

Sex Differences in Mental Rotation Ability Are a Consequence of Procedure and Artificiality of Stimuli

Maryanne L. Fisher¹ · Tami Meredith² · Melissa Gray³

© Springer International Publishing AG 2017

Abstract Mental rotation is a cognitive process that involves performing rotations on visual images or objects, which has played a significant role in humans' evolutionary past. Sex differences in mental rotation ability have been extensively assessed using the Vandenberg and Kuse (1978) Mental Rotations Test. This test produces consistently higher scores for men than women, which has led numerous researchers to conclude that males have superior mental rotation ability. The causes of this sex difference have been widely debated, and research remains inconclusive. Various researchers have challenged the legitimacy of this male advantage by investigating moderating factors that are part of the assessment process. Here we show, through the use of photographs and three-dimensional models, that the form of the stimuli can eliminate the sex difference. Our results suggest that the sex difference found on this test is not due to a male advantage in spatial ability, but is an artifact of the stimuli.

Keywords Mental rotation · Sex difference · Representation effects · Time limit effects

Everyone needs to understand and represent themselves and objects in their surrounding environment. Moreau et al. (2010)

✉ Maryanne L. Fisher
mlfisher.99@gmail.com

¹ Department of Psychology, Saint Mary's University, 923 Robie Street, Halifax, NS B3H 3C3, Canada

² Department of Math & Computer Science, Saint Mary's University, Halifax, Canada

³ Student Development & Counseling, SAIT Polytechnic, Calgary, Canada

point out that the ability to move oneself and objects in environments, or to imagine doing so, requires spatial coding. However, they argue that people vary in their efficiency to perform this coding, which leads to individual differences in speed and accuracy on spatial tasks. One task that is routinely used to examine performance is mental rotation. Mental rotation is the visuospatial ability to quickly and accurately rotate objects in one's mind (see Shepard and Metzler 1971).

Mental rotation has important ties to humans' evolutionary history, given that it involves cortical areas directly linked to perception, tracking objects in motion, and determining spatial relations (see for a partial review, Parsons et al. 2004). It is considered to be related to intelligence (Kaufman 2007) and performing motor actions (Moreau et al. 2012), and critical for executing daily tasks such as orientation (Pazzaglia and Moè 2013). In light of the importance of mental rotation for daily life, and its links to intelligence and other cognitive factors, we propose that the sexes should respond similarly on tasks tapping into this ability. That is, both sexes would benefit from being able to orient themselves in space, relative to other objects, and hence, evolutionary explanations need to examine similarities between the sexes rather than exclusively focus on differences.

Indeed, the overwhelming majority of evolutionary-based research has focused on men's superior performance, relative to women. For example, Silverman et al. (2000) tie mental rotation to human's evolutionary past by proposing that the ability to three-dimensionally mentally rotate is linked to evolved hunting skills and closely linked to wayfinding ability, leading men to perform better on related tasks. Vashro and Cashdan (2015) agree that mental rotation is tied to navigational ability and directly link it to male reproductive success. Among Tve and Tjimba men, those with better mental rotation have larger traveling ranges and father more children with more women.

A large body of research has established a male advantage on mental rotation through the use of one test in particular: the Vandenberg and Kuse Mental Rotations Test (1978, MRT) (e.g., Geary et al. 1992; Hedges and Nowell 1995; Linn and Peterson 1985; Masters 1998; Masters and Sanders 1993; Maccoby and Jacklin 1974; Vandenberg and Kuse 1978; Voyer et al. 1995). The MRT is a paper-and-pencil version of the Shepard and Metzler mental rotation task (1971). In the original Shepard and Metzler test, participants were presented with pairs of drawings and asked whether the items were rotated, mirror images, or the same as a target. In contrast, the MRT consists of two-dimensional line drawings of three-dimensional block figures that are rotated around all three axes. When completing this task, participants are required, under limited time, to choose two of the four block figures that match a target figure but that are rotated differently. The drawings are isometric (i.e., not orthographic and do not decrease in size with distance) and do not contain shadows or any other indicators of realism. This test has produced the largest magnitude of sex differences in cognitive functioning (Halpern 2012; Kimura 1999; Parsons et al. 2004; Voyer and Saunders 2004). This male advantage has been largely interpreted to suggest that men are more skilled, than women, in the act of cognitively rotating an object and visualizing how that object would appear from another perspective.

Researchers have long debated which factors account for the origin of the sex difference on the MRT. This debate has largely focused on biological, evolutionary, and environmental causes that exhibit a determining influence over the development of spatial abilities (Voyer 1997). For instance, some have argued that factors such as handedness (Annett 1992), brain lateralization (Gur et al. 2000), or hormone levels (Silverman and Phillips 1993) have a large impact on this sex difference. Evolutionary theories suggest that men's superior spatial ability evolved for navigating (Geary 1995; Silverman et al. 2000), as a result of males' intrasexual competition for resources related to warfare or hunting (Symons 1979), or as a consequence of males having larger home ranges than females (Gaulin and FitzGerald 1989). In contrast, others have argued that environmental factors have shaped this ability, such as gender role socialization (Saucier et al. 2002), the differential participation of males and females in spatially oriented activities (Voyer et al. 2000), and the number of stereotypically masculine spatial activities engaged in while a youth (Nazareth et al. 2013). These environmental and social explanations may in part account for large cultural differences in MRT performance (e.g., between Oman and Germany, Jansen et al. 2016). Researchers generally conclude that there is a range of factors that may explain this sex difference (Halpern 2012).

