

Confidence Mediates the Sex Difference in Mental Rotation Performance

Zachary Estes · Sydney Felker

Received: 5 March 2011 / Revised: 29 September 2011 / Accepted: 1 October 2011 / Published online: 1 December 2011
© Springer Science+Business Media, LLC 2011

Abstract On tasks that require the mental rotation of 3-dimensional figures, males typically exhibit higher accuracy than females. Using the most common measure of mental rotation (i.e., the Mental Rotations Test), we investigated whether individual variability in confidence mediates this sex difference in mental rotation performance. In each of four experiments, the sex difference was reliably elicited and eliminated by controlling or manipulating participants' confidence. Specifically, confidence predicted performance within and between sexes (Experiment 1), rendering confidence irrelevant to the task reliably eliminated the sex difference in performance (Experiments 2 and 3), and manipulating confidence significantly affected performance (Experiment 4). Thus, confidence mediates the sex difference in mental rotation performance and hence the sex difference appears to be a difference of performance rather than ability. Results are discussed in relation to other potential mediators and mechanisms, such as gender roles, sex stereotypes, spatial experience, rotation strategies, working memory, and spatial attention.

Keywords Confidence · Gender roles · Mental rotation · Sex differences · Spatial abilities · Stereotype threat and lift

Introduction

Of all cognitive sex differences, the mental rotation of abstract figures in 3-dimensional space is the most robust (Halpern, 2000; Hines, 2004; Linn & Petersen, 1985; Maccoby & Jacklin, 1974). Females typically respond less accurately and more slowly on mental rotation tasks than do males (Lippa, Collaer, & Peters, 2010; Lohman, 1986; Maylor et al., 2007; Peters, 2005; Voyer, Voyer, & Bryden, 1995), though the variability within each sex is greater than the difference between sexes (Kail, Carter, & Pellegrino, 1979; Resnick, 1993). Meta-analyses indicate a medium effect size of this sex difference in mental rotation across age groups (Cohen's $d = .73$, Linn & Petersen, 1985) and among adults more specifically ($d = .66$, Voyer et al., 1995). The single largest study of mental rotation ($N = 255,100$) also revealed a medium effect ($d = .53$, Peters, Manning, & Reimers, 2007). Given the complexity of the task and the magnitude of the sex difference, it likely has multiple causes or mediators. Purely biological explanations (for review, see Kimura, 1999) have received little empirical support, with no clear relationship between mental rotation ability and endogenous levels of sex hormones either prenatally (Collaer & Hines, 1995; Hines et al., 2003; Rahman, Wilson, & Abrahams, 2004) or in adulthood (Halari et al., 2005). Moreover, a biological constraint on mental rotation ability would not preclude mediation of performance by sociocognitive factors (e.g., Casey, 1996; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). Here, we examined whether one such sociocognitive factor, namely participants' confidence, contributed to this sex difference in mental rotation performance. Although this presumed relation between confidence and mental rotation performance has received little empirical attention, related research on gender roles, sex stereotypes, and stereotype threat provides a rich source of supportive evidence.

Z. Estes (✉)
Department of Marketing, Bocconi University, Via Roentgen 1,
20136 Milan, Italy
e-mail: estes@unibocconi.it

S. Felker
Counseling and Psychiatric Services, University of Georgia Health
Center, Athens, GA, USA

Sex Stereotype Effects

Gender role beliefs and traits may partially explain the sex difference in mental rotation performance. Individuals who hold traditional beliefs about gender roles and who engage in gender-typical behaviors might believe in the common stereotype that men are superior to women at spatial skills. This belief might then induce or accentuate the sex difference in mental rotation performance. Indeed, people are generally aware of the stereotype that females have poorer spatial and mathematical abilities than males and, in fact, nearly half of all females endorse this stereotype to some extent (Blanton, Christie, & Dye, 2002). Females perform better on a mental rotation task when asked to imagine themselves as a stereotypical male than as a stereotypical female ($d = .56$, Ortner & Sieverding, 2008) and, more generally, mental rotation ability is associated with more masculine gender role traits ($r = +.32$) and less feminine gender role traits ($r = -.26$, Saucier, McCreary, & Saxberg, 2002; see also Signorella, Jamison, & Krupa, 1989). Performance on spatial tasks thus is clearly related to gender role beliefs and traits.

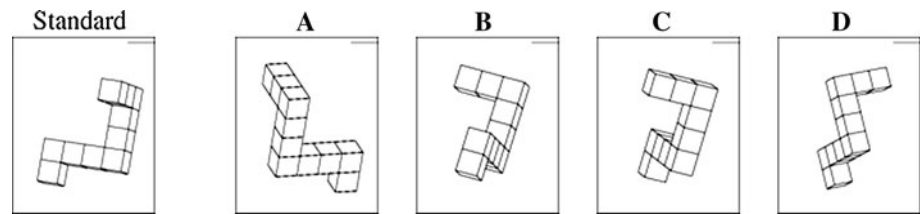
Mental rotation performance may also be affected by mere awareness of, rather than belief in, the stereotype that men are superior to women on spatial tasks. Stereotype threat is the tendency for members of a negatively stereotyped group to underperform on tasks relevant to the stereotype (e.g., Steele, 1997; Steele & Aronson, 1995; for review, see Maass & Cadinu, 2003; Schmader, Johns, & Forbes, 2008). In this case, the stereotype that women have poor spatial skills could induce stereotype threat in women, thereby accentuating any decrement in mental rotation performance that may or may not occur otherwise. McGlone and Aronson (2006) directly tested whether stereotype threat affected mental rotation performance. Prior to completing a mental rotation task, males and females at a private university identified themselves via a series of questions. Some participants answered questions about their gender. If stereotype threat affects mental rotation performance, then the sex difference should be observed in this condition. Other participants answered questions about attending a private university. This condition, which highlighted participants' scholastic achievement, should attenuate the sex difference in mental rotation. These predictions were supported. In fact, females performed better when identified as a "private college student" than when identified as a female ($d = 1.38$). Evidently, activating females' achieved scholastic identity alleviated the stereotype threat and allowed them to perform to their potential. In contrast, males performed better when identified as a male than as a private college student ($d = .88$). This result indicates a complementary effect, stereotype lift (Walton & Cohen, 2003), which is an improvement in performance due to knowledge that an outgroup is negatively stereotyped (see also Shih, Ambady, Richeson, Fujita, & Gray, 2002; Shih, Pittinsky, & Ambady, 1999). In this case, the stereotype that men are superior at spatial tasks may

bolster their confidence and subsequently improve their performance.

