction was not significant, $F(5, 276) < 1$, $p > .10$, planned comparisons revealed that on average boys performed better than girls in each age group from 4 years 6 months to 6 years 11 months, but not in the youngest age group. The mean performance level of the youngest age group (4 years 5 months to 4 years 11 months) was significantly higher than the chance score of 8, $t(47) = 3.93$, $p < .0001$, but performance level in this group may not have been sufficiently high to allow any existing sex difference to be detected because of decreased variance.

### Table 2

**Spatial Transformation Task Scores by Sex and Age Group**

<table>
<thead>
<tr>
<th>Age group</th>
<th>Age range</th>
<th>Boys</th>
<th></th>
<th></th>
<th>Girls</th>
<th></th>
<th></th>
<th>All</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M SD</td>
<td>M SD</td>
<td>M SD</td>
<td>M SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4 yr to 4 yr 5 mo</td>
<td>9.96</td>
<td>2.99</td>
<td>9.96</td>
<td>3.93</td>
<td>9.96</td>
<td>3.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4 yr 6 mo to 4 yr 11 mo</td>
<td>13.21</td>
<td>4.62</td>
<td>11.79</td>
<td>4.29</td>
<td>12.50</td>
<td>4.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5 yr to 5 yr 5 mo</td>
<td>18.79</td>
<td>5.01</td>
<td>14.92</td>
<td>6.85</td>
<td>16.85</td>
<td>6.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5 yr 6 mo to 5 yr 11 mo</td>
<td>21.33</td>
<td>6.08</td>
<td>18.96</td>
<td>5.89</td>
<td>20.15</td>
<td>6.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6 yr to 6 yr 5 mo</td>
<td>22.17</td>
<td>5.67</td>
<td>20.38</td>
<td>5.33</td>
<td>21.27</td>
<td>5.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6 yr 6 mo to 6 yr 11 mo</td>
<td>22.83</td>
<td>5.21</td>
<td>21.79</td>
<td>5.44</td>
<td>22.31</td>
<td>5.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All ages</td>
<td></td>
<td>18.02</td>
<td>6.84</td>
<td>16.30</td>
<td>6.89</td>
<td>17.16</td>
<td>6.91</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Maximum score = 32.
The main effect of problem type, \( F(3, 828) = 13.60, p < .0001 \), reflected better performance on the two translation item types than on the two rotation item types. Tukey's HSD tests revealed that scores on both of the translation problem types were significantly higher than scores on both of the rotation problem types (\( p < .05 \) in each case). However, scores on horizontal and diagonal translations did not significantly differ from each other, and scores on horizontal and diagonal rotations did not significantly differ from each other. Table 3 lists the means and standard deviations for each problem type for boys and girls, collapsed across age. The Sex \( \times \) Problem Type interaction was not significant, indicating that the male advantage on items involving mental rotation versus mental translation did not differ, \( F(3, 276) < 1, p > .10 \).

The main effect of target shape, \( F(1, 276) = 577.33, p < .0001 \), was highly significant, reflecting better performance on bilaterally symmetric target shapes than on unilaterally symmetric target shapes (\( M = 10.22, SD = 5.54 \) vs. \( M = 6.92, SD = 6.22 \)). This difference is most likely attributable to the redundant cues provided by the bilaterally symmetric target shapes. Target shape did not interact with any other variable.

Finally, there was a highly significant main effect of test half, \( F(1, 276) = 24.12, p < .0001 \), reflecting better performance on the second half of items than the first half of items (\( M = 9.00, SD = 3.99 \) vs. \( M = 8.17, SD = 3.47 \)). This main effect was modified by a significant interaction of Test Half \( \times \) Age Group, \( F(5, 276) = 3.30, p < .007 \), as shown in Figure 4. Tests of simple effects showed that children in the four oldest age groups, who were 5 years and over in age, performed significantly better on the second half of trials, but children in the youngest two age groups did not (Group 1, \( p = .92 \); Group 2, \( p = .84 \); Group 3, \( p < .002 \); Group 4, \( p < .08 \); Group 5, \( p < .0001 \); Group 6, \( p < .0001 \)). As was suggested for the absence of a sex difference in the youngest age group, the absence of a test-half effect for younger participants may be attributable to their low performance level and decreased variance. Test half did not interact with sex or problem type (\( F < 1 \) in both cases). It should be noted that because only one fixed random order of items was used in this experiment, it is possible that the easier items ended up in the second half of the test trials by chance. In this case, practice would not be responsible for the better performance on the second half of items. Experiment 2 was a control experiment designed to disambiguate the nature of the test-half effect.