Procedural Issues

These explanations might be somewhat unnecessary if the MRT is simply not accurate for assessing mental rotation ability, a possibility that is suggested by the differing results caused by various procedural and measurement manipulations (e.g., Goldstein et al. 1990; Moè 2016; Voyer et al. 1995). Past research shows that when performing the MRT, time restriction causes women to perform at a lower level than they would otherwise (e.g., Peters 2005; Voyer 1997), possibly because they work more slowly and cautiously than men (Goldstein et al. 1990), or because they are more detail-oriented rather than holistic (Boone and Hegarty 2017). This possibility has gained limited support though; the removal of the standard time limit of 10 min to complete the MRT has led some researchers to conclude that there is no sex difference (Goldstein et al. 1990; Voyer 1997), while others report a reduction in magnitude (study 1, Peters 2005), and others continue to find a male advantage (Masters 1998; Resnick 1993). Consequently, the role of time restraints remains inconclusive and warrants more investigation. The ability of procedural factors such as time restrictions to influence the pervasiveness of this sex difference questions the legitimacy of past findings and provides reasons for the continued investigation of assessment factors.

Stimuli Characteristics

In an effort to advance such an inquiry, researchers have examined the persistence of this sex difference when real three-dimensional objects are employed instead of abstract two-dimensional depictions (Kaushall and Parson 1981; McWilliams et al. 1997; Parsons 1995; Robert and Chevrier 2003). McWilliams et al. (1997) compared performance on a paper-and-pencil test against the use of physical three-dimensional models. Participants were asked to state whether the two figures (adapted from Shepard and Metzler 1971) were the same or different. They found that a sex difference emerged for the paper-and-pencil task, but not the three-dimensional models. The use of such stimuli has modified this sex difference, either by reducing its magnitude (Robert and Chevrier 2003) or eliminating it (Kaushall and Parson 1981; McWilliams et al. 1997; Parsons 1995).

These findings suggest that the move from abstract to more realistic stimuli has the potential to reduce or eliminate any sex difference. However, although informative, this past research fails to identify the point when the sex difference is removed and when the stimuli shifts from two-dimensional to three-dimensional representation. For example, perhaps the sex differences still exists for a two-dimensional photograph but not when participants can actually see the three-dimensional blocks mounted on a board. Past researchers have examined

one situation only, rather than a range of stimuli that vary according to their level of abstraction.

Moreover, the depiction of dimensionality could cause the established sex difference on the MRT. Voyer and Hou (2006, see also Boone and Hegarty 2017) examined the possibility that items containing structurally different foils (i.e., incorrect options) are easier to complete, as these questions only involve object recognition, whereas questions that include foils that are mirror images involve both object recognition and mental rotation. There was no evidence to suggest that questions containing structurally different foils produced the sex difference, and hence, the existing sex difference could not be attributed to difference in mental rotation ability. To explore this issue further, they examined the occlusion of items on the MRT. Parts of the object are obstructed from view as a result of the rotation of the object. Occlusion may make these questions more difficult, as the shape of these objects may be misperceived. Given that occlusion is a by-product of the three-dimensional nature of the stimuli used on the MRT, the authors argued that if larger sex differences were found on occluded items versus non-occluded items, these differences could be linked to the three-dimensional nature of the task. Results indicated that occluded items did produce poorer performance than non-occluded items and that there was a larger sex difference found on occluded items. Accordingly, they suggest that the three-dimensional nature of the task may account for the sex difference. This finding is at odds with those obtained when using three-dimensional models, in which case, no sex difference is reported (e.g., Kaushall and Parson 1981). This discrepancy also warrants further investigation.

Aims and Hypotheses

To fully explore the influence of the form of the stimuli, we performed two experiments to assess the influence of three-dimensional stimuli on the reported sex difference on the MRT. In experiment 1, we attempted to replicate the existing sex difference for the traditional paper-and-pencil version of the test, using the same procedures (i.e., time limit and negative scoring). We also sought to explore the issue of time restrictions and, hence, included an additional untimed condition. We further sought to explore the issue of stimuli format by including two conditions where the stimuli were photographs of real, three-dimensional block models. To provide a complete model, we tested participants in both timed and untimed conditions when viewing the photograph version. In experiment 2, we continue to address the issue of stimuli form by asking participants to complete the same MRT task using one of five forms of the stimuli that varied in their levels of realism, including mounted three-dimensional blocks that could be gently touched while blindfolded, concluding with models that they could touch and rotate while sighted. To date,

we know of no study that has examined both time effects and this continuum of stimuli formats, varying from two dimensional to three dimensional.

In keeping with some of the past work (e.g., Voyer 1997), we hypothesized (hypothesis 1) that the removal of a time restriction would lead to the elimination of the sex difference. Similar evolutionary-relevant cognitive demands are on both women and men, such that both sexes need to be able to rotate objects, including themselves, in the real world. Thus, if the test is measuring this ability, even in part, then there should not be any sex difference when time is removed. We expect that when there is a time restriction, men will outperform women due to their reliance on a holistic strategy, whereby they see the global shape, which is faster than using detail-oriented strategies such as counting blocks (Boone and Hegarty 2017; Pletzer 2014).

