

The locus of sex differences in spatial ability

ROBERT KAIL, PHILIP CARTER, and JAMES PELLEGRINO
University of Pittsburgh, Pittsburgh, Pennsylvania 15260

College men and women judged whether pairs of stimuli were identical or mirror images. One stimulus of a pair was presented upright; the other was rotated 0° - 150° from the vertical. The stimuli were either alphanumeric symbols or unfamiliar letter-like characters of the type found on the Primary Mental Abilities Spatial Relations Test. For each individual, the linear function relating response latency to degree of rotation was computed. The slope of this function was steeper for women than for men. Further, the distribution of slopes was more variable among women, with approximately 30% falling outside the range of distribution for men. Women and men were quite similar in the accuracy of their judgments, the intercepts of the latency functions, and the precision with which the linear function characterized the latency data. It is suggested that the sex difference in the slope of the rotation function may reflect differences in strategies of mental rotation.

Spatial ability is one of the few aspects of cognitive functioning in which the evidence consistently indicates a sex difference in performance: When a task involves the ability to imagine what an object will look like when it is moved or rotated in space, males' performance generally surpasses that of females (Maccoby & Jacklin, 1974). Despite the consistency with which this sex difference is found, there is surprisingly little evidence regarding the processes underlying the difference. That is, solution of problems like those appearing on tests of spatial ability involves a series of component processes; sex differences in performance on spatial tasks might be attributable to any of these processes.

Consider, for example, the Spatial Relations Test from the Primary Mental Abilities (PMA) battery (Thurstone, 1958). A letter-like stimulus is presented with six rotated versions of the stimulus. Some of these comparison stimuli, when rotated in the picture plane, would be identical to the standard, while others would be mirror images of it. Subjects are asked to select those comparison stimuli that would be identical to the standard when rotated. Males typically solve these problems more accurately and more rapidly than females (Thurstone, 1958).

Solution of this kind of problem seems to involve four general processes (Cooper & Shepard, 1973).

This research was supported by grants to the first author from Sigma Xi and from the Faculty of Arts and Sciences of the University of Pittsburgh. It was also supported by the Learning Research and Development Center of the University of Pittsburgh, which is funded in part by the National Institute of Education, U.S. Department of Health, Education and Welfare. We are grateful to Jeff Bisanz for his helpful comments on a previous draft of this paper, and to Jack Joseph, Rose LaSpada, Barbara Pazak, Teresa Smith, Carol Vesely, Mary Welsh, and Regina Wessel for their help in testing subjects and in scoring the data. Requests for reprints should be sent to Robert Kail, Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907.

First, an individual encodes the target and comparison stimuli. Second, s/he mentally rotates the comparison stimulus into congruence with the standard. Third, s/he compares the target with the rotated representation of the comparison stimulus. Fourth, s/he responds. Sex differences in performance on the PMA spatial test might reflect differences between the sexes in any or all of these processes.

The present research was conducted to identify information processing parameters that might be related to sex differences in performance on spatial tasks. The task was derived from the spatial test of the PMA. Adults were asked to discriminate letters and letter-like forms from their mirror images. Two versions of a stimulus were shown on a trial. The standard, on the left, was presented in the normal, upright orientation. The comparison stimulus, on the right, was rotated 0° - 150° from the standard. Subjects determined if the standard and comparison stimuli would be identical or mirror images when the comparison stimulus was rotated mentally.

From latencies on these judgments, we can estimate the amount of time required to execute the component processes described above. Specifically, the slope of the function relating response time to degree of stimulus rotation reflects the rate of mental rotation, while the intercept of this function reflects the times to (a) encode the stimuli, (b) compare the rotated stimulus, and (c) respond (Cooper & Shepard, 1973).

Thus, one purpose of our experiment was to determine if sex differences would be found in the slope of the latency function, the intercept, or in both parameters. In other words, one goal of the present research was to examine encoding, rotation, comparison, and response processes as possible sources of sex differences in one form of psycho-

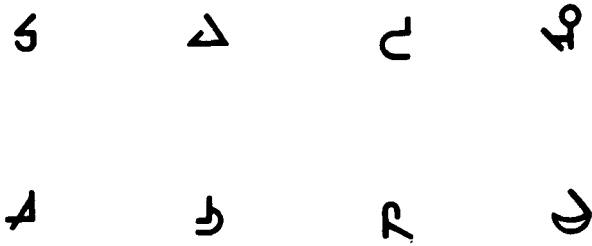


Figure 1. Letter-like characters from the Spatial Relations Test of the PMA.