We performed a multivariate analysis of variance on the 273 children who performed all three tasks (spatial transformation task, WPPSI-R Mazes subtest, WPPSI-R Vocabulary subtest). Scores on the three tests were dependent variables, and sex was a between-subjects variable. Scores on the Mazes and Vocabulary subtests are standard scores that are not expected to vary with age. For comparability, we converted scores on the spatial transformation task to z scores within each age group. Mean standard scores and standard deviations for the Mazes and Vocabulary subtests and mean z scores and standard deviations for the spatial transformation task are reported in Table 4 for boys and girls in each age group. A multivariate test revealed a significant main effect of sex, \( F(3, 269) = 3.25, p < .025 \), based on Wilk's lambda. Univariate F tests for sex were significant for the spatial transformation task, \( F(1, 271) = 6.50, p < .02 \), and for WPPSI-R Mazes, \( F(1, 271) = 6.08, p < .02 \), but not for WPPSI-R Vocabulary (\( F < 1, p > .10 \)). The absence of a sex difference on the Vocabulary test indicated that the male performance advantage on the two spatial tasks was not attributable to a general intellectual advantage of our sample of boys compared with our sample of girls. A calculation of Cohen's d yielded an effect size of .30 on the WPPSI Mazes test and .10 on the WPPSI Vocabulary test. Cohen's d for the spatial transformation task was .31 when calculated on z scores for the 273 participants who were administered all three tasks. The somewhat higher effect size for z scores than for raw scores for the spatial transformation task was attributable to the lower variance that results because z scores are calculated within age group.

We carried out correlational analyses to examine the relation of participants' scores on the three tasks administered. The zero-order correlations were low but significant: Mazes and Vocabulary tests, \( r(1, 271) = .17, p < .005 \); Vocabulary test and spatial transformation task, \( r(1, 271) = .25, p < .0001 \); Mazes test and spatial transformation task, \( r(1, 271) = .30, p < .0001 \). Partial correlations revealed that the relation of Vocabulary and spatial transformation scores, \( r(1, 271) = .21, p < .0001 \), and the relation of Mazes and spatial transformation scores, \( r(1, 271) = .26, p < .0001 \), remained significant when each was adjusted by removing the variance attributable to the other task. Performance on the Mazes subtest independently accounted for 7.1% of variance on the spatial transformation task, and performance on the Vocabulary subtest independently accounted for 4.7% of the variance on the spatial transformation task. A small amount of additional variance in performance on the spatial transformation task (1.3%)
EARLY SEX DIFFERENCES IN SPATIAL SKILL

Table 4

<table>
<thead>
<tr>
<th>WPPSI-R Mazes and Vocabulary Subtests Scaled Scores and Spatial Transformation Task z Scores by Sex and Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4 yr to 4 yr 5 mo</td>
</tr>
<tr>
<td>4 yr 6 mo to 4 yr 11 mo</td>
</tr>
<tr>
<td>5 yr to 5 yr 5 mo</td>
</tr>
<tr>
<td>5 yr 6 mo to 5 yr 11 mo</td>
</tr>
<tr>
<td>6 yr to 6 yr 5 mo</td>
</tr>
<tr>
<td>6 yr 6 mo to 6 yr 11 mo</td>
</tr>
<tr>
<td>All ages</td>
</tr>
</tbody>
</table>

Note. WPPSI-R = Wechsler Preschool and Primary Scale of Intelligence—Revised. Scaled scores on WPPSI-R subtests have a mean of 10 and a standard deviation of 3.