We also hypothesized (hypothesis 2) that the sex difference could be partly due to the artificial nature of the paper version of the test and that the removal of abstraction and artificiality would decrease the sex difference. Our reasoning is largely based on research in computer programming (Turkle and Papert 1990) and computer code comprehension (Fisher et al. 2006), where evidence shows that women prefer to work using a “bottom-up” approach and that men are more likely to use a “top-down” approach. In this domain, computer source code, which is highly concrete and displays clear functionality, is considered as the bottom, and an abstract domain, such as banking, manufacturing, or image manipulation, is considered as the top. Thus, there is interdisciplinary evidence that women may prefer to work with more concrete concepts than with more abstract ones. Given that the paper-and-pencil version of the MRT lacks many details related to realism (e.g., not orthographic, no decrease in size due with increased distance) and that women may also believe that such artificial tasks are outside of their abilities (see Moè 2016), the sex difference may decrease as realism increases.

Experiment 1

Methods

Participants We tested 132 women (age, in years, $M = 21.19$, $SD = 4.05$) and 107 men (age, in years, $M = 21.80$, $SD = 4.57$). Approximately 75% of the participants were Caucasian. All were university students in at least one psychology course at any year of study and received a small course credit for their participation. The university was a moderate-sized public institution in Canada.

Study Design and Measures Our independent variables were participant sex (female vs. male), timing (timed vs. untimed),

and stimuli type (drawings vs. photographs), and total score was as the dependent variable.

Our stimuli were either the Vandenberg and Kuse MRT (1978) or photographs of models made to match the MRT. For the first two conditions (paper-and-pencil based, timed and untimed), we used a photocopied version of Vandenberg and Kuse's instrument with no modifications. Then, we created 125 wooden block models (24 test items and one practice item) from 1-in square wooden strips, painted them white, and drew black lines, at 1-in intervals, to mimic the line drawings. We reproduced the original figures such that each model precisely matched a specific drawing from the paper-based version of the MRT. These models were then professionally photographed with a standard light gray background. Each photograph was of a single model rotated so that its orientation was as close as possible to a single figure in the MRT. The photographs were then arranged such that one page contained a photograph reproduction (five photographs) of one item from the paper-based MRT (see Fig. 1). The only difference between the photograph and paper version was the number of items per page (1 vs. 6) and the size of each individual figure (i.e., the photographs were roughly twice as high and wide, but with the same height-width ratio, as the line drawings). The larger size for the photographed stimuli was simply a layout decision; having small stimuli on letter-sized sheets seemed odd to the point of distraction for pilot-testing participants ($n = 4$). There was a letter placed above each individual image, and participants were provided with a response sheet. The first page of the booklet of photograph stimuli contained the example item used in the paper version, and the same instructions were provided.

Procedures Participants were tested in small groups of up to five individuals. The administration of the paper survey was identical to that used previously (e.g., Vandenberg and Kuse 1978), except in the untimed conditions when participants were informed that they had as much time as they needed. In the timed, photograph condition, participants completed a photographed version of the MRT with a time limit of 11 min. An extra minute was added to the original time restriction to compensate for the increased number of pages that would have to be flipped (24 vs. 4), as each page only contained one rotation problem. Furthermore, unlike the original MRT, participants in the photograph conditions were not able to record their answers directly on the test booklet, thus the extra minute also accounted for the additional time needed for participants to record their answers on a separate response sheet. This decision was made for financial reasons as the photograph items were printed on high-quality photograph paper, laminated to prevent damage, and then bound into a booklet. The instructions were otherwise the same as the original version using isometric drawings. Participants were randomly

placed in each of the four conditions (drawings vs. photographs, timed vs. untimed).

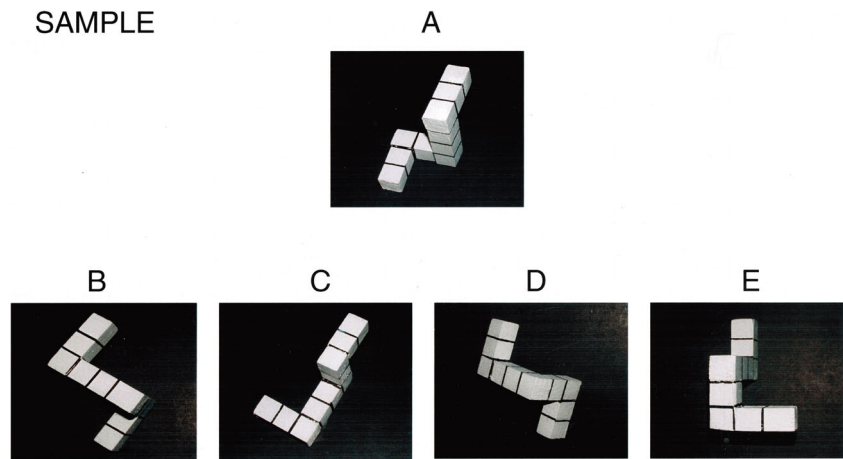
Results

A $2 \times 2 \times 2$ analysis of variance (ANOVA) was performed with participant sex, timing, and stimuli type as the independent variables and total score (maximum 48) as the dependent variable. For all comparisons, two-tailed tests with a significance level of $\alpha = 0.05$ was used. Descriptive statistics, broken down by the independent variables, can be found in Table 1.

This model yielded a significant main effect for participant sex, $F(1, 231) = 33.09, p < 0.001, \eta_p^2 = 0.12$, such that men ($M = 30.39, SD = 13.57$) performed better than women ($M = 20.36, SD = 12.20$). There was also a significant main effect for stimuli, $F(1, 231) = 5.32, p = 0.02, \eta_p^2 = 0.02$, such that scores were higher for the drawings ($M = 26.56, SD = 3.52$) than for the photographs ($M = 22.96, SD = 13.80$). Timing was also a significant factor, $F(1, 231) = 13.65, p < 0.001, \eta_p^2 = 0.07$, such that untimed scores ($M = 28.66, SD = 14.50$) were higher than timed scores ($M = 22.12, SD = 12.52$).