metrically measured spatial ability. In addition, we were interested in the possibility that a sex difference in any of these components might be specific to a particular class of stimuli. Notably, the stimuli in the PMA, like those on many tests of spatial ability, are unfamiliar and do not have readily available verbal labels (see Figure 1). Given females' superiority on a host of verbal tasks (Maccoby & Jacklin, 1974), it seemed possible that the use of highly familiar verbal stimuli might considerably reduce the sex difference in spatial performance, or perhaps eliminate it altogether. To evaluate this notion, subjects judged pairs of unfamiliar stimuli like those in Figure 1 on half of the trials and familiar, alphanumeric stimuli on the remaining half.

METHOD

Participants

One hundred and four undergraduates at the University of Pittsburgh, 51 men and 53 women, participated to satisfy a course requirement.

Material

The stimuli were eight asymmetrical alphanumeric symbols (4, 5, F, G, J, L, P, R) and eight characters from the PMA (shown in Figure 1). Nine slides were prepared for each stimulus. In six, the standard and comparison were identical. The comparison was depicted in a different orientation on each, rotated 0°, 30°, 60°, 90°, 120°, or 150° from the standard. In the remaining three slides, the standard and comparison stimuli were mirror images. Each mirror-image comparison stimulus was presented in three different orientations selected arbitrarily from the set of 0°-150°, such that across all eight stimuli within a category, the six orientations were represented equally.

The resulting set of 144 slides thus consisted of 96 slides in which the stimuli were identical and 48 in which they were mirror images.¹ The 96 slides depicting identical pairs, in turn, consisted of 6 degrees of rotation (0, 30, 60, 90, 120, 150) × 2 types of stimuli (alphanumeric, PMA characters) × 8 replications. The 48 mirror-image slides consisted of 6 degrees of rotation × 2 types of stimuli × 4 replications. Slides were ordered randomly with the constraints that (a) pairs of identical stimuli corresponding to each combination of degree of rotation and type of stimulus appeared once in every block of 18 slides, and (b) pairs of mirror image stimuli corresponding to each combination of degree of rotation and type of stimulus appeared once within each block of 36 slides.

Procedure

Subjects were asked to judge when stimuli were identical and when they were mirror images. Several pairs of each type were

presented to demonstrate the distinction and to show how stimuli could be rotated ("turned in your head") to determine the correct response. Twelve practice trials were followed by the test series of 144 slides.

On each trial, a slide was projected onto a screen approximately 1 m from the subject. Presentation of the slide activated a photocell, which started a Lafayette clockcounter. Subjects responded by pressing one of two buttons with the index finger of the preferred hand, which stopped the clock. The next trial began immediately after the experimenter had recorded the response time. Testing required approximately 15 min.

RESULTS

For each combination of degree of rotation, type of stimulus, and type of response, a median latency was calculated from an individual's correct responses. Shown in Figure 2 are means of these medians, plotted as a function of degree of rotation. The result of interest in Figure 2 is that while response latency consistently increases as a function of degree of rotation, the increase seems to be greater for women than for men, particularly on the PMA characters. This pattern of results is seen for both "same" and "different" judgments. However, the functions for "different" judgements are less linear than those obtained for "same" judgments, particularly when functions are calculated for individuals. Consequently, in the remainder of this paper, we will consider only data from "same" judgments.

In all analyses reported here, effects are significant at $p < .01$, unless indicated otherwise.

Errors

The error rates for men and women, 6% and 5%, did not differ significantly ($F = 1.56$). Furthermore, all interactions involving sex were nonsignificant ($F_s \leq 1$). Subjects made fewer errors on alphanumerics (3%) than on PMA characters (8%) [$F(1,102)$

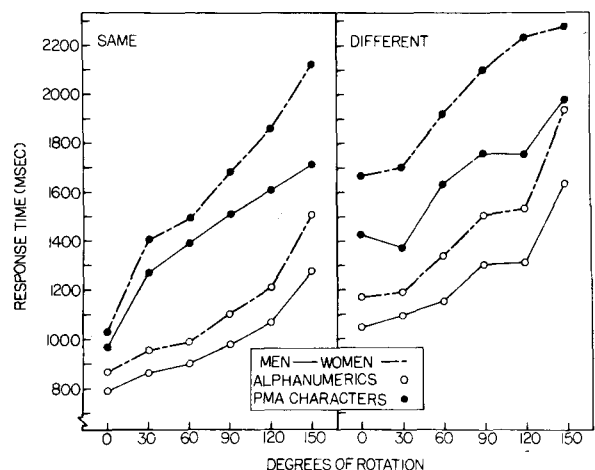


Figure 2. Response time in milliseconds as a function of the degree of rotation, for "same" judgments (left panel) and "different" judgments (right panel).

