CrossMark

Ability and sex differences in spatial thinking: What does the mental rotation test really measure?

Mary Hegarty[1]

Abstract Spatial ability tests are often interpreted as measuring facility with imagined spatial transformations of objects. But some spatial ability tests can be solved by analytic strategies as well as imagery transformation strategies. In the present study, participants gave verbal protocols while completing items on the Vandenberg and Kuse (*Perceptual & Motor Skills, 4,* 599–604, 1978) mental rotation test, and/or reported the strategies they had used on the test. Most participants used both imagery transformation and analytic strategies (i.e., feature-based, orientation-independent strategies) to solve the test items. Use of one analytic strategy, the global-shape strategy, was positively correlated with accuracy. Specifically, some of the most successful students used this strategy to eliminate answer choices, reducing the need for mental imagery. Men outperformed women, as is typical on this test, and were more likely than women to use the global-shape strategy, in particular, and more holistic strategies, in general. These results argue against the mental rotation test as a measure of spatial imagery alone and suggest that the ability to discover and use more efficient analytic strategies may be an important additional component of what this test measures.

In recent years, interest in individual differences in spatial thinking has increased. We now know that spatial abilities predict success in science, technology, engineering, and mathematics (STEM) education and careers, independently of verbal and mathematical abilities (Wai, Lubinski, & Benbow, 2009), and that spatial skills can be trained (Uttal et al., 2013). But if we are to understand how to foster the development of spatial thinking skills, it is important to identify the fundamental differences between individuals who perform well and poorly on tests of spatial abilities and skills.

Here, I focus on individual differences in performance on the mental rotation test (Vandenberg & Kuse, 1978), which I will refer to as the "MRT." As a task, mental rotation has been important in demonstrating the functional significance of mental imagery in thinking (Shepard & Metzler, 1971). It is also an important task in psychometric studies of intelligence, as one of the most commonly used measures of spatial ability, which is one of the major dimensions of intelligence (Carroll, 1993). Mental rotation is also notable as the cognitive ability with the largest documented sex difference in favor of men. Meta-analyses have indicated a large sex difference (between 0.7 and 1 standard deviation) in the MRT (Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995).

Following classic laboratory mental rotation tasks (Shepard & Metzler, 1971), the MRT is usually assumed to measure the ability to mentally manipulate images. This raises questions about the fundamental differences between individuals who perform at different levels on this test. One possibility is that performance on the MRT reflects differences in visual spatial working memory capacity (Shah & Miyake, 1996). Specifically, people with more working memory might be able to maintain more complex visual spatial images and maintain the integrity of these representations under mental transformations such as rotation (Just & Carpenter, 1985; Shah & Miyake, 1996). But, is image rotation the only cognitive process measured by the MRT, and can differences in working memory capacity explain all of the variance in this test? In one recent study, spatial working memory differences accounted for performance

✉ Mary Hegarty
mary.hegarty@psych.ucsb.edu

1   Department of Psychological & Brain Sciences, University of California Santa Barbara, Santa Barbara, CA 93106-9660, USA

in various spatial ability tests but were not sufficient to explain all of the variance in the MRT results (Kaufman, 2007). Moreover, sex differences in working memory tend to be much smaller than those observed in the MRT, suggesting that working memory is unlikely to be the only cause of the sex difference (Voyer, Voyer, & Saint-Aubin, 2017).

Besides mental capacity, another possibility is that performance differences in the MRT reflect strategy differences. A number of studies have shown that people use a variety of strategies on spatial ability tests (Hegarty & Waller, 2005; Lohman, 1988) and laboratory mental rotation tasks (Bethell-Fox & Shepard, 1988; Folk & Luce, 1987; Just & Carpenter, 1985; Takano, 1989; Yuille & Steiger, 1982). Besides mental rotation of holistic images, these strategies include piecemeal rotation, perspective taking, and feature-based, viewpoint-independent strategies.