There was a significant interaction for sex with stimuli, $F(1, 231) = 5.23, p = 0.02, \eta_p^2 = 0.01$. For the drawings, men ($M = 33.93, SD = 12.32$) performed better than women ($M = 20.66, SD = 11.44$), confirmed by a two-tailed independent samples t test, $t(124) = 6.26, p < 0.001$. The same pattern emerged for the photograph conditions; men ($M = 26.51, SD = 13.93$) performed better than women ($M = 20.03, SD = 13.08$), $t(111) = 2.54, p = 0.01$. Men performed significantly better in the drawing conditions than in the photograph conditions, $t(105) = 2.92, p = 0.004$. However, women's performance was equivalent across condition, $t(130) = 0.29, p = 0.77$.

There was also a significant interaction for stimuli with timing, $F(1, 231) = 4.77, p = 0.03, \eta_p^2 = 0.03$. For the photographs, independent samples t tests revealed a significant difference for timed ($M = 18.51, SD = 11.26$) versus untimed scores ($M = 28.56, SD = 14.74$), $t(111) = 4.11, p < 0.001$. In contrast, there was no significant difference between timed ($M = 25.11, SD = 12.79$) versus untimed scores ($M = 28.76, SD = 14.41$) for the drawings, $t(111) = 1.49, p = 0.14$. Moreover, within the timed conditions, scores on the drawings ($M = 25.11, SD = 12.79$) were higher than for the photographs ($M = 18.51, SD = 11.26$), $t(137) = 3.20, p = 0.002$. Contrariwise, there was no significant difference in the untimed conditions, as scores on the drawings ($M = 28.76, SD = 14.74$) were similar to those for the photographs ($M = 28.56, SD = 14.74$). There was no significant interaction between participant sex and timing, $F(1, 231) = 0.04, p > 0.05$, or participant sex, stimuli type, and timing, $F(1, 231) = 0.003$,

Fig. 1 Photograph version of mental rotation stimuli

$p > 0.05$. It should be noted that a t test between male and female scores for the photograph untimed condition did not reveal a significant difference; $t(48) = 1.32, p = 0.20$.

Discussion

Overall, men had higher scores than women, and untimed conditions resulted in higher scores than the timed conditions. Scores for the photographs were lower than for the drawings, an effect that appears to be driven by men's decrease in the photograph conditions. Timing is important; across participants, when untimed, scores for the photographs are not significantly different than for the drawings. Thus, it seems that there is no noticeable improvement due to the increased realism provided in the photographs, and when participants have sufficient time, there is no difference in scores on the drawings versus the photographs.

An inspection of the descriptive statistics shows that mental rotation scores for the timed photograph condition were the lowest of the four conditions, possibly because the added detail (e.g., cues of depth) meant that extra visual and visuospatial processing had to occur. It is also possible that the extra minute was not sufficient for participants to flip the pages and

go to the next item. However, an analysis of the attempted items shows no significant difference with the drawing timed condition, $t(237) = 1.42, p = 0.16$.

We did not find support for our hypothesis that the addition of realism eliminates the sex difference. Collapsed across timed versus untimed, women's scores were almost identical for the drawings versus the photographs, while men's scores surprisingly decreased in the photograph conditions. We partially explore this finding in experiment 2, as well as further investigate the issue of realism.

Experiment 2

We had two goals in experiment 2; first, we sought to replicate the effects obtained in experiment 1. Second, we wanted to further explore how performance in mental rotation is linked with the form of the stimuli. In the untimed conditions of experiment 1, women's performance was almost identical in the drawing and photograph conditions, although the photograph condition introduced several cues for depth, as provided by the orthographic nature of the photographs. Moreover, past research (e.g., Robert and Chevrier 2003; McWilliams et al. 1997) has found that the use of three-dimensional models led to no sex difference in performance, as compared to the paper-and-pencil test. This finding is interesting given that past researchers have argued that mental rotation and manual rotation share a similar cognitive process (e.g., Gardony et al. 2013; Wohlschälger and Wohlschälger 1998). Consequently, we predict there must exist a stage or tipping point where the sex difference is noticeably decreased, and this stage will fall between the two-dimensional representation (i.e., drawings) and a full physical representation where participants can manually manipulate the blocks. This last point is important, as manual rotation leads to no sex difference (e.g., Gentaz and Hatwell 1995), and thus, we expect that as the MRT becomes more like a test of manual rotation, the sex difference will be eliminated.

Table 1 Descriptive statistics for independent variables of experiment 1

Condition	Timing	Sex	Number	M	SD
Drawings	Timed	Female	45	19.78	10.66
		Male	31	32.84	11.73
	Untimed	Female	25	22.24	12.79
		Male	25	35.28	13.12
Photographs	Timed	Female	37	16.11	10.76
		Male	26	11.27	11.27
	Untimed	Female	25	25.84	14.24
		Male	25	31.28	15.01

Methods

Participants We tested 116 women (age, in years, $M = 21.27$, $SD = 2.63$) and 115 men (age, in years, $M = 21.96$, $SD = 4.29$). They came from the same participant pool as experiment 1, and hence, approximately 75% of the participants were Caucasian, and all were students at a university in Canada. All participants were sighted (i.e., not blind or hard of sight) and right-handed, which is critical for the experimental conditions described below.