Rather than manipulating participants' salient identity (e.g., McGlone & Aronson, 2006), Moe and Pazzaglia (2006) manipulated the stereotype itself. They first had participants complete a block of mental rotation trials, then they informed participants either that men were better or that women were better at the task, and finally they had those same participants complete another block of mental rotation trials. Women performed significantly worse after being told that men were better at the task ($d = .44$) and significantly better after being told that women were better ($d = .35$). Conversely, men performed significantly better after being told that men were better at the task ($d = .80$) and significantly worse after being told that women were better ($d = .78$; see also Wraga, Duncan, Jacobs, Helt, & Church, 2006). Massa, Mayer, and Bohon (2005) also manipulated the sex stereotype, and additionally examined its interaction with gender role beliefs. They found that women with masculine gender role beliefs scored higher on a spatial task when they were told that it measured spatial skills than when told that it measured empathy ($d = 1.85$), whereas women with feminine gender role beliefs scored higher when told that it measured empathy than when told that it measured spatial skills ($d = 1.31$). Lippa et al. (2010) examined mental rotation and line angle judgments across 53 nations that varied in egalitarianism. Males outperformed females in every nation in both mental rotation (mean $d = .47$) and line angle judgment (mean $d = .49$) and, somewhat surprisingly, these sex differences were larger in highly egalitarian nations (e.g., Norway) than in less egalitarian nations (e.g., Pakistan; mental rotation $r = +.47$; line angle $r = +.41$). Lippa et al. attributed this finding to greater awareness of sex stereotypes and/or greater susceptibility to stereotype threat in egalitarian nations. So, in summary, beliefs about and awareness of sex stereotypes are both related to the sex difference in mental rotation performance. But how exactly might sex stereotypes affect performance?

Confidence as a Potential Mediator

Much of the research on gender role and sex stereotype effects assumes confidence as a potential cognitive mechanism by which those social factors exert their effect. For instance, Steele (1997) explained stereotype threat in terms such as self-regard, self-efficacy, and self-confidence. Walton and Cohen (2003) similarly explained stereotype lift thus: "By comparing themselves with a socially devalued group, people may experience an elevation in their self-efficacy...[which] may be important to maintaining confidence and motivation" (p. 456). The belief that one (or one's social group) is skilled or poor at a given task may well affect one's confidence when approaching that task and this effect on confidence may have cascading effects on

Fig. 1 A trial of the MRT

basic cognitive skills, such as attention, memory, and judgment (e.g., Schmader et al., 2008), which ultimately would affect performance.

In their seminal study, Steele and Aronson (1995) demonstrated that describing a difficult verbal test as diagnostic of intellectual ability significantly increased self-doubt and decreased performance among Black students but not among White students. Stereotype threat also increased negative performance-related thoughts among women, and these negative thoughts mediated performance on a math test (Beilock, Rydell, & McConnell, 2007; Cadinu, Maass, Rosabianca, & Kiesner, 2005; Schmader, Forbes, Zhang, & Mendes, 2009). And, conversely, women who self-affirmed another valued trait, such as their creativity or humor, exhibited no stereotype threat effect on a math test (Martens, Johns, Greenberg, & Schimel, 2006). Self-doubt, negative thoughts, and self-affirmation are closely related to the more general construct of confidence. Thus, sex stereotypes are widely thought to affect performance on cognitive tasks indirectly, partially by influencing participants' confidence. We therefore tested whether confidence mediated mental rotation performance.

Preliminary evidence suggests that confidence might indeed underlie the sex difference in mental rotation. Females reported less confidence than males on cognitive tasks in general (Beyer & Bowden, 1997; Maccoby & Jacklin, 1974) and on mental rotation tasks in particular ($d = 1.04$, Cooke-Simpson & Voyer, 2007; see also Pallier, 2003). Moreover, because confidence is gauged before the judgment is made (Baranski & Petrusic, 1998), it may affect that judgment (Petrusic & Baranski, 2003). Indeed, confidence has been shown to predict performance on other cognitive tasks, such as mathematical problem solving (Casey, Nuttall, & Pezaris, 1997; Schmader et al., 2009) and semantic categorization (Estes, 2004; Pasterski, Zwierzynska, & Estes, 2011). So given these sex differences in confidence and mental rotation, and given that confidence mediates performance on some cognitive tasks, confidence might mediate the sex difference in mental rotation. This presumed relation between confidence and spatial abilities, however, has not been thoroughly explored and the results are mixed. Gonzales, Blanton, and Williams (2002) found that participants' self-evaluations of task competence did not predict their scores on a test of mathematical and spatial abilities. In contrast, Cooke-Simpson and Voyer (2007) found that participants' confidence ratings reliably predicted their mental rotation scores ($r = +.69$), with

more confident men and women outperforming their less confident peers.

The most common measure of mental rotation performance is the Mental Rotations Test (MRT; Vandenberg & Kuse, 1978), which is based on the 3-dimensional block figures introduced by Shepard and Metzler (1971) and updated by Peters et al. (1995). Each trial of the MRT consists of one standard and four alternative figures. Exactly two of the alternatives are rotated versions of the standard; the other two are non-matching distracters (see Fig. 1). Thus, each trial has two correct answers. Critically, though, participants may omit one or both responses. Females tend to provide fewer responses than males ($d = .30$, Voyer, Rodgers, & McCormick, 2004; see also Voyer & Saunders, 2004), but statistically correcting for differential response rates does not eliminate the sex difference in performance (Masters, 1998; Peters, 2005; Peters et al., 1995; Resnick, 1993). Nevertheless, the mere possibility of omitting responses supports the hypothesis that confidence mediates the sex difference in accuracy on the MRT. The decision to respond or abstain on any given trial of the MRT is presumably based on one's confidence in knowing the correct response. Because confidence appears to be an important component of the MRT, it stands as a plausible mediator of performance.