Importantly, critical differences between the MRT and laboratory mental rotation tasks make the MRT open to additional strategies that are not available for the laboratory task. In the MRT, participants view a depiction of a three-dimensional target figure and four test figures, as is shown in Fig. 1. Their task is to determine which two of the four test figures show the same object as the target. In contrast, the Shepard and Metzler (1971) laboratory task asks solvers to make same–different judgments about pairs of shapes. In the psychometric test, participants are given a time limit of 3 min to complete ten items, whereas in the laboratory task, each item is individually timed. Critically, in the laboratory task the foils are always mirror images of the target object (*mirror foils*), whereas in the psychometric test, half of the foils are mirror images of the standard, and the others are structurally different objects (*structure foils*). Structure foils can be eliminated on

the basis of orientation-independent aspects of shape, so that mental rotation is not necessary (Jolicœur, 1990; Takano, 1989). For example, in the item in Fig. 1b, the foils can be eliminated by noticing that the two terminal arms of the standard figure are parallel to each other, whereas in the foils they are perpendicular.

Previous research on performance factors in the MRT has focused on manipulating the test instructions or examining response patterns to understand sex differences in this task. For example, giving participants a more generous time limit reduces but does not eliminate the sex difference (Voyer, 2011). Researchers have also examined effects of the type of foil (mirror vs. structure) on MRT performance. Some studies have shown no difference between the sizes of the sex differences for these two types of items (Voyer & Doyle, 2010; Voyer & Hou, 2006), although these studies did not impose the usual time limits. In a study that focused on individual differences more generally, Geiser, Lehmann, and Eid (2006) identified a group of otherwise low-performing participants who performed well on items with structure foils, who they referred to as *nonrotators*. Hegarty (2010) had students report their strategies after they had completed the MRT, and found that reporting the orientation-independent strategy was associated with better performance.

Study 1

The present Study 1 focused on analyses of the Vandenberg and Kuse (1978) test based on concurrent and retrospective verbal protocols. Concurrent verbal protocols reveal aspects of performance that cannot be studied by manipulation of the
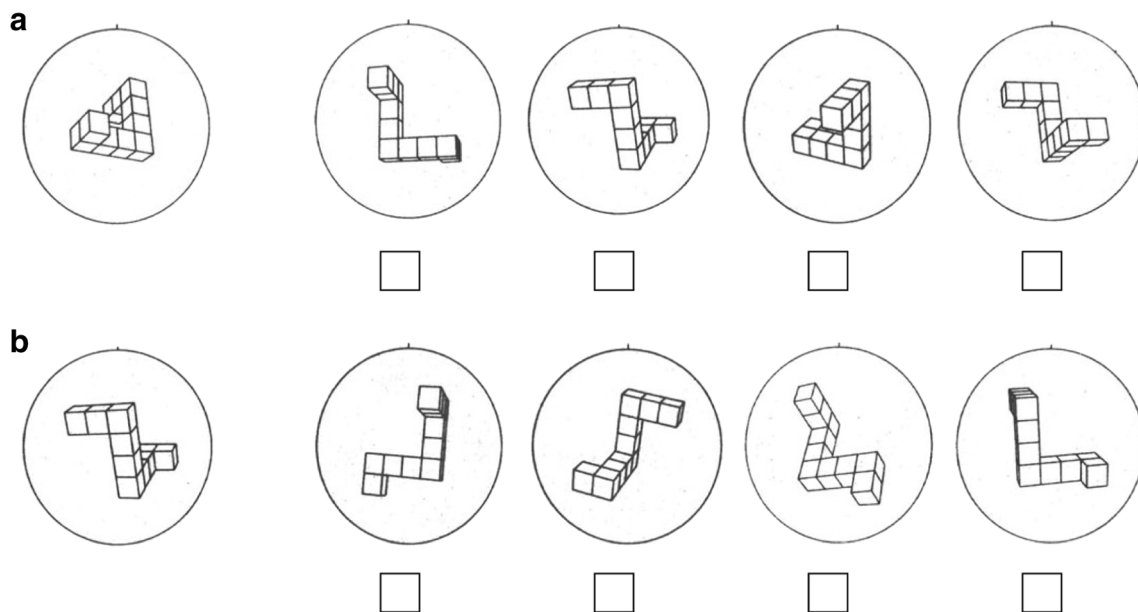

Fig. 1 Sample items from the Vandenberg and Kuse (1978) mental rotation test (MRT). For item A, the foils are mirror images. For item B, the foils are structurally different.

instructions or test-taking conditions. For example, such protocols can reveal which stimulus features individuals attend to when solving test items, discriminate between imagery and other strategies, and reveal whether people use multiple strategies. The goal of this study was to identify the full range of strategies used in this test, sex differences in strategy use, and the relationships between test performance and strategy use.