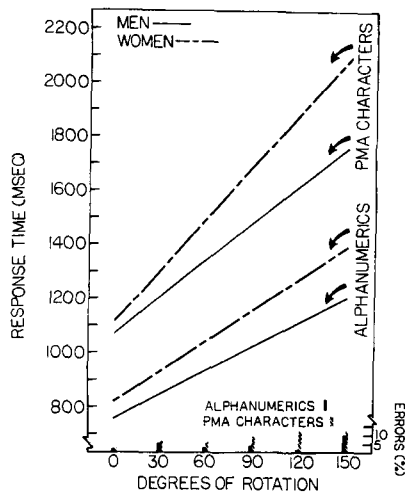


Figure 3. Average functions relating response time in milliseconds to degree of rotation (left ordinate) and errors as a function of degree of rotation (right ordinate).

= 85.63, $MSe = .52$]. Further, errors were related to increases in the degree of rotation (1%, 4%, 4%, 5%, 7%, 11%) [$F(5,510) = 39.59$, $MSe = .39$]. Finally, the interaction of type of stimulus and degree of rotation was also significant [$F(5,510) = 8.65$, $MSe = .33$], reflecting the fact that judgments of alphanumerics were less affected by degree of rotation than were the PMA characters (see Figure 3). Most important, however, is the high degree of similarity found for males' and females' ability to make the correct judgments.

Response Times

The linear function relating response time to degree of rotation was computed for each individual, separately for alphanumerics and PMA characters. Mean functions are shown in Figure 3. For men, equations relating response time (RT) to degree of rotation were $RT = 2.93X + 760$ and $RT = 4.61X + 1,064$ for alphanumeric and PMA characters, respectively. Corresponding equations for women were $RT = 3.88X + 814$ and $RT = 6.66X + 1,101$ (where X corresponds to degrees of rotation).

r^2 . To assess the extent to which the linear functions described subjects' response latencies, r^2 was calculated for each subject. The overall mean was .73, indicating a reasonably accurate fit to the linear model. A 2 (sex) by 2 (type of stimulus) ANOVA revealed no significant main effects, but a significant interaction [$F(1,102) = 7.69$, $MSe = .0327$]. For males, latencies from alphanumeric characters were fit slightly more accurately by the linear model than were PMA characters (r^2 s of .75 and .72), while the reverse was true for females (r^2 s of .66 and .77). Perhaps the most important finding, however, is that for both types of stimuli and both sexes, the linear

model accounts for a large proportion of the variability in subjects' response times.²

Slopes. The slopes of the functions depicted in Figure 3 provide estimates of the rate of mental rotation. Analyses of these slopes revealed that men rotated stimuli more rapidly than women [$F(1,102) = 6.20$, $MSe = 18.97$] and that alphanumerics were rotated more rapidly than PMA characters [$F(1,102) = 62.03$, $MSe = 4.17$]. Finally, the Sex by Type of Stimulus interaction was marginally significant [$F(1,102) = 3.82$, $p = .053$]. Men rotated alphanumerics .95 msec/deg more rapidly than women and rotated PMA characters 2.05 msec/deg more rapidly.

Intercepts. As described in the introduction, the intercept presumably reflects the times necessary to (a) encode the pairs, (b) compare them, and (c) respond. Assuming that response time (i.e., c) is constant across the two types of stimuli, differences between intercepts reflect differences in the encoding and comparison stages. In fact, intercepts for PMA characters were much larger (1,048 msec) than those for alphanumerics (726 msec) [$F(1,102) = 278.87$, $MSe = 16,312.04$]. Alphanumeric stimuli are encoded and compared much more readily than the PMA characters. Of primary interest, however, is the finding that all effects involving sex were nonsignificant ($F_s \leq 1.11$). Thus, the sexes apparently do not differ in the rate with which they encode and compare stimuli and respond.