The present study therefore investigated whether confidence mediated accuracy on this most common and robust measure of mental rotation performance. Specifically, if confidence mediates mental rotation, then (1) confidence should predict mental rotation scores not only between sexes, but also within sexes, (2) rendering confidence irrelevant to the task should attenuate the sex difference, and (3) manipulating participants' confidence should affect their mental rotation performance. We tested these three predictions across four experiments.

Experiment 1

In Experiment 1, we tested whether confidence predicted mental rotation performance between sexes, within each sex, and within individuals. Participants completed a standard MRT, but they additionally rated their confidence in each response. Following common procedure, the MRT was administered under a time constraint of 15 s per trial, which allows greater experimental control without influencing the magnitude of the sex difference (Peters, 2005; Voyer et al., 2004).

Cooke-Simpson and Voyer (2007) provided tentative evidence that confidence predicted MRT performance, but that study had several critical limitations. As described above, each item on the MRT includes four alternative figures, exactly two of which are rotated versions of the standard. Participants in Cooke-Simpson and Voyer's study only rated their confidence in each *item*, which included between zero and two *responses* (depending on how many responses the participant omits on a given item). Unfortunately, this methodology likely decreased the accuracy of participants' confidence ratings. That is, because participants provided confidence ratings for each *pair* of responses, participants must not only gauge their confidence but also choose a decision rule and apply the chosen computation to determine their confidence rating. Should the confidence rating be an average over the two responses, should it be the minimum of the two or should it be the maximum? Conversely, this methodology also limits the precision of possible conclusions from the research: Because participants provided a single confidence rating for each pair of responses, it is unclear to which response a given rating refers. Finally, Cooke-Simpson and Voyer examined the relation between confidence and performance only across individuals. That is, they calculated each participant's mean confidence rating and overall accuracy score, and they tested whether highly confident individuals outperformed less confident individuals. While such an analysis is informative, it fails to test the potential relation between confidence and performance within an individual.¹ Is a given participant more likely to respond correctly when she is highly confident than when she is less confident?

To address these limitations, in Experiment 1, we required participants to rate their confidence after each individual response (rather than after each pair of responses). By removing the complexity and ambiguity of judging confidence over multiple responses, this procedure may elicit more accurate ratings and more precise conclusions. Furthermore, in addition to examining the relation between confidence and performance across individuals, Experiment 1 also examined this relation within individuals. This allowed us to investigate not only whether highly confident individuals tend to outperform less confident individuals (as in

¹ Cooke-Simpson and Voyer (2007) also computed two additional measures of the relation between confidence and accuracy (i.e., Brier scores and "confidence relative to performance"), but again their analyses were at the level of the individual rather than the individual response. That is, for each participant they calculated the mean confidence and the mean accuracy, *collapsed across trials*, and then they compared the group means. They did not examine the relation between confidence and accuracy *on each trial*. Moreover, both of their additional measures required the assumption that confidence ratings on a 1-to-7 scale map directly and evenly onto a probability scale. For instance, those measures assumed that a confidence rating of 1 indicated a probability judgment of 0. This assumption is questionable. Presumably, if a participant believes that there is zero chance that his response is correct, then he would either omit that response or else change it to a different response. The important point for our purposes here is that Cooke-Simpson and Voyer did not examine the relation between confidence and accuracy on a trial-by-trial basis, as we did in Experiment 1.

Cooke-Simpson & Voyer, 2007), but also whether a given participant was more likely to respond correctly when she was highly confident than when she was less confident. That is, we examined the relation between confidence and performance on a trial-by-trial basis. If confidence mediates mental rotation performance, then confidence ought to predict accuracy on the MRT across sexes, within each sex, and possibly even within individuals.

Method

Participants

All participants in each of the experiments reported herein were undergraduates at a large North American university, most were between the ages of 17 and 23 years, all received partial course credit for participation, and none participated in more than one of the experiments. Seventy undergraduates (35 females, 35 males) participated in Experiment 1.

Measure

The 24-item version of the MRT was used. Following standard procedures for administration of the MRT, participants were informed that each standard figure had two matching alternatives, and they were instructed not to respond unless they were sure of the answer (see Voyer & Saunders, 2004).

Procedure

In each of the present experiments, participants were tested individually in a sound attenuated room and the entire experiment (including the instructions) was administered via computer. Each trial of Experiment 1 began with presentation of the standard and four alternatives aligned horizontally onscreen (see Fig. 1). The figures remained onscreen for 15 s and were then replaced with the prompt "Please enter your first choice." Participants either pressed the A, B, C or D key or else pressed the spacebar if they were unsure of the answer. Participants were then asked "How confident are you in this choice?" The numbers 1 ("not at all") through 7 ("extremely") were the only valid choices. The prompt "Please enter your second choice" then appeared, followed again by "How confident are you in this choice?" On trials where the participant omitted a response (by pressing the spacebar), they were instructed to press any number for the confidence rating and those ratings were excluded from all analyses. After the second confidence rating had been entered, there was a 1 s inter-trial interval prior to presentation of the next set of figures. Four practice trials preceded the 24 experimental trials, which were presented in random order. After completion of all experimental trials, participants were prompted to press the M key or the F key to indicate that they were male

or female, respectively, and then to enter their age into a text-box.

Analyses

For each of the experiments, we adopted the relatively strict criterion that any participant who was more than 2.5 SD beyond the group mean for any of the measures was considered an outlier and was, therefore, excluded from analyses. In the present experiment, this led to the exclusion of three males.

Accuracy was defined as the total percent correct (i.e., number correct/number possible). Scores were also corrected for individual differences in response rate by excluding omitted trials (i.e., number correct/number responses). Corrected scores were highly correlated with total scores, $r(67) = +.86, p < .001$. Some researchers score a trial as correct only if two correct responses are provided. These scores were also highly correlated with the total scores, $r(67) = +.97, p < .001$. Given these high intercorrelations (see also Masters, 1998; Resnick, 1993; Voyer et al., 1995) and for the sake of simplicity, hereafter we report only the total percent correct.