Method

Participants The participants were 47 undergraduate students (24 male, 23 female), recruited from undergraduate psychology and chemistry classes, who received either course credit or payment for participation. The age range of the students in these classes is 18–22 years. An a priori power analysis determined that this sample size would allow the detection of a sex difference of 0.75 (which is typical on this test; Voyer et al., 1995) with .8 power. All participants were treated in accordance with APA ethical guidelines in both studies reported.

Materials The data were collected in the context of a broader study on strategies in spatial ability tests. Participants completed three psychometric tests: the paper-folding test (Eckstrom, French, Harman, & Dermen, 1976), the MRT (Vandenberg & Kuse, 1978), and a perspective-taking test.[1] This article focuses on performance on the MRT, which is composed of two parts with ten items each (see the examples in Fig. 1). The items in Part 1 of the test were presented in their usual format (four problems to a page). The items in Part 2 were printed on individual pages for the protocol trials. They were covered by plastic sheets so that participants could not make any marks on the test, and the four answer choices for each item were labeled A, B, C, and D.

Procedure After completing ten trials of the paper-folding test, participants were given instructions for the MRT and were allowed 3 min to complete Part 1 of this test (ten items), followed by the perspective-taking test. Then they were given instructions on how to give a concurrent verbal protocol. These instructions (see Ericsson & Simon, 1984, p. 378) asked participants to report everything that came to mind from when they first saw the problem to when they answered, and included "warm-up" exercises in which they practiced thinking aloud on a mental arithmetic problem and a mental imagery task ("How many windows are in your parent's house?"). Next, participants solved ten items of the paper-folding test

---

[1] Thirteen participants completed the spatial orientation test (Hegarty & Waller, 2004), and the remaining participants completed an adapted version of the visualization of viewpoints test (Guay & McDaniels, 1976). Both of these perspective-taking tests and the paper-folding tests were solved by a variety of strategies, but no systematic sex differences emerged in the strategies used on these tests, and no specific strategies were reliably correlated with success.

while giving a concurrent protocol. Then they were shown the items in Part 2 of the MRT, one at a time, and were instructed to think aloud while solving the items. They gave their answers by saying the labels (A, B, C, D) of their answer choices aloud. After completing these ten items, they provided a retrospective protocol—that is, they reported what they had been thinking while solving the items, including any strategies they had used. A videotape captured their verbalizations and gestures during the concurrent and retrospective protocols.

Coding The MRT was scored out of 20, with each item scored as correct if both correct answer choices were marked, and scored as incorrect otherwise. Two mental imagery strategies (mental rotation and perspective taking) and three feature-based, orientation-independent strategies (counting cubes, local turns, and global shape) were identified. *Mental rotation* was coded when participants described "rotating," "turning," or "flipping" the objects, gestured picking up the object and turning it, or used their hand to represent and rotate the object (cf. Chu & Kita, 2011). *Perspective taking* was coded if the participant described imagining looking at the object from a different orientation or mapping their whole body to the object. *Counting cubes* was coded if the participant counted the number of cubes in the different arms of the object. *Local turns* was coded if participants used directional terms (up, to the right, toward me, etc.) to refer to the relative directions of the arms of the figure before any attempt to rotate it. Finally, *global shape* was coded if participants eliminated answer choices on the basis of whether the two terminal arms of the figure were parallel or perpendicular or whether or not three sides of the object formed a plane. Table 1 gives examples of

Table 1 Examples of retrospective strategy reports for each of the five spatial strategies used in the mental rotation test

---

Mental Rotation:
"I just . . . uh . . . go number by number and see if I can **rotate** it in my head this one to be superimposed on top of the other ones . . ."