Variability within Sex

To this point, we have focused exclusively on differences between the sexes in mean level of performance. However, analyses of individual variation within each sex reveal an interesting pattern of differences between males and females. Frequency distributions of slopes and intercepts for both alphanumeric and PMA characters are shown in Figure 4. On the intercepts, the variability among men and women was similar. However, on the slopes, the range of individual differences was greater for women than for men. This difference in variability was significant for both alphanumeric characters [$F \max(2,102) = 4.66$] and PMA characters [$F \max(2,102) = 5.41$].

The shapes of the frequency distributions for slopes also reveal interesting differences between males and females. For men, the distributions of both alphanumeric and PMA characters are generally normal with some positive skew. For women, the distribution of alphanumerics differs significantly from the normal distribution [$\chi^2(10) = 34.72$], due to considerable positive skew. A similar pattern, in extreme form, is found in women's distribution of slopes for PMA characters. Again, the distribution differs significantly from normality [$\chi^2(10) =$

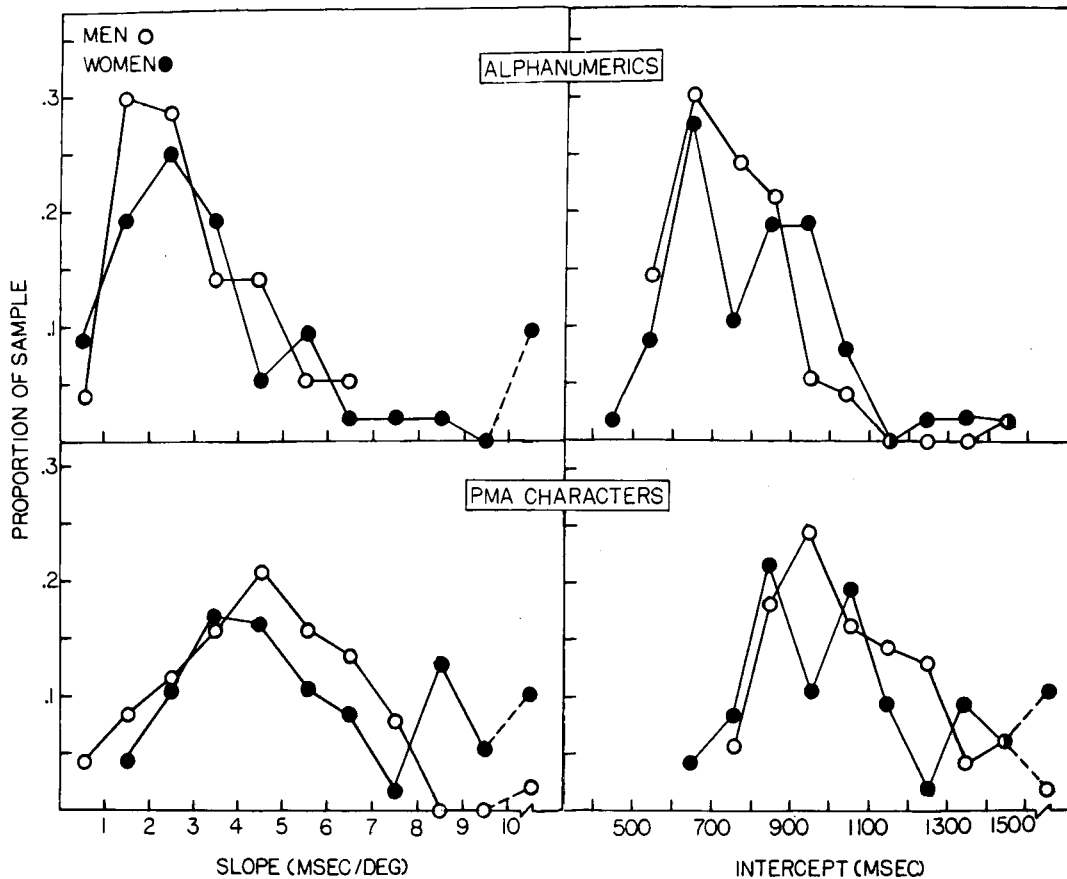


Figure 4. Frequency distributions of slopes and intercepts on alphanumeric and PMA characters, plotted separately for men and women.

38.18]. Further, the distribution is clearly bimodal. One mode, 3.5 msec/deg, approximates the male mode of 4.5 msec/deg. The second mode, 8.5 msec/deg, is essentially outside the range of the male distribution (only one male had a slope greater than 8.5 msec/deg). Stated another way, 37 of 53 women (70%) have slopes in the range of the male distribution, while 16 women (30%) fall outside of that distribution.