Results and Discussion

As expected, an independent samples *t*-test revealed that males ($M = 5.61, SD = 1.02$) were more confident than females ($M = 4.62, SD = 1.41$), $d = .74, t(65) = 3.26, p < .01$. The effect size of this sex difference in confidence was comparable to that observed in prior research ($d = 1.04$; Cooke-Simpson & Voyer, 2007). As shown in Table 1, males were also significantly more accurate than females, $t(65) = 2.44, p < .05$, and the effect size ($d = .58$) was comparable to that observed in other studies ($d = .66$, Voyer et al., 1995). Males ($M = 6, SD = 8$) and females ($M = 8, SD = 8$) did not differ significantly in the percentage of responses omitted, $t(65) = 1.10$.

As illustrated in Fig. 2, confidence predicted accuracy across both sexes, $r(67) = +.56, p < .001$, and the strength of this

Table 1 Accuracy (% correct) as a function of sex (Experiments 1–4)

Experiment	Condition	Females			Males			<i>d</i>
		<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
1	Omission	35	68	12	32	75	12	.58*
2	Omission	41	63	12	43	73	14	.72**
	Commission	42	73	12	43	75	12	.19
3	Commission	39	80	13	39	80	15	.03
	Confidence	35	75	17	30	93	6	1.16***
4	High confidence	35	81	13	35	90	11	.74**
	Low confidence	35	77	15	34	85	16	.45†

d effect size (Cohen's *d*) of the sex difference in accuracy. Participants were given 15 s per trial in Experiments 1, 2 and 4, and were given unlimited time in Experiment 3

† $p < .06$; * $p < .05$; ** $p < .01$; *** $p < .001$

correlation was comparable to that observed in prior studies ($r = +.69$; Cooke-Simpson & Voyer, 2007). Confidence also predicted accuracy among females, $r(35) = +.45, p < .01$, and males $r(32) = +.62, p < .001$. The slope of this relationship did not differ between males and females, $z = .94$. This suggests that confidence may contribute not only to the sex difference in mental rotation performance, but also to individual differences within each sex (see also Cooke-Simpson & Voyer, 2007). Confidence was unrelated to the percentage of omitted trials, $r = -.14$. Rather, omissions were negatively correlated with accuracy, $r(67) = -.47, p < .001$; the fewer responses a participant provided, the lower the overall score.

Figure 2 illustrates the relationship of confidence and accuracy across participants; individuals who were highly confident also tended to be highly accurate. Figure 3 illustrates the relationship of confidence and accuracy within participants. That is, for each participant, we calculated the accuracy at each point on the confidence scale (i.e., 1–7). Because most participants used a restricted range of confidence ratings, and because the ratings were not uniformly distributed across the scale, there were unequal numbers of observations across the confidence points. Despite this limitation, a linearly increasing function was observed between confidence and accuracy for both males and females. That is, as one's confidence increased, so did one's accuracy (see Fig. 3). This result indicates that participants were successful at gauging their own relative performance on individual trials of the MRT. So, together, Figs. 2 and 3 reveal that males and females were similarly successful at calibrating their confidence to their accuracy (Fig. 3), though males tended toward the upper part of the distribution on both confidence and accuracy (Fig. 2).

These analyses establish a relationship among sex, confidence, and accuracy. However, they do not discriminate between alternative models of mediation. The question of interest is whether confidence mediated the effect of sex on accuracy, as

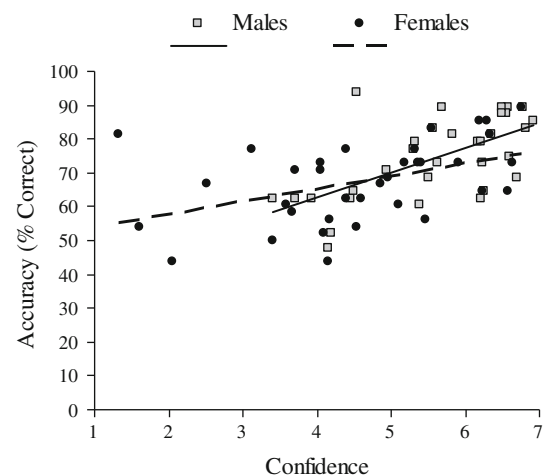


Fig. 2 Accuracy as a function of confidence across participants, Experiment 1

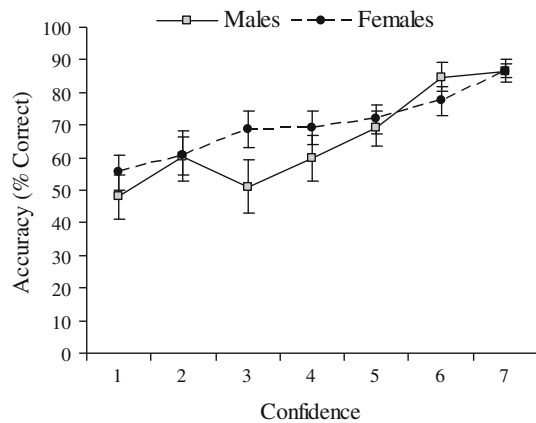


Fig. 3 Accuracy as a function of confidence within participants, Experiment 1

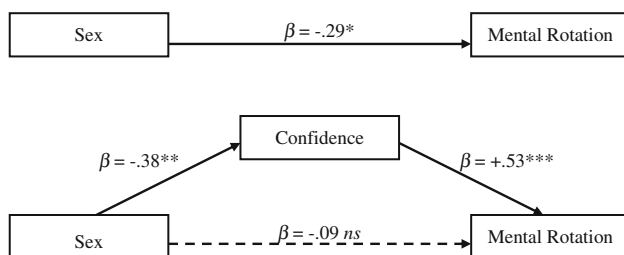


Fig. 4 Confidence as a mediator of the sex difference in mental rotation performance, Experiment 1. * $p < .05$; ** $p < .01$; *** $p < .001$