Design, Measures, and Procedures The five conditions (drawings, photographs, mounted, blindfold, vs. touch) were an independent variable, and participant sex (female vs. male) was the second independent variable; total score, out of 48, was the dependent variable. All tests were completed without a time restriction, as it was not possible to uniformly and reliably present the stimuli within tight time limits. The drawing (stimuli condition 1) and the photograph conditions (stimuli condition 2) were the same as experiment 1. In the mounted condition (stimuli condition 3), the blocks were professionally mounted using long screws to thin gray boards, such that they replicated the original MRT drawings. It should be noted that the drawings defy gravity in some instances, and hence, the screws went through the board and back of the blocks in such a way as to hold them securely in the correct position, but be invisible to the eye. The board was marked with a thick, black line between the target item and the remaining four blocks, and letters (i.e., A thru D) matching the response sheet were written below the blocks. Participants were provided with an example board, matching the example on the paper-and-pencil version. A wooden holder was made, such that the board was held at a 45° angle on a tabletop in front of the participants, and presented in a well-lit area to minimize shadows cast by the blocks. The board was presented 16 in away from participants. Participants were not allowed to touch the models nor move the board closer to them. All participants were asked if they could see the blocks clearly and were allowed to move their chair up or down slightly, in height, if needed. Participants recorded their answers on a response sheet. In this condition, then, participants could see but not touch the three-dimensional blocks.

In the blindfold condition (stimuli condition 4), participants were allowed to touch the blocks while wearing a black, thick, blindfold. Note that the blocks were unmounted in this condition. The blindfold condition temporally preceded the mounted condition, as the holes caused by the screws were irreversible and could interfere with the accuracy of the mental rotations. However, conceptually the blindfold condition was to come after the mounted condition as it involves manual rotation and should yield no sex difference; hence, it is discussed as having occurred after the mounted condition. A

guide board was made, such that the target block was placed in a shallow box with a felted bottom, and there were four additional “boxes” to the right, created by adding thin wooden stripping for dividers. The four remaining blocks to be matched to the target were placed, in the order of the original MRT, from left to right in these boxes. The participant read the instructions (the same as those used in the other conditions), and then, they were blindfolded. Participants were advised to keep their left hand on the target and use their right hand for exploring the other blocks, as pilot testing revealed that participants lost track of the target block if they were allowed to move the blocks freely. Also, they were allowed to pick up the blocks to rotate them, but not more than an inch from the bottom of the box. An experimenter watched the testing from a nearby table to ensure that this restriction was enforced. This procedural step stopped participants from moving two blocks close together and exploring whether they fit together, which could be a cue that they matched. Participants first completed an example item and then told the researcher which blocks they believed were correctly matched to the target. The researcher told them whether they were correct or incorrect, as per the original MRT instructions, and permitted the participant to investigate, while still blindfolded, the items again before proceeding. During the testing session, participants informed the researcher which items were correct, and the researcher recorded their answers, removed the blocks, and then laid out the next blocks in a manner that closely resembled, as much as possible, the MRT. Once they had completed half the items, participants were asked if they would like a short rest, and the majority (82%) agreed. In this blindfold condition, participants could therefore touch, but not see, the three-dimensional blocks. Thus, while participants could physically rotate the blocks (i.e., manual rotation), we propose that they had to mentally visualize the blocks, then rotate and match their visualizations.

In the touch condition (stimuli condition 5), participants were allowed to use the same shallow box as the blindfold condition and manually manipulate the blocks. The only difference to the blindfold condition is that participants were able to record their own scores on a response sheet. In this condition, participants could touch, rotate, and see the blocks, but the same movement restriction (i.e., no more than an inch) was enforced. Similar to the blindfold condition, the touch condition temporally preceded the mounted condition, but conceptually, it was to follow it, and hence, it will be described as having occurred later.

Results

An ANOVA model was created with the five conditions and participant sex as independent variables and total score as the dependent variable. Descriptive statistics broken down by

independent variable can be found in Table 2. For all comparisons, a significance level of $\alpha = 0.05$ was used.

There was a main effect for participant sex, $F(1, 221) = 6.59$, $p = 0.01$, $\eta_p^2 = 0.05$, with men generally performing better than women. There was also a main effect due to condition, $F(4, 221) = 25.72$, $p < 0.001$, $\eta_p^2 = 0.30$. Tukey HSD post hoc analysis revealed many significant findings at $p < 0.001$. The comparisons that were significant were the drawing versus mounted, blindfold, and touch conditions and photograph versus mounted, blindfold, and touch conditions. Thus, there was a split, such that the two-dimensional stimuli forms (i.e., drawing and photograph conditions) were not significantly different from each other, and the three-dimensional stimuli forms (i.e., blindfold, mounted, and touch) were not significantly different from each other, but the two versus three-dimensional forms were different.

As well, there was a significant interaction between participant sex and condition, $F(4, 221) = 2.38$, $p = 0.05$, $\eta_p^2 = 0.03$. Independent samples t tests, shown in Table 2, revealed that women and men significantly differed for the drawing condition. However, there were no sex differences in any other condition. There was no significant difference between women and men for the photograph condition, the mounted condition, the blindfold condition, or the touch condition.