Perspective Taking:
"I just like . . . imagine myself like standing . . . and then imagining like . . . a piece on the ground next to me, and then like looking up in which direction does it . . . seem to go toward . . ."

Counting Cubes:
"I counted . . . the . . . numbers . . . of the squares . . . Because it was always either three, three, four, three . . . or three, four, three, three . . ."

Local Turns:
"for example like I was saying this has like three cubes going down, and then there were four going to the left, and then three going down, and then three going . . . more forward toward me . . . So I would look . . . for like that sort of diagram in the other answers."

Global Shapes:
"Well the first thing I'd do is, like . . . figure out how many things were sticking off in, like . . . other directions . . . from each other. . . . So on this one, it's like . . . all **planar** except for one thing that sticks off . . . but, uh . . . on this one, you can't eliminate any, but on some of 'em you can eliminate from that . . ."

the retrospective protocols of participants who were coded as using each strategy.

Concurrent protocols for each trial were coded for evidence of each of the five strategies (the measure for each strategy was the number of trials out of ten on which the strategy was observed; three participants' concurrent protocols could not be reliably coded and are not included in analyses of these measures). Each participant's retrospective protocol was coded as either reporting or not reporting each of the five strategies. Two raters independently scored the concurrent protocols of 14 participants and the retrospective protocols of 33 participants; interrater agreement was 93.8% for the concurrent protocols and 93.9% for the retrospective protocols. Disagreements were resolved by the author.

Results

The mean score on the MRT was 12.12, $SE = 0.59$, 95% CI [10.93, 13.32], out of a maximum of 20 (chance performance is 3.33, assuming that participants mark two answer choices for each item). Not surprisingly, participants solved more items on Part 2 of the test ($M = 7.30$, $SE = 0.30$, 95% CI [6.70, 7.90]) than on Part 1 ($M = 4.83$, $SE = 0.40$, 95% CI [4.02, 5.63]), $t(46) = 6.44$, $p < .001$, $d = 1.01$, given that there was no time limit for Part 2. The accuracy for items with mirror foils ($M = 5.81$, $SE = 0.34$) was marginally lower than that for items with structure foils ($M = 6.31$, $SE = 0.31$), $t(46) = 1.96$, $p = .06$.

Frequency of strategy use As Table 2 shows, mental rotation was the dominant strategy, reported by 43 participants and observed on more than 75% of the trials. However it was not the only strategy used. Participants reported an average of 2.15 strategies ($SE = 0.15$, 95% CI [1.8, 2.5]) in their retrospective reports, and the concurrent protocols revealed that they used an average of 1.87 strategies per trial ($SE = 0.09$, 95% CI [1.69, 2.06]). Perspective taking was the least frequently used strategy. Although several participants counted cubes and analyzed the directions of the turns in the figure, a minority (13) noticed the differences in global shape for structure foils and used this feature to eliminate answer choices.[2]

Correlations between the measures from the concurrent and retrospective protocols (see Table 3) are in the medium to high range, indicating good but not perfect agreement between the measures. This is to be expected, given that the retrospective protocols measured only strategies that participants were conscious of and remembered using (Ericsson & Simon, 1984). Correlations (Spearman's rho) between the different strategies, computed on the basis of the concurrent protocols, indicated that use of the mental rotation strategy was negatively related with counting cubes (rho = –.36, $p = .016$) and using the local-turns strategy (rho = –.43, $p = .003$). Other correlations between the strategies were not statistically significant (rho < .3, $p > .05$ in all cases).

Correlations of strategy use with performance Table 3 shows correlations of accuracy on the MRT with the use of each strategy and with the total number of strategies used. Critically, accuracy was positively correlated with use of the global-shape strategy (for concurrent strategy coding, this correlation only reached significance for Part 2 of the test, when concurrent protocols were given). Use of the local-turns strategy was negatively correlated with accuracy in the concurrent protocols. The other strategies were not significantly correlated with accuracy, nor was the total number of strategies used (see Table 3).