we hypothesized, or whether accuracy mediated the effect of sex on confidence. We tested these alternative models via Baron and Kenny's (1986) regression method for simple mediation. The mediation model is illustrated in Fig. 4. For these regressions, the sex factor was coded as 1 for male and 2 for female, so that negative coefficients indicate lower scores for females. As established above, sex negatively predicted both confidence ratings, $\beta = -.38$, and mental rotation scores, $\beta = -.29$, whereas confidence positively predicted mental rotation scores, $\beta = +.56$. Most critically, when sex and confidence were both included as predictor variables in the same regression, confidence remained a strong positive predictor of mental rotation, $\beta = +.53$, $t(65) = 4.73$, $p < .001$, whereas sex failed to predict mental rotation scores, $\beta = -.09$, $t(65) < 1$. Stated differently, when the effect of confidence was taken into account, the sex difference in mental rotation scores was eliminated. That confidence reduced the predictive validity of the sex factor to nearly zero (i.e., $\beta = -.09$) indicates not only that confidence mediated the sex difference in mental rotation performance, but indeed that this mediation was nearly complete. Finally, a reverse mediation analysis tested the alternative model that mental rotation scores mediated the effect of sex on confidence. With sex and mental rotation scores as the predictor variables and confidence as the criterion, sex significantly predicted confidence, $\beta = -.23$, $t(65) = 2.23$, $p < .05$. Thus, mental rotation performance did not mediate the sex difference in confidence, but rather confidence

strongly mediated the sex difference in mental rotation performance.

In conclusion, Experiment 1 corroborated the finding that confidence predicted mental rotation performance both across and within sexes (Cooke-Simpson & Voyer, 2007). Experiment 1 further demonstrated, for the first time, that confidence predicted mental rotation performance within individuals: Participants were more accurate on trials for which they were more confident. These results thus provide the most precise evidence to date of the relation between confidence and mental rotation. Finally, Experiment 1 also provided the first evidence of the direction of this relationship: Mediation analyses revealed that confidence mediated the sex difference in mental rotation performance whereas mental rotation performance did not mediate the sex difference in confidence. Nonetheless, such mediation analyses provide only indirect evidence of the nature of this relationship. Experimental manipulations of confidence would provide more direct evidence of its relation to mental rotation performance. Experiments 2–4 thus manipulated confidence and examined its influence on mental rotation performance.

Experiment 2

In Experiment 2, we sought to attenuate the sex difference in mental rotation performance by rendering confidence irrelevant to the task. One group of participants completed the standard MRT, in which they were permitted to omit trials at their discretion ("omission" group). This condition was identical to the preceding experiment, except that confidence ratings were not collected. Another group of participants also completed the MRT, but were required to respond on every trial ("commission" group). Our rationale was that when participants may omit trials, then one's confidence on each trial determines whether to respond (commit) or abstain (omit), and hence confidence is highly relevant to the task. This omission group should therefore replicate the sex difference in mental rotation that was observed in Experiment 1 and elsewhere. In contrast, when omissions are not permitted, the efficacy of evaluating one's confidence is eliminated, and hence we hypothesized that confidence would have a diminished effect on performance. So if indeed confidence contributes to the sex difference in mental rotation, then requiring a response on every trial should attenuate that difference by rendering confidence irrelevant to the task.

Alternatively, requiring participants to respond on every trial could conceivably render participants' confidence even more salient, in which case this commission group might exhibit an even larger sex difference in mental rotation than that observed in the omission group. Thus, if confidence affects mental rotation, then we should observe an interaction such that the omission group should exhibit a sex difference that is either larger or smaller than that of the commission group. Critically, an interaction in either direction would suggest that confidence

mediates mental rotation. If confidence was unrelated to mental rotation, then the sex difference should be equivalent across groups (i.e., no interaction should occur).

Method

A total of 174 undergraduates (85 females, 89 males) participated. Three outlying males and two outlying females were excluded from all analyses on the basis of the criteria established above. Materials were identical to those of Experiment 1. For participants in the omission group, the procedure was identical to that of Experiment 1, with the exception that confidence ratings were not collected. The procedure of the commission condition was identical to the omission condition, except that participants were instructed to provide two responses on each trial. They were instructed to provide their best guess if they were unsure of an answer.

Results and Discussion

Results are summarized in Table 1. The sex difference in accuracy was replicated in the omission condition but not in the commission condition. A 2 (Sex) by 2 (Condition) analysis of variance (ANOVA) confirmed a significant interaction in accuracy, $F(1, 165) = 4.19, p < .05$. That is, in the omission group, males significantly outperformed females, $t(82) = 3.54, p < .01$. The effect size ($d = .72$) was comparable to other studies that have used this standard “omission” instruction with the MRT ($d = .66$; Voyer et al., 1995). In contrast, males and females in the commission group did not differ in accuracy, $t(83) < 1$. To examine the reliability of this null sex difference in the commission group, we conducted post hoc power analyses (see Faul, Erdfelder, Lang, & Buchner, 2007). In their meta-analysis of 42 published studies of the sex difference on the MRT among participants over 18 years of age, Voyer et al. (1995, Table 4) found an effect size of $d = .66$. In Experiment 1 and in the omission group of Experiment 2, we obtained similar effect sizes of .58 and .72, respectively (see Table 1). We therefore adopted .66 as our estimate of effect size. Using the standard alpha of .05, the achieved power in the commission group was .92, where power of .80 or higher is typically considered good. Thus, despite good statistical power to detect a sex difference on the MRT, no such difference was observed in the commission group.

We suggest that the mere possibility of omitting a response renders confidence efficacious, because presumably the decision to respond or omit was based on confidence. But when required to respond, confidence was no longer efficacious and hence its effect was attenuated. However, females ($M = 14\%$, $SD = 12$) also omitted more responses than males ($M = 7\%$, $SD = 7$), $d = .63, t(82) = 3.04, p < .01$, and this effect size was somewhat larger than that observed in prior research ($d = .30$, Voyer et al., 2004). The sex difference in accuracy thus could

be attributable to omissions rather than confidence per se. We therefore held omissions constant via analysis of covariance, and the sex difference in accuracy remained significant, $F(1, 81) = 4.86, p < .05$. Thus, the sex difference in mental rotation was attributable to confidence rather than omissions.