It was shared by the Vocabulary and Mazes tasks. The relation of the two IQ subtests to the spatial transformation task did not significantly differ for boys and girls.

Discussion

A significant male advantage was found on the spatial transformation task by 4.5 years of age. In addition, consistent with previous reports in the literature (Fairweather & Butterworth, 1977; Wilson, 1975), a sex difference in favor of boys emerged on the WPPSI-R Mazes subtest between 4 and 5 years of age. The effect size of the sex difference on these two spatial tasks was comparable. The absence of a sex difference on the WPPSI-R Vocabulary subtest indicates that the male advantage on the two spatial tasks is not attributable to generally higher test scores in our sample of boys than in our sample of girls.

Results of the present study, together with earlier findings of sex differences in preschoolers on a variety of spatial tasks (Cronin, 1967; Fairweather & Butterworth, 1977; McGuiness & Morley, 1991; Rosser et al., 1984; Siegel & Schadler, 1977; Uttal et al., 1999; Wilson, 1975), indicate that children begin to show sex differences on at least a subset of spatial tasks by the preschool years. These findings should put to rest claims that adolescence marks the onset of sex differences in spatial skill. However, the finding that the average performance level of boys is better than that of girls as early as the preschool years on certain spatial tasks does not preclude the possibility that the magnitude or even the nature of sex differences on spatial skills changes over the course of development. In fact, a recent meta-analysis suggests that the magnitude of sex differences does increase between childhood and young adulthood (Voyer et al., 1995).

The magnitude of the sex difference on our spatial transformation task is quite similar to that shown by older children and adults on certain tasks involving mental rotation but somewhat smaller than that shown on other tasks involving mental rotation (weighted estimator of effect size on Card Rotation Test = .31; Vandenberg Mental Rotation Test = .67; generic Shepard–Metzler mental rotation task = .37; and Spatial Relations subtest of the Primary Mental Abilities Test = .44; Voyer et al., 1995). Because different tasks have been used to assess spatial transformation skill in young children and adults, it is not possible to determine whether there is a developmental increase in the magnitude of sex differences in spatial skill. Just as the magnitude of the sex difference on different mental rotation tasks varies for adults, the difference in the magnitude of effect sizes for young children versus adults may be attributable to varying task demands and the extent to which these demands tap aspects of spatial skills that show sex differences in the different age groups (Linn & Petersen, 1985).

Although use of the same task to assess spatial transformation skill in young children and adults would seem to be advantageous, it is difficult in practice. In particular, the tasks that are at an appropriate level of difficulty for adults are typically too difficult for young children, and those that are at an appropriate level of difficulty for young children are typically too easy for adults. Although reaction time techniques may offer a way to use the same task across age groups, such measures are noisy with young children. Furthermore, as Linn and Petersen (1985) pointed out, even using the same tests of spatial ability does not ensure that the same underlying processes are being tapped at different developmental time points.

An interesting and potentially important pattern of results from the present study is the lack of a difference in the magnitude of sex differences on rotation versus translation items from the spatial transformation task as well as the lack of a difference in the magnitude of sex differences on the spatial transformation task versus the WPPSI-R Mazes task. These findings raise the possibility that sex differences in spatial skill during the preschool years are no more robust for spatial tasks involving mental rotation than for other types of spatial tasks, a pattern that differs from that which has been reported in adults (e.g., Linn & Petersen, 1985; Voyer et al., 1995). It is of course possible that there is a mental rotation task other than the rotation items on our spatial transformation task on which preschool children would show the adult pattern of a differentially large sex difference. A hint that this may be the case is provided by Uttal et al.’s (1999) finding that the sex difference in favor of 5-year-old boys was particularly large on items on which the map was rotated with respect to the real-world space. Alternatively, it is possible that the more robust sex difference reported on mental rotation tasks in adults may actually extend to other types of spatial tasks, such as those involving nonrotational transformations.
Experiment 2

We performed a control experiment to determine whether the effect of test half found in Experiment 1 was attributable to practice or to the chance placement of easier items on the second half of the task.