Sex differences As in previous studies (Voyer et al., 1995), women ($M = 10.91$, $SE = 0.71$, 95% CI [9.44, 12.39]) had lower MRT scores than men ($M = 13.29$, $SE = 0.89$, 95% CI [11.46, 15.13]), $t(45) = 2.08$, $p = .04$, $d = 0.59$. As Table 4 shows, there were also sex differences in the strategies. In the retrospective protocols, men reported the global-shape strategy more than women (Fisher's exact test, $p = .008$), and women reported the local-turns strategy more than men ($p = .015$). Similarly, in the concurrent protocols men more frequently used the global-shape strategy (Mann–Whitney $U = 149$, $p = .018$), and women more frequently analyzed local turns (Mann–Whitney $U = 112$, $p = .004$). For both measures, women tended to use the perspective-taking strategy more ($ps < .08$ for both measures), whereas no sex differences emerged in mental rotation or counting cubes ($ps > .20$ for both measures). There was no significant difference between men ($M = 1.79$, $SE = 0.14$, 95% CI [1.49, 2.09]) and women ($M = 1.96$, $SE = 0.12$, 95% CI [1.71, 2.20]) in total number of strategies used per trial, $t(42) = 0.87$, $p = .39$.

Table 2 Strategy use on the mental rotation test in Study 1

Strategy	Retrospective Protocols: Number of Participants Reporting Strategy N (%)	Concurrent Protocols: Proportion of Trials on Which Strategy Was Used Mean [95% CI]	Correlation: Concurrent & Retrospective Protocols
Mental rotation	43 (91%)	.76 [.67, .86]	.54, $p < .001$
Perspective taking	9 (19%)	.09 [.02, .15]	.76, $p < .001$
Counting cubes	20 (43%)	.37 [.25, .48]	.60, $p < .001$
Local turns	16 (34%)	.49 [.34, .60]	.60, $p < .001$
Global shapes	13 (28%)	.19 [.10, .28]	.45, $p = .003$

The correlation coefficient is the phi coefficient. To compute this correlation, participants were first categorized as using each strategy if the strategy was observed in their concurrent protocols for at least three of the ten trials, and as not using this strategy otherwise.

[2] The concurrent protocols revealed no significant differences in the strategies attempted on mirror- versus structure-foil items.

Table 3 Correlations of strategies (coded on the basis of retrospective and concurrent strategies) with accuracy on the mental rotation test in Study 1

Strategy	Retrospective Protocols		Concurrent Protocols	
	Total Score	Part 2	Total Score	Part 2
Mental rotation	−.05	.16	.12	.19
Perspective taking	−.25	.21	−.03	.03
Counting cubes	−.16	.04	−.16	.10
Local turns	−.10	.15	−.24	−.32[*]
Global shapes	.37[**]	.40[***]	.25	.31[*]
Number of strategies used	−.06	.05	−.13	.04

Correlations are shown for both the total score on the test and the score only on Part 2 of the test (during which participants gave concurrent protocols). For retrospective protocols, the correlation coefficient reported is the point-biserial correlation; for concurrent protocols and number of strategies used, the correlation coefficient reported is Spearman's rho. [*] $p < .05$, [**] $p < .01$, [***] $p = .005$.

Discussion

Study 1 revealed that a number of strategies besides mental rotation are used to solve items on the MRT. The concurrent and retrospective protocols were quite consistent, supporting the validity of both measures. Sex differences were found in both performance and strategies. Specifically, men used the global-shape strategy more often, whereas women more often analyzed local aspects of the shapes. Moreover, the global-shape strategy was associated with more accurate performance, suggesting that this orientation-independent strategy might be a source of the sex difference in performance.

Study 2

A possible concern regarding Study 1 was that the act of thinking aloud made participants more aware of their strategies and may have altered these strategies. In Study 2, participants did not report their strategies until after they had completed the MRT, so the reports could not have affected their strategies. They completed a posttask questionnaire, which asked them to report whether they had used a piecemeal or a holistic mental rotation strategy, in addition to the strategies identified in Study 1. A holistic strategy involves rotating a mental image of the object as a whole. A piecemeal strategy involves decomposing the mental image into pieces, mentally rotating one part of the image into congruence with the comparison figure, and then applying the same rotation to the other parts of the image to see if they match (Just & Carpenter, 1985). Previous studies using the Shepard and Metzler same–different paradigm have suggested that good imagers use holistic mental rotation strategies, whereas poor imagers use piecemeal strategies (e.g., Bethell-Fox & Shepard, 1988; Khooshabeh, Hegarty, & Shipley, 2013), and that men use holistic strategies more than women do (Heil & Jansen-Osmann, 2008).