For exploratory purposes, we split the sample by sex and analyzed the five stimuli form conditions using one-way ANOVAs. For men only, $F(4, 114) = 6.88$, $p < 0.001$. Post hoc LSD comparisons revealed that scores in the drawing and photograph conditions were equivalent and significantly different from the rest (values excluded for brevity). The touch, blindfold, and mounted conditions did not differ from each other, but are significantly different from the drawing and photograph conditions. Thus, there was a clear division between two-dimensional and three-dimensional tests for men with respect to their performance. For women only, $F(4, 115) = 21.79$, $p < 0.001$, and post hoc analyses indicate that the drawing and photograph conditions were equivalent and significantly different from the rest. The blindfold and touch condition were equivalent, but significantly different from the rest. Last, the mounted condition was significantly different from all other conditions. Therefore, for women, there is a division between two-dimensional, the mounted condition, and the manual rotation conditions, with performance improving in that order.

It may be interesting to mention that the number of attempted items versus correct items did not show sex differences within each condition (analyses omitted for brevity). Moreover, for all conditions, the number of attempted items significantly exceeded the number of correctly solved items. Paired samples t tests split by condition revealed for the paper timed condition, $t(75) = 11.58$, $p < 0.001$, paper untimed condition $t(62) = 10.92$, $p < 0.001$, photograph timed condition $t(62) = 12.400$, $p < 0.001$, photograph untimed condition $t(57) = 10.01$, $p < 0.001$, mounted condition $t(39) = 6.77$,

Table 2 Descriptive and t test statistics of experiment 2

Condition	Sex	Number	M	SD	t test	p value
Drawing	Female	32	22.69	12.76	3.12	.003
	Male	31	33.35	14.34		
Photograph	Female	30	26.00	14.10	1.39	.17
	Male	28	31.29	14.77		
Mounted	Female	18	35.89	9.51	1.49	.15
	Male	17	40.65	9.40		
Blindfold	Female	19	43.11	5.22	.96	.34
	Male	21	44.38	3.01		
Touch	Female	17	46.71	1.72	1.99	.06
	Male	18	44.00	5.36		

$p < 0.001$, blindfold condition $t(40) = 4.33$, $p < 0.001$, and touch condition $t(35) = 2.63$, $p = 0.013$.

Discussion

There were significant sex differences for the drawings, which replicated the findings of experiment 1. However, there were no sex differences in any of the other conditions (i.e., when the stimuli were presented in photograph form, in three-dimensional form mounted on boards, or when participants were allowed to touch the models while sighted or while blindfolded). In fact, the means for each sex, by condition, show a consistently decreasing difference until such time as women score higher than men, on average, in the sighted and touch condition. Thus, we also replicate the findings of Gentaz and Hatwell (1995) who show that manual rotation has no sex difference.

Both the blindfold and touch conditions introduce kinetic visuospatial processing. The descriptive statistics revealed that men did not decline in their performance, but rather that women improved. Interestingly, women's performance steadily improved across the conditions, between the two-dimensional, to the mounted, to the blindfolded and touch conditions. Even more interesting, according to the effect size statistics, condition accounted for 30% of the variance in participants' performance, which presumably indicates that the abstract stimuli are more difficult than the concrete ones for both sexes. Meanwhile, sex accounted for only 2%. From this finding, we suggest that sex difference in mental rotation is, at least in part, an artifact of the stimuli.

General Discussion

At the start of the introduction, we stated that mental rotation pertains to the ability to orient or move oneself and objects in a given environment, or to imagine such action (Moreau et al. 2010). While we do not dispute claims that individuals vary

on their ability to perform these types of actions, we argue that the clear evolutionary advantages linked to being able to orient oneself and objects in a context suggest that sex differences should be minimal. Our findings suggest that this claim has merit and that previously documented sex differences rest on stimuli artificiality and procedural issues, such as time restrictions.

The results of the present study confirm our hypothesis that more realistic stimuli tend to reduce the sex difference found on the MRT. As the stimuli became more realistic, the sex difference was increasingly reduced. Furthermore, the sex difference was eliminated in the conditions that involved handling of the actual models, as expected based on prior research involving manual rotation (Gentaz and Hatwell 1995). Together, these findings indicate that the previously documented sex difference in mental rotation, as measured by the MRT, may be an artifact of the stimuli.

In experiment 1, male and female scores for the photograph, untimed condition did not differ significantly. This result parallels that of experiment 2, where for photographs, in untimed condition, no sex difference was found. It is likely that the lack of interaction is caused by the significant drop in men's scores when presented with the photographs, whereas women's scores decreased in the timed condition and increased slightly in the untimed condition, but neither change was significant. We suggest that in the timed condition, the extra visual information provided by the photographs required additional time to process and consequently slowed down male participants and lowered their score. Conversely, it may be a lack of visual information that caused women to go slower for the drawings, and thus, they were not significantly influenced by the need for extra processing time of the photographs. This possibility needs to be tested by future research. Further, the way that responses were recorded in the photograph condition may be a potential limitation of the current study, as it introduced delays by having participants write their responses on a separate sheet.

It is known that, in general, women prefer to avoid taking risks, as compared to men (Daly and Wilson 2001), and the negative scoring of the MRT involves a form of risk. Perhaps women find it less risky to select an answer for the stimuli forms where there are multiple sources of information, such as when additional corroborating visual evidence like depth, decreased size according to distance, and slight shadow is available. It is also known that women, more than men, are influenced by feedback (Helgeson 2005), and feedback may be at least partly provided by this additional visual information. That is, women might be confirming their initial guesses at the correct items by using these cues, and thus providing feedback to themselves as to the correctness of their guess. Moreover, women have a tendency to double-check their answers (Hirnsstein et al. 2009), which may also provide feedback. This feedback presumably decreases feelings of risk, in

that one can obtain some information that they are using the right process or strategy to solve a problem, for example. Boone and Hegarty (2017) report that women, more than men, are likely to skip an item on the MRT if they cannot determine the answer, rather than consider the item in detail. Risk aversion may help explain this finding.