Method

Participants. An additional group of 19 children (9 boys and 10 girls), aged 5 years to 5 years 5 months, was tested on the spatial transformation task in order to disambiguate the finding of significantly better performance on the second half of test items in Experiment 1. This age group was chosen because it was the youngest age group in Experiment 1 to show a significant test-half effect and because it was readily available to us. The children were students in parochial schools in the Chicago area.

Materials. Materials for the spatial transformation task were identical to those used in Experiment 1.

Design and procedure. Participants were given the spatial transformation task from Experiment 1. In Experiment 2, however, the first half of trials in Experiment 1 was administered second and the second half of trials in Experiment 1 was administered first. Thus, this new group of participants received Items 17–32 first (as numbered in Experiment 1), followed by Items 1–16 (as numbered in Experiment 1). All other variables were counterbalanced across participants as in the original sample. Participants in Experiment 2 were administered only the spatial transformation task and were seen for only one test session.

Results

An ANOVA revealed a significant main effect of test half, $F(1, 17) = 13.44, p < .002$, reflecting the finding that these children, like those in Experiment 1, performed better on the second half of items. Also similar to the results of Experiment 1, the main effects of problem type, $F(3, 51) = 4.24, p < .01$, and target shape, $F(1, 17) = 48.39, p < .01$, were significant, and the main effect of sex was marginally significant, $F(1, 17) = 3.44, p < .08$.

A second ANOVA with sex and group (Experiment 1: children aged 5 years to 5 years 5 months; Experiment 2: children aged 5 years to 5 years 5 months) as between-subjects variables and problem type, target shape, and test half (first half, second half) as within-subjects variables revealed a main effect of test half, $F(1, 64) = 20.01, p < .0001$, and no Group × Test Half interaction ($F < 1, p > .10$, see Figure 5). Thus, both groups performed better on the second half of trials when the order of item administration was counterbalanced across halves: In Experiment 1, for children aged 5 years to 5 years 5 months, first half $M = 7.77$ ($SD = 4.69$) and second half $M = 9.08$ ($SD = 4.76$); in Experiment 2, for children aged 5 years to 5 years 5 months, first half $M = 7.16$ ($SD = 4.63$) and second half $M = 9.05$ ($SD = 4.20$). These results indicated that practice, rather than differential difficulty of the first and second half of items, was responsible for the better performance on the second half of test items. The main effects of sex, $F(1, 64) = 8.32, p < .005$, problem type, $F(3, 192) = 5.76, p < .001$, and target shape, $F(1, 64) = 107.62, p < .0001$, were significant, as was the case in Experiment 1.

Discussion

The findings of the control experiment indicate that the higher performance of children aged 5 years and over on the second half of trials of the spatial transformation task is attributable to practice. The comparability of this improvement for boys and girls is consistent with a meta-analysis of training studies that showed that training benefits males' and females' performance on spatial tasks to an equal extent (Baenninger & Newcombe, 1989). Further research is needed to (a) identify the types of input that occur in natural settings, such as school and home, that are most effective in enhancing particular spatial skills and (b) determine whether the effectiveness of such input differs for males and females.

General Discussion

The present finding of sex differences in spatial skill during the preschool years focuses our attention on early rather than late correlates of these cognitive differences. Existing research has suggested that both biological differences and input differences are related to sex differences in spatial skill. Although none of these differences has been shown to be causally related to sex differences in spatial skill, we briefly discuss early biological and input differences that have been shown to be related to spatial skill differences. Biological and input variables related to sex differences in spatial skill are most frequently examined separately in the literature, even though these classes of variables are richly intertwined. A clear demonstration of this relation was provided by Casey and Brabeck's (1990) study, which showed that the spatial skills of females with non-right-handed relatives are more likely to benefit from spatially relevant experiences than are the spatial skills of females whose relatives are right-handed. These findings illustrate the interaction of biological and input variables in the development of spatial skills.