DISCUSSION

The pattern of results from the present study helps to specify the locus of sex differences in one spatial task. Men and women were quite similar in (a) their use of mental rotation to judge pairs of stimuli, (b) the accuracy with which they judged pairs, (c) the precision with which the linear model fit their latency data, and (d) the times needed to encode and compare pairs of familiar and unfamiliar stimuli. The single significant sex difference in performance was that men mentally rotated stimuli more rapidly than women. This sex difference in means was, in turn, attributable to greater variability among women. About 70% of the women

mentally rotated stimuli at rates comparable to men, while 30% of the women did so at a slower rate.

The most important question posed by these data concerns the factor(s) underlying the comparatively slow rate of mental rotation by a sizeable minority of women. One possibility is that these women used a different strategy of mental rotation than the rest of the sample. That is, two general strategies of mental rotation can be distinguished. In "holistic" mental rotation, the entire comparison stimulus is rotated simultaneously into congruence with the target. A second strategy is one in which the individual features of the comparison stimulus are rotated separately. That is, the mental rotation process is applied iteratively until all features of a stimulus have been rotated. If rate of rotation is constant, regardless of the complexity of the information being rotated (i.e., a single feature vs. an entire stimulus), then use of the feature strategy would account for the steep slopes found in 30% of the women in our sample.

One aspect of our data suggest that this interpretation is at least plausible. If features and complete stimuli are rotated at the same rate, then multiplying the slope for subjects using holistic rotation

by some integer (representing the number of different features rotated) should yield the slope for subjects using the feature rotation strategy. In fact, the women in question rotated PMA characters approximately twice as slowly as the rest of the sample (8.5 msec/deg vs. 4 msec/deg), suggesting that these women may have rotated two separate features of the PMA characters.

The intercepts of men and women were consistently similar, in both average performance and variability. However, since the intercept is a composite measure reflecting three separate processes, a sex difference in only one of these processes might not be evident in our data. More fine-grained analyses will be necessary to determine the reliability of our findings concerning the absence of sex differences in encoding, comparison, and response processes.

In conclusion, the results of the present study demonstrate two important points regarding the nature of sex differences in spatial ability. First, the sexes were strikingly similar in most aspects of their performance on this spatial task. Males' superiority on at least this spatial task seems to be linked to a single component of processing. Second, sex differences in spatial ability are generally viewed as differences between the sexes in terms of average level of spatial ability. From the present results, it appears that the important difference between the sexes is not the average level of ability, but in the variability within each sex.

REFERENCES

COOPER, L. A., & SHEPARD, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

MACCOBY, E. E., & JACKLIN, C. N. *The psychology of sex differences*. Stanford: Stanford University Press, 1974.
THURSTONE, T. G. *Manual for the SRA Primary Mental Abilities*. Chicago: Science Research Associates, 1958.

NOTES

1. The ratio of identical pairs to mirror image pairs was 2:1. We chose this rather than a 1:1 ratio for the following reason. Portions of these data were collected in the context of a developmental study of mental rotation in which young children would cooperate for about 20 min before they tired of the task. This limited us to approximately 150 trials. As described in the Results section, a median RT was computed for each combination of degree of rotation and type of stimulus. Using a 1:1 ratio with 144 trials, each of these medians would have been based on a maximum of six responses. Since the functions relating RT to degree of rotation are virtually identical for the two types of responses (e.g., Cooper & Shepard, 1973), a 2:1 ratio was used so that half of the medians (those for "same" judgments) would be based on a slightly larger pool of responses (eight rather than six).

Another possibility to achieve a balanced 1:1 ratio would have been to extend testing over several sessions. We chose not to do this because of our interest in comparing these results with typical performance on the PMA and comparable psychometric tests of spatial ability, in which testing occurs in a single brief session.

2. Our interpretation of slopes and intercepts depends on the strength of the linear relation between response latency and degree of rotation. While the average fit was quite good for both males and females, it is also the case that within each sex, individuals differed in the extent to which the linear model adequately described their data. One could argue that we can be confident that mental rotation has occurred *only* when an individual's linear fit is significant. Consequently, we analyzed the data separately for the 41 men and 39 women who had significant r^2 values (i.e., $r^2 \geq .53$) on both types of stimuli. The results were identical to those reported here for the entire sample.

(Received for publication February 12, 1979;
revision accepted June 18, 1979.)