Method

Participants The participants were 50 students (25 female, 25 male) in Organic Chemistry classes at the University of California, Santa Barbara. The age range of students in these classes is 18–25 years.

They were recruited through flyers and announcements in their classes and were paid for participation.

Procedure The students participated in groups (four maximum). They were administered three spatial ability tests: the paper-folding test, the Vandenberg MRT, and the visualization of views test, with the usual time limits. Then they completed a strategy choice questionnaire for each of the spatial ability measures. The questionnaires showed a sample item from the test at the top of the page and a list of possible strategies. Table 5 lists the eight strategy questionnaire items for the MRT. Participants were instructed to rank any strategies that they used, from most important to least important and not to rank any strategy they did not use.

Table 4 Strategy use for women and men in Study 1

Strategy	Retrospective Protocols: Number of Participants Reporting Strategy		Concurrent Protocols: Proportion of Trials on Which Strategy Was Used	
	Women N (%)	Men N (%)	Women M [95% CI]	Men M [95% CI]
Mental rotation	20 (87%)	23 (96%)	.69 [.53, .84]	.85 [.74, .95]
Perspective taking	7 (30%)	2 (08%)	.14 [.03, .26]	.02 [.01, .05]
Counting cubes	9 (39%)	11 (46%)	.38 [.21, .55]	.36 [.19, .52]
Local turns	12 (52%)	4 (17%)	.65 [.48, .83]	.27 [.10, .43]
Global shapes	2 (09%)	11 (46%)	.09 [.02, .16]	.31 [.15, .47]

Table 5 Strategies included in the questionnaire in Study 2 and numbers (and percentages) of participants reporting each strategy among their top three strategies

Strategy	Number (Percent) of Participants Reporting
Holistic Mental Rotation: "I imagined myself being stationary and the whole object rotating with respect to me."	31 (62%)
Piecemeal Mental Rotation: "I first imagined rotating a part of the object and then checked whether the rest of the object could be rotated in the same way to match the target."	26 (52%)
Perspective Taking: "I imagined the objects being stationary as I moved around them to view them from different perspectives."	12 (24%)
Embodied Strategies: "I imagined that the object was an animal (e.g., a snake) and where its head or arms or tail would be."	3 (6%)
Counting Cubes, All Arms: "I counted the number of cubes in the four straight segments of the object."	11 (22%)
Counting Cubes, End Arms: "I counted the number of cubes in just the two end segments of the object."	6 (12%)
Local Turns, All Arms: "I examined the directions of the four segments of the object with respect to each other."	31 (62%)
Global Shapes: "I examined the directions of the of the two end segments with respect to each other."	23 (46%)

Results

The mean score on the MRT was 10.80, $SE = 0.65$, 95% CI [9.51, 12.13]. There were no significant differences between performance on Part 1 ($M = 5.46$, $SE = 0.35$, 95% CI [4.73, 6.19]) and Part 2 ($M = 5.36$, $SE = 0.34$, 95% CI [4.68, 6.04]) of the test, or on items with mirror foils ($M = 5.26$, $SE = 0.36$, 95% CI [4.53, 5.99]) versus structure foils ($M = 5.56$, $SE = 0.33$, 95% CI [4.90, 6.23]).

Frequency of strategy use As Table 5 shows, the majority of participants reported either holistic or piecemeal mental rotation (11 reported both) but, as in Study 1, several other strategies were also reported. Participants reported an average of 3.64 strategies ($SE = 0.17$, 95% CI [3.30, 3.98]).

Correlation of strategy use with performance Participants were categorized as using a strategy if it was ranked as one of their top three strategies. Table 6 shows the correlations of strategy use with MRT score. Accuracy was negatively correlated with use of the perspective-taking strategy, and marginally positively correlated with the global-shape strategy. Accuracy was also positively correlated with holistic mental rotation, and negatively correlated with counting cubes in all arms of the figures. As in Study 1, accuracy was not correlated with the total number of strategies reported ($r = -.07$, $p > .25$).