A variety of interrelated biological variables, notably differences in gonadal hormones and differences in regional brain maturation, have been implicated in the early emergence of sex differences in cognition (e.g., Bachevalier & Hagger, 1991; Goy & McEwen, 1980; Levy & Heller, 1992; Williams & Meck, 1991). Gonadal hormone levels have been related to the development of spatial skill through studies of humans with hormonal abnormalities (see review by Levy & Heller, 1992). For example, males...
with idiopathic hypogonadotropic hypogonadism who have andro-
genesis deficiency early in life have low spatial ability compared
with normal males and with males who developed androgen de-
cency during puberty (Hier & Crowley, 1982). Furthermore, fe-
male with congenital adrenal hyperplasia who have high andro-
gen levels during prenatal development and early in life have high
spatial ability compared with normal controls (Berenbaum, 1993;
Hampson, Rovet, & Altman, 1998; Resnick, Berenbaum, Gottes-
man, & Bouchard, 1986). A significant association between level of
masculinizing hormones and spatial skill has not been found in
“normal” young children, but it is possible that the measures of
both hormones and spatial skill were not sufficiently sensitive to
detect a relation (Finegan, Niccols, & Sitarenios, 1992; Jacklin,
Wilcox, & Maccoby, 1988). It should be noted that certain studies
suggest that genetic variables, rather than hormonal differences per-
se, are related to sex differences in these hormonally abnormal
populations (e.g., Arnold, 1996; Nass & Baker, 1991).

A variety of studies provide evidence of sex differences in brain
development early in life. Most of these studies support earlier
development of the right hemisphere in males than females (e.g.,
de Lacoste, Horvath, & Woodward, 1991; Diamond, Dowling, &
Johnson, 1981; Gratton, De Vos, Levy, & McClintock, 1992;
Shucard, Shucard, Cummins, & Campos, 1981; Shucard, Shucard,
& Thomas, 1984). Both anatomical and functional evidence of
faster right hemisphere development in male than female infants
has been reported. In a study of human fetal brains, de Lacoste et
al. (1991) reported greater volume of the right hemisphere in males
and either no asymmetry or a slight left hemisphere volumetric
advantage in females. Male infants also are reported to show more
mature auditory-evoked responses over the right than the left
hemisphere. In contrast, female infants transiently show more
mature auditory-evoked responses over the left hemisphere (e.g.,
Shucard et al., 1981; Shucard et al., 1984). Asymmetries in motor
functioning also support earlier maturation of right-sided brain
structures in male than female infants. In particular, newborn
males are reported to show greater leftward lower limb reflexes
than newborn females (Gratton et al., 1992). Because the right
hemisphere has been shown to be more involved than the left
hemisphere in many spatial processing tasks (e.g., Levine, Banich,
& Koch-Weser, 1988; Levy & Reid, 1978; Sergent, Ohta, &
MacDonald, 1992; Warrington & Taylor, 1973), this evidence of
earlier right hemisphere development in males has been hypothe-
sitized to be related to the spatial skill advantage in males. However,
at the present time, this connection is based on conjecture rather
than on evidence.

Other studies have focused on the possible relation of sex
differences in environmental input to sex differences in spatial
skill. Existing studies report differences in spatially relevant input
to girls and boys at school and at home. At school, preschool
teachers are reported to spend more time with boys than girls and
usually interact with boys in the block, construction, sand play, and
climbing areas and with girls in the dramatic play area (Ebbeck,
1984). At home, preschool boys are reported to more frequently
engage in spatial activities than preschool girls, both alone and in
conjunction with their caregivers (Newcombe & Sanderson, 1993).
Boys have greater access to so-called male toys (e.g., Lincoln
Logs, Legos) than girls, and this accounts for at least a portion of
the sex difference in performance on the Block Design subtest of
the Wechsler Intelligence Scale for Children (Serbin, Zelkowitz,
References


Shepard, R. N. (1975). Form, formation, and transformation of internal

Received March 11, 1997
Revision received November 11, 1998
Accepted November 11, 1998

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the American Psychologist and APA Monitor. High School Teacher and International Affiliates receive subscriptions to the APA Monitor, and they may subscribe to the American Psychologist at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the Publication Manual of the American Psychological Association, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.