Studies on problem solving have shown that women tend to use algorithmic approaches, while men are more likely to use novel and creative approaches (Gallagher and De Lisi 1994). For the drawings, which are isometric, it is difficult to apply an algorithmic approach, because they are abstract and lack features that one would typically use, such as shadows or diminishing size due to distance. Thus, the first condition may prevent women from using their preferred problem solving style and instead demand novelty and risk-taking. However, the more realistic conditions, being more like situations encountered in everyday life, are less unique and more readily addressed using real-world mental rotation skills. This possibility is supported by Boone and Hegarty (2017) who suggest that "the sex difference in performance of so-called mental rotation tasks is not only in the mental rotation process but also in discovery and application of alternative solution strategies" (p. 1015).

Recently, Boone and Hegarty (2017) report that another variable that needs to be considered in future research is the type of instruction and the use of foils. They show that the sex difference is no longer evident when all foils are structural rather than mirror reflections, and when participants are trained to look for structural foils. Moreover, while both sexes use multiple strategies, men tend to outperform women when instructed to examine the overall shape, but the sex difference is removed when instructed to instead count the number of units that are used to construct the blocks. The sex difference was also entirely removed when both sexes were instructed to examine the arm direction as a strategy. The authors conclude that the MRT is measuring something other than mental rotation and that women focus on the details of the blocks (e.g., by counting units), whereas men are using more global strategies (e.g., reviewing overall shape). This global strategy may be faster to apply, and hence result in large sex differences in timed tests.

Indeed, training in strategies seems key. When confidence is manipulated, and when they are motivated to perform well and believe that they have the expertise to do so, women perform on par with men (confidence Estes and Felker 2012, motivation Moè 2016). Moè (2016) documented that women's performance after receiving an hour training session focused around motivation (to reduce the belief in male advantage in spatial tasks) and/or solution strategies was higher or on par with men's performance before training. If the sex difference in mental rotation ability was due to biological reasons, it should not be eradicated by training.

We recognize that we did not fully explore the issue of timing in that we did not have both timed and untimed conditions in experiment 2. However, the large number of participants, for both experiments and across all conditions (470), came close to using every available and willing member of the participant pool over a 3-year period. It is doubtful that we could have readily located additional 200 or more participants with similar educational and demographic backgrounds. In part, we believe that a major strength of the study is that all conditions were tested on a limited population that prevents confounding factors such as education from arbitrarily influencing the results. We acknowledge that the sample sizes in some of the conditions are not large and replication is needed.

These results have significant implications for the larger debate in psychology concerning sex differences in mental rotation ability, and they directly challenge the contention that men outperform women on mental rotation. Our findings support the prediction that factors involving the testing procedures and stimuli form have an impact on the sex difference that has been found in mental rotation. Based on our findings, as well as those of other studies, it is clear that in some conditions, a sex difference does exist. However, this difference in mental rotation ability has been magnified by the procedure and stimuli, such as using a tight time restriction and isographic line drawings. We have found evidence that there is no significant sex difference in mental rotation ability when the stimuli possess a high degree of realism, which indicates that prior research that suggests a large sex difference is not entirely accurate. Given the reviewed importance of mental rotation for daily functioning, and its presumed associated evolutionary advantages, there is no reason to expect a sex difference. What is now needed is additional research into how the stimuli affects men and women and, consequently, further exploration of the actual factors underlying the differences between men's and women's performance on the original MRT.