Experiment 3

Experiment 3 provided a further test of whether the sex difference in performance is better explained by confidence or by omissions. One group of participants replicated the commission condition of Experiment 2 (“commission” group). Another group was also required to respond on every trial but, critically, they also provided confidence ratings on each trial (“confidence” group). If the sex difference in performance is due to omissions, then neither group should exhibit a sex difference, because both groups were disallowed from omitting responses. Alternatively, if the sex difference is due to confidence, then only the confidence group should exhibit a sex difference, because confidence is irrelevant to the commission group. To generalize the results across timing conditions, participants were given unlimited time to complete each trial (see Lohman, 1986; Masters, 1998; Peters, 2005; Voyer et al., 2004).

Method

A total of 148 undergraduates (76 females, 72 males) participated. Three outlying males and two outlying females were excluded from all analyses on the basis of the criteria established above. Materials were identical to those of Experiment 1. For participants in the commission group, the procedure was identical to the commission condition of Experiment 2, except that participants were given unlimited time to complete each trial. Thus, the stimuli remained onscreen until the participant provided two responses. The procedure of the confidence condition was identical, except that confidence ratings were also collected after each response, as in Experiment 1.

Results and Discussion

Results are summarized in Table 1. Relative to Experiment 2, the unlimited time allowed on each trial in Experiment 3 appears to have increased accuracy (see also Peters, 2005). But, most importantly, the confidence group replicated the sex difference in accuracy that was obtained in Experiment 1 whereas the commission group replicated the null sex difference that was obtained in Experiment 2. A 2 (Sex) by 2 (Condition) ANOVA confirmed a significant interaction in accuracy, $F(1, 139) = 15.94, p < .001$. That is, males and females in the commission group did not differ in accuracy, $t(76) < 1$, despite good statistical power to detect such a difference (with $d = .66$ and $\alpha = .05$, power = .89). In the confidence group, however, males were

significantly more accurate than females, $t(63) = 5.72, p < .001$. This sex difference was large ($d = 1.16$), but within the normally observed range of effect sizes on this task (Voyer et al., 1995). Males ($M = 6.35, SD = .63$) were also more confident than females ($M = 5.19, SD = 1.19$), $d = 1.04, t(63) = 4.84, p < .001$, and the effect size was comparable to that observed in prior studies ($d = 1.04$, Cooke-Simpson & Voyer, 2007). Thus, when confidence was irrelevant to the task (i.e., commission group), the sex difference in accuracy was eliminated. But when confidence was reinstated as relevant to the task (i.e., confidence group), the sex difference in accuracy re-emerged. Experiment 3 therefore supported the hypothesis that mental rotation is mediated by confidence.

In fact, as illustrated in Fig. 5, confidence ratings predicted accuracy across sexes, $r(65) = +.70, p < .001$; individuals who were highly confident were also highly accurate. The strength of this correlation was comparable to that observed in prior studies ($r = +.69$, Cooke-Simpson & Voyer, 2007). The correlation was also significant among females, $r(35) = +.63, p < .001$, but not among males, $r = +.23$, and these slopes differed significantly, $z = 1.94, p = .05$. This differential relation between confidence and accuracy might indicate that, under these circumstances, females gauged their own relative performance more accurately than males did. However, no difference in calibration was observed in Experiment 1. The lack of correlation among males in the present study is more likely attributable to highly restricted ranges of both confidence and accuracy (see Fig. 5). Indeed, relative to Experiment 1, both the confidence and the accuracy of both males and females increased substantially. This systematic increase was most likely due to the unlimited time allowed for responding in Experiment 3, compared to 15 s per trial in Experiment 1. As in Experiment 1, we additionally examined the relationship of confidence and accuracy within participants. For both males and females, accuracy was a linearly increasing function of confidence (see Fig. 6). This

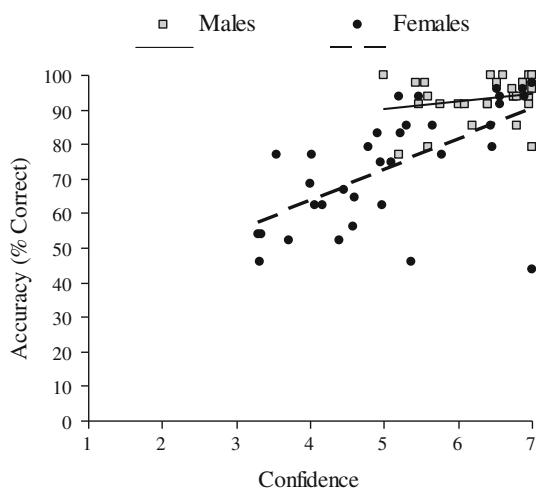


Fig. 5 Accuracy as a function of confidence across participants, Experiment 3

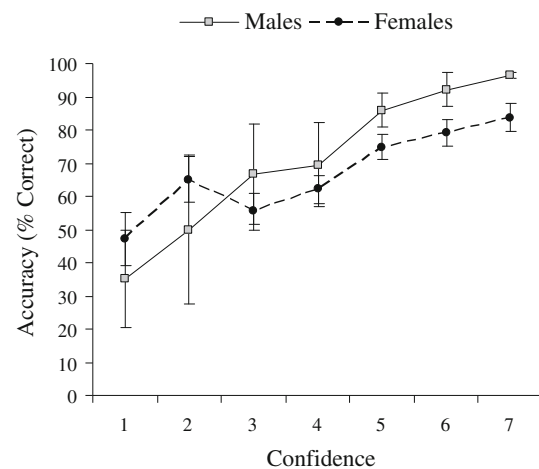


Fig. 6 Accuracy as a function of confidence within participants, Experiment 3

result replicates the general pattern observed in Experiment 1 (Fig. 3), and indicates that both males' and females' confidence was indeed calibrated to their performance.

Experiment 4

In Experiment 4, we manipulated participants' confidence prior to administration of the MRT. All participants first completed a line judgment task that was intentionally difficult, so that participants would be unable to gauge their performance. On each trial of this task, one line was presented in a vertical orientation and another line was presented horizontally. The task was to judge whether the two lines were of the same length. As expected, performance on this task was near chance ($M = .56, SE = .01$), with males and females performing equally poorly, $t(137) < 1$.

Upon completion of the line judgment task, participants were randomly informed that their performance on the line judgment task was either above average ("high confidence" condition) or below average ("low confidence" condition). All participants then immediately completed the standard (omission) version of the MRT, with 15 s per trial. Because the line judgment task required comparison of lines at different orientations, we assumed that participants would interpret the evaluation of their performance on this task as relevant to the subsequent MRT. If confidence mediates mental rotation performance, then participants in the high confidence condition should outperform their counterparts in the low confidence condition.

Method

Participants

A total of 153 undergraduates (76 females, 77 males) participated. Eight outlying males and six outlying females were

excluded from all analyses on the basis of the criteria established above.

Measures

The line judgment task consisted of three standard lines and five alternative lines for each standard. The “large” standard was 4.5 in., the “medium” was 33% shorter (3.02 in.), and the “small” was 67% shorter than the “large” (1.49 in.). For each of the three standards, one alternative was identical, one was 10% shorter, one was 5% shorter, one was 5% longer, and one was 10% longer. Materials of the MRT were identical to those of Experiment 1.

Procedure

Participants were initially informed that the experiment would consist of two parts and that the first part was a line judgment task. On each trial of the line judgment task, the standard was presented in vertical orientation. The alternative always appeared in horizontal orientation at the midpoint of the standard, separated horizontally by 1 in. Each standard was presented eight times. On four of those presentations, the alternative was of the same length. On the other four presentations, the alternative was of a different length (i.e., 10% shorter, 5% shorter, 5% longer, or 10% longer). The lines were presented in the upper half of the computer display. After 2 s, the prompt “Same (S) or different (D)?” appeared in the lower part of the display while the lines remained onscreen. Participants indicated by keypress whether the lines were of the same length.

Participants were randomly assigned to either a high confidence or a low confidence condition. At the conclusion of the line judgment task, the following statement was presented in the upper part of the display for 3 s: “EVALUATION: Your performance indicates that you are...” After 3 s, participants in the high confidence condition saw “ABOVE AVERAGE” in the middle of the display whereas participants in the low confidence condition saw “BELOW AVERAGE.” After a 1 s delay, “on the line judgment task” appeared in the lower part of the screen. The entire statement of evaluation remained onscreen simultaneously for three additional seconds. Finally, the prompt “Please press the spacebar to proceed to the second part of the experiment” appeared near the bottom of the display. The procedure of the MRT was identical to that of the omission condition of Experiment 2.

Results and Discussion

Results are summarized in Table 1. A 2 (Sex) by 2 (Condition) ANOVA confirmed an overall sex difference in accuracy, $F(1, 135) = 12.59, p < .001$, with males again outperforming females. The effect sizes in the low and high confidence conditions were

comparable to those observed in prior studies ($d = .66$, Voyer et al., 1995). The interaction was not significant, $F(1, 135) < 1$. Most critically, there was a significant main effect of condition on accuracy, $F(1, 135) = 3.79, p = .05$, with participants in the high confidence group significantly outperforming those in the low confidence group. That is, manipulating participants’ confidence affected their mental rotation: Participants scored higher on the MRT after being randomly informed that they were above average on a line judgment task than after being informed that they were below average on the line judgment task. Notably, females in the high confidence group and males in the low confidence group did not differ in accuracy, $t(68) = 1.09$, despite good statistical power to detect such a difference (with $d = .66$ and $\alpha = .05$, power = .86).

A 2 (Sex) by 2 (Condition) ANOVA on the percentage of omitted responses yielded no main effect of condition, $F(1, 135) < 1$, and no interaction, $F(1, 135) < 1$, indicating that the effect of condition on accuracy was not attributable to a difference in omissions between conditions. However, as in Experiment 2, there was an overall sex difference in omissions, $F(1, 135) = 16.74, p < .001$. Females ($M = 5, SD = 6$) again omitted more responses than males ($M = 1, SD = 2$), and the effect size ($d = .66$) was comparable to that observed in Experiment 2 ($d = .63$) and in other studies ($d = .30$, Voyer et al., 2004). We therefore held omissions constant via analysis of covariance, as in Experiment 2, and the sex difference in accuracy remained significant, $F(1, 134) = 4.36, p < .05$. Thus, the sex difference in mental rotation was attributable to confidence rather than omissions.

General Discussion

Given that men are more confident than women on mental rotation tasks (Cooke-Simpson & Voyer, 2007) and that confidence mediates performance on other cognitive tasks (Casey et al., 1997; Estes, 2004), we hypothesized that confidence would mediate the sex difference in mental rotation performance. This hypothesis seemed plausible for two reasons. First, it is consistent with much prior research demonstrating that sociocognitive factors, such as gender role beliefs (e.g., Massa et al., 2005), sex-typedness (e.g., Saucier et al., 2002), sexual orientation (e.g., Peters et al., 2007), and salience of sex stereotypes (e.g., McGlone & Aronson, 2006), are related to mental rotation performance. Confidence may serve as a common mediator by which these distal factors affect mental rotation. Second, because one is allowed to omit responding on any given trial of the MRT, one’s choice to respond or abstain presumably derives from one’s confidence in knowing the correct response. Hence, confidence is efficacious in that it determines whether to respond or abstain. So confidence is relevant to mental rotation performance and because confidence can affect

performance on cognitive tasks (Petrucci & Baranski, 2003), it was a plausible mediator.

Experiment 1 replicated the sex difference in mental rotation performance, but also showed that confidence mediated this sex difference and taking confidence into account eliminated the sex difference in mental rotation scores. Confidence predicted performance between sexes, within each sex, and even within individuals. The commission condition of Experiment 2 removed the efficacy of confidence by requiring participants to respond on every trial. When confidence was thus rendered irrelevant, the sex difference in accuracy that was observed in the standard omission condition was eliminated in this commission condition. In Experiment 3, the commission condition once again eliminated the sex difference in accuracy. However, the confidence condition reinstated the relevance of confidence by additionally requiring participants to judge their confidence in each response. With confidence thus emphasized, the sex difference in accuracy re-emerged. Finally, in Experiment 4, we manipulated participants' confidence by randomly informing them that they were either above or below average on an extremely difficult line judgment task. Participants subsequently performed better and worse, respectively, on the MRT. Moreover, women in the high confidence condition performed as well on the MRT as men in the low confidence condition. Thus, in each of the four experiments, we replicated and eliminated the sex difference in mental rotation performance by controlling or manipulating participants' confidence.

Understanding the source(s) of the sex difference ultimately may facilitate the bridging of the gender gap in mental rotation skills. Most directly, boosting females' confidence in their mental rotation abilities appears to improve their actual performance (see also Moe & Pazzaglia, 2006; Wraga et al., 2006). Potentially effective methods for achieving this outcome include rejecting the negative stereotype that women have poor spatial skills, encouraging women to view spatial skills as learnable, encouraging females to engage in more spatial tasks, and providing positive feedback when they do so. Such methods have proven effective for combating the effects of negative stereotypes on spatial tests and other performance measures (e.g., Aronson, Fried, & Good, 2002; Feng, Spence, & Pratt, 2007; Johns, Schmader, & Martens, 2005; Martens et al., 2006; Moe, Meneghetti, & Cadinu, 2009). Another promising method is to encourage women to reappraise the arousal they experience when performing under stereotype threat (Jamieson, Mendes, Blackstock, & Schmader, 2010; Johns, Inzlicht, & Schmader, 2008; Schmader et al., 2009). The present research suggests that merely rendering confidence irrelevant to the task might also improve females' mental rotation performance.

A corollary implication of this research is that the MRT may accentuate the sex difference in performance. Because participants may omit responses, the MRT implicitly induces participants to evaluate their confidence on each trial, thereby

affecting performance. Thus, it is no coincidence that the MRT exhibits the largest and most robust cognitive sex difference, with effect sizes (Cohen's *d*) ranging from .45 to 1.16 in the present study (see Table 1) and an average effect size of .66 across studies (Voyer et al., 1995). The large magnitude of this sex difference may be a direct consequence of the fact that confidence is particularly efficacious for performance on the MRT. This is not to say that the sex difference is merely a methodological artifact; rather, it may explain why the magnitude of the sex difference varies across different tests of mental rotation (see Voyer et al., 1995). This raises the question of whether confidence also mediates the sex difference observed on other tests of mental rotation (e.g., the Spatial Relations subtest of the Primary Mental Abilities test) (Thurstone & Thurstone, 1949) and spatial abilities more generally (e.g., the Rod-and-Frame test) (Witkin & Asch, 1948). Thus, the full implications of this research for the general class of sex differences in spatial ability are yet to be determined.

The large magnitude of the sex difference in mental rotation suggests that it may well have multiple causes. Moreover, those causes may be described at multiple levels of analysis, from distal factors (i.e., mediators) such as gender role beliefs to proximal factors (i.e., mechanisms) such as working memory capacity. Unfortunately, with a few notable exceptions (e.g., Beilock et al., 2007; Johns et al., 2008; Rydell, McConnell, & Beilock, 2009; Schmader & Johns, 2003; Schmader et al., 2008), investigations of this sex difference rarely extend across multiple levels of analysis. The present study was no exception. Because our primary purpose was to establish as simply as possible whether confidence mediates mental rotation performance, the present experiments did not attempt to relate this factor to other potential mediators or mechanisms.

Other Potential Mediators

In the introduction, we reviewed several studies demonstrating that sex stereotypes can affect mental rotation performance. Because we focused instead on confidence as a potential mediator of mental rotation, the present experiments did not manipulate or assess participants' beliefs in or awareness of sex stereotypes. However, many prior studies on gender role beliefs, stereotype threat, and stereotype lift explicitly appeal to confidence as a likely mediator between stereotypes and behavior (see e.g., Schmader et al., 2008; Steele, 1997; Walton & Cohen, 2003). In the domain of mathematics, for instance, stereotype threat increases self-doubt (Steele & Aronson, 1995) and negative performance-related thoughts (Beilock et al., 2007; Cadinu et al., 2005) among women, whereas boosting females' self-evaluation eliminates the stereotype threat effect (Martens et al., 2006; Rydell et al., 2009). In spatial tasks, women perform significantly better when they self-identify with scholastic

achievement (McGlone & Aronson, 2006), when told that the test measures empathy (Massa et al., 2005), when told that women are better than men at the task (Moe & Pazzaglia, 2006), and when confident in their ability on masculine tasks (Moe et al., 2009). All of these findings implicate confidence as a likely mediator between sex stereotypes and performance on spatial and mathematical tests.

Experience with spatial tasks is also related to mental rotation performance (Baenninger & Newcombe, 1995; Casey, 1996; Levine et al., 2005; Stericker & LeVesconte, 1982). Males tend to engage in more spatial activities than females, such as sports and action video games, and this factor predicts performance on spatial tasks (Newcombe, Bandura, & Taylor, 1983; Quaiser-Pohl, Geiser, & Lehmann, 2006; Terlecki & Newcombe, 2005; Voyer, Nolan, & Voyer, 2000). However, the correlation between spatial experience and performance is small (Baenninger & Newcombe, 1989; see also Scali, Brownlow, & Hicks, 2000). If spatial experience mediates spatial ability, then training on spatial tasks should improve performance. Indeed, extensive training improves mental rotation performance and attenuates the sex difference (Feng et al., 2007; Lizarraga & Ganuza, 2003), but gains in performance on trained stimuli typically do not generalize to untrained stimuli (Kail & Park, 1990; Lohman & Nichols, 1990; Sims & Mayer, 2002). Thus, the improvement in performance may be attributable to stimulus familiarity rather than task experience (Bethel-Fox & Shepard, 1988; Sims & Mayer, 2002; see also Kail et al., 1979). These effects of spatial experience and