Sex differences The overall score of women ($M = 9.00$, $SE = .88$, 95% CI [7.19, 10.81] was lower than that of men ($M = 12.64$, $SE = 0.83$, 95% CI [10.93, 24.35]), $t(48) = 3.02$, $p = .004$, $d = 0.79$. There were also sex differences in strategies used, as is shown in Table 6. Chi-square tests of independence indicated that men reported the global-shape strategy more than women did, $\chi^2(1, N = 50) = 9.74$, $p = .002$ (replicating Study 1), and tended to report more holistic mental rotation, $\chi^2(1, N = 50) = 4.16$, $p = .04$. In contrast, women tended to report perspective

taking more than men did, $\chi^2(1, N = 50) = 3.95$, $p = .05$. Men ($M = 3.64$, $SE = 0.21$) and women ($M = 3.64$, $SE = 0.27$) did not differ in the number of strategies reported.

General discussion

Across two studies, concurrent and retrospective strategy reports revealed that people use a variety of strategies to solve items on the Vandenberg and Kuse (1978) MRT. Although mental rotation is the most common strategy, it is clearly not the only strategy used. Most people reported using other strategies, including perspective taking, counting cubes, local turns, and global shapes.

Overall, use of a mental rotation strategy was not correlated with success on this test, likely because almost all participants reported some form of mental rotation as one of their strategies. However, reporting holistic rotation (rotating the figure as a whole), as opposed to a piecemeal strategy (rotating the

Table 6 Correlations of strategies reported in Study 2 with accuracy on the mental rotation test, along with numbers (and percentages) of men and women reporting those strategies

Strategy	Correlation With Accuracy	Men N (%)	Women N (%)
Holistic mental rotation	.38, $p = .01$	19 (76%)	12 (48%)
Piecemeal mental rotation	.05, $p > .25$	11 (44%)	15 (60%)
Perspective taking	−.35, $p = .01$	3 (12%)	9 (36%)
Embodied strategies	−.03, $p > .25$	2 (8%)	1 (4%)
Counting cubes (all arms)	−.48, $p < .001$	3 (12%)	8 (32%)
Counting cubes (end arms)	.09, $p > .25$	3 (12%)	3 (12%)
Local turns (all arms)	.06, $p > .25$	15 (60%)	16 (64%)
Global shapes	.25, $p = .08$	17 (68%)	6 (24%)

The measures of strategy use are dichotomous (whether or not the participant reported each strategy). The correlation coefficient reported is the point-biserial correlation.

figure piece by piece; see Table 5), was correlated with success. This result is consistent with previous studies using the Shepard and Metzler same–different paradigm (Bethell-Fox & Shepard, 1988; Khooshabeh et al., 2013), suggesting that the ability to maintain a holistic image of a three-dimensional structure under spatial transformations is one source of individual differences in performance on this test.

The use of the global-shape strategy was also correlated with success. This strategy involves comparing orientation-independent features of the objects (cf. Jolicœur, 1990; Takano, 1989), obviating the need for mental rotation on trials with structure foils. Geiser at al. (2006) identified this strategy on the basis of patterns of performance across items and characterized it as a strategy used by low-spatial individuals who are unable to use mental rotation. In contrast, the present study reveals that it is also used by some of the highest scorers on the MRT and is positively correlated with success on this test.

The present study reveals that strategy choice can also negatively impact performance on the MRT. Some participants use time-consuming but ineffective strategies, such as counting cubes or encoding the relative directions of all arms of the figures. Counting cubes is particularly ineffective, because the shapes cannot be discriminated on the basis of number of cubes in the successive arms. Perspective taking is also an ineffective strategy for this test.

The present studies replicated the usual sex difference in the MRT. Importantly, there were qualitative differences in the strategies used by men and women to approach this test. Notably, women were more likely to pay attention to local details (counting cubes or encoding the relative directions of all arms) and to report perspective taking. In contrast, men were more likely to notice aspects of the overall shape of the object (global shape) that could be used to quickly identify structure foils, and to report rotating the object as a whole, rather than piecemeal. These results suggest that men and women may differ in their preferences or abilities for global or holistic as opposed to piecemeal processing of visual spatial stimuli.

A possible concern with any protocol study is that asking participants to think aloud might influence the strategies that they adopt. Although this might have been a factor in Study 1, note that the concurrent protocol instructions (see Ericsson & Simon, 1984, p. 378) did not ask participants to analyze their thinking processes in any way. In both studies participant were asked to report their strategies (in retrospective protocols in Exp. 1 and in strategy questionnaires in Exp. 2) *after* they had completed the test, so strategy reports were unlikely to have influenced test performance.

It is somewhat surprising that no significant differences in accuracy emerged between problems with mirror versus structure foils. The concurrent protocols also revealed no significant differences in attempted strategies on these items. Some participants attempted the global-shape strategy initially on all trials, including those with mirror foils, turning to other strategies only if the global-shape strategy did not yield an answer. In terms of accuracy, the MRT is a timed test, and most people do not finish in the time allotted. Using the global-shape strategy to quickly solve a structure-foil item also allowed participants more time to complete the mirror-foil trials. This, and the relatively small number of items on the test, may account for the nonsignificant differences between item types. In a recent study, Boone and Hegarty (2017) created two versions of the test, one with all structure foils and another with all mirror foils, and found significantly better performance for the structure-foil version. Moreover, they generalized this finding to versions of the Shepard and Metzler (1971) same–different task and demonstrated that when the foils were structurally differel and participants were taught the global-shape strategy, performance improved and sex differences were no longer evident, providing preliminary evidence for a causal connection between use of this strategy and performance.

In summary, success on the MRT appears to reflect both basic capacities for encoding and transforming spatial images and the ability to notice orientation-independent aspects of shape. Although some individuals appear to have large spatial working memory capacities that enable them to mentally rotate mental images of objects, other successful solvers of this test capitalize on efficient analytic strategies when possible. Less successful solvers waste time encoding stimulus features, such as numbers of cubes, that do not differentiate the test items. As we consider how best to enhance spatial thinking, it is important to include training in optimal strategy choice, which may be more malleable than improving students' basic working memory capacities for storing and transforming visual spatial representations.

References

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 12–23. doi: 10.1037/0096-1523.14.1.12

Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1005–1019. doi:10.1037/xlm0000370

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

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140,* 102–116. doi:10.1037/a0021790

Eckstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 393–404.

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating" rotators" from" nonrotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research, 41,* 261–293.

Guay, R., & McDaniels, E. (1976). *The visualization of viewpoints*. West Lafayette, IN: Purdue Research Foundation.

Hegarty, M. (2010). Components of spatial intelligence. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 265–297). San Diego, CA: Academic Press.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32,* 175–191.

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121–169). Cambridge, UK: Cambridge University Press.

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology, 61,* 683–689. doi:10.1080/17470210701822967

Jolicœur, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind & Language, 5,* 387–410.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92,* 137–172. doi:10.1037/0033-295X.92.2.137

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence, 35,* 211–223.

Khooshabeh, P., Hegarty, M., & Shipley, T. F. (2013). Individual differences in mental rotation: Piecemeal vs. holistic processing. *Experimental Psychology, 60,* 164–171.

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

Lohman, D. (1988). Spatial abilities as traits, processes and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181–248). Hillsdale NJ: Erlbaum.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27. doi:10.1037/0096-3445.125.1.4

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

Takano, Y. (1989). Perception of rotated forms: A theory of information types. *Cognitive Psychology, 21,* 1–59.

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

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 4,* 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., & Doyle, R. A. (2010). Item type and gender differences on the Mental Rotations Test. *Learning and Individual Differences, 20,* 469–472.

Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the Mental Rotations Test. *Canadian Journal of Experimental Psychology, 60,* 91–100. doi:10.1037/cjep2006010

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. doi:10.1037/0033-2909.117.2.250

Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in visual–spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review, 24,* 307–334.

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. doi:10.1037/a0016127

Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics, 31,* 201–209. doi:10.3758/BF03202524