References

- Annett, M. (1992). Spatial ability in subgroups of left- and right-handers. *British Journal of Psychology*, *47*, 730–747.
- 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*(7), 1005–1019.
- Daly, M., & Wilson, M. (2001). Risk-taking, intrasexual competition, and homicide. *Nebraska Symposium on Motivation*, *47*, 1–36.
- Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Archives of Sexual Behavior*, *41*, 557–570.
- Fisher, M., Cox, A., & Zhao, L. (2006). Using sex differences to link spatial cognition and program comprehension. *Proceedings of the IEEE International Conference on Software Maintenance*, 289–298.
- Gallagher, A., & De Lisi, R. (1994). Gender differences in scholastic aptitude test—mathematics problem solving among high-ability students. *Journal of Educational Psychology*, *86*, 204–211.
- Gardony, A. L., Taylor, H. A., & Brunyé, T. T. (2013). What does physical rotation reveal about mental rotation? *Psychological Science*, *25*(2), 605–612. <https://doi.org/10.1177/0956797613503174>.
- Gaulin, S. J., & FitzGerald, R. W. (1989). Sexual selection for spatial learning ability. *Animal Behavior*, *37*, 322–331.
- Geary, D. C. (1995). Sexual selection and sex differences in spatial cognition. *Learning and Individual Differences*, *7*, 289–301.
- Geary, D. C., Gilger, J. W., & Elliot-Miller, B. (1992). Gender differences in three-dimensional mental rotation: a replication. *Journal of Genetic Differences*, *153*, 115–117.
- Gentaz, E., & Hatwell, Y. (1995). The haptic “oblique effect” in children's and adults' perception of orientation. *Perception*, *24*, 631–646.
- Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: the role of performance factors. *Memory & Cognition*, *18*, 546–550.
- Gur, R. C., Alsop, D., Glahn, D., Petty, R., Swanson, C. L., Maldjian, J. A., Turetsky, B. I., Detre, J. A., Gee, J., & Gur, R. E. (2000). An fMRI study of sex differences in regional activation to a verbal and spatial task. *Brain and Language*, *74*, 157–170.
- Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed.). NY: Psychology Press.
- Hedges, L. V., & Nowell, A. (1995). Gender differences in mental test scores, variability and numbers of high-scoring individuals. *Science*, *269*, 41–45.
- Helgeson, V. (2005). *Psychology of gender* (2nd ed.). Upper Saddle River: Prentice-Hall.
- Hirnstein, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences*, *19*(2), 225–228.
- Jansen, P., Zayed, K., & Osmani, R. (2016). Gender differences in mental rotation in Oman and Germany. *Learning and Individual Differences*, *51*, 284–290.
- 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. <https://doi.org/10.1016/j.intell.2006.07.009>.
- Kaushall, P., & Parson, L. M. (1981). Optical information and practice in the discrimination of three-dimensional objects. *Canadian Journal of Psychology*, *46*, 170–190.
- Kimura, D. (1999). *Sex and cognition*. Cambridge: MIT Press.
- Linn, M. C., & Peterson, A. C. (1985). Emergence and characterization of gender differences in spatial abilities: a meta-analysis. *Child Development*, *56*, 1479–1498.
- Masters, M. S. (1998). The gender difference on the Mental Rotations Test is not due to performance factors. *Memory and Cognition*, *26*, 444–448.
- Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, *23*, 337–341.
- Maccoby, E. E., & Jacklin, C. N. (1974). *Psychology of sex differences*. Stanford: Stanford University Press.
- McWilliams, W., Hamilton, C. J., & Muncer, S. J. (1997). On mental rotation in three dimensions. *Perceptual and Motor Skills*, *85*, 297–298.
- Moè, A. (2016). Teaching motivation and strategies to improve mental rotation abilities. *Intelligence*, *59*, 16–23.
- Moreau, D., Clerc, J., Mansy-Dannay, A., & Guerrien, A. (2012). Enhancing spatial ability through sport practice: evidence for an effect of motor training on mental rotation performance. *Journal of Individual Differences*, *33*, 83–88. <https://doi.org/10.1027/1614-0001/a000075>.
- Moreau, D., Mansy-Dannay, A., Clerc, J., & Guerrien, A. (2010). Academic program and mental rotation performance: evidence for

- a developmental effect on individual differences in early adulthood. *GESJ: Education Science and Psychology*, 2(17), 21–28.
- Nazareth, A., Herrera, A., & Pruden, S. M. (2013). Explaining sex differences in mental rotation: role of spatial activity experience. *Cognitive Processing*, 14(2), 201–204.
- Parsons, L. M. (1995). Inability to reason about an object's orientation using an axis and angle of rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1259–1277.
- Parsons, T. D., Larson, P., Kratz, K., Thiebaut, M., Bluestein, B., Buckwalter, J. G., & Rizzo, A. A. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia*, 42(4), 555–562.
- Pazzaglia, F., & Moè, A. (2013). Cognitive styles and mental rotation ability in map learning. *Cognitive Processing*, 14(4), 391–399. <https://doi.org/10.1007/s10339-013-0572-2>.
- Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, 57, 176–184.
- Pletzer, B. (2014). Sex-specific strategy use and global-local processing: a perspective toward integrating sex differences in cognition. *Frontiers in Neuroscience*, 8, 425. <https://doi.org/10.3389/fnins.2014.00425>.
- Resnick, S. (1993). Sex differences in mental rotations: an effect of time limits? *Brain and Cognition*, 21, 71–79.
- Robert, M., & Chevrier, E. (2003). Does men's advantage in mental rotation persist when real three-dimensional objects are either felt or seen? *Memory and Cognition*, 31, 1136–1145.
- Saucier, D. M., McCreary, D. R., & Saxberg, J. K. J. (2002). Does gender role socialization mediate sex differences in mental rotations? *Personality and Individual Differences*, 32, 1101–1111.
- Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., & Olshansky, E. (2000). Evolved mechanisms underlying wayfinding: further studies on the hunter-gatherer theory of spatial sex differences. *Evolution and Human Behavior*, 21(3), 201–213.
- Silverman, I., & Phillips, K. (1993). Effects of estrogen changes during the menstrual cycle on spatial performance. *Ethology and Sociobiology*, 14, 257–270.
- Symons, D. (1979). *The evolution of human sexuality*. Oxford: Oxford University Press.
- Turkle, S., & Papert, S. (1990). Epistemological pluralism: styles and voices within the computer culture. *Signs*, 16, 128–157.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599–604.
- Vashro, L., & Cashdan, E. (2015). Spatial cognition, mobility, and reproductive success in northwestern Namibia. *Evolution and Human Behavior*, 36(2), 123–129.
- Voyer, D. (1997). Scoring procedure, performance factors and magnitude of sex differences in spatial performance. *American Journal of Psychology*, 110, 259–276.
- 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.
- Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles*, 43, 891–915.
- Voyer, D., & Saunders, K. (2004). Gender differences on the mental rotations test: a factor analysis. *Acta Psychologica*, 117, 79–94.
- 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.
- Wohlschälger, A., & Wohlschälger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 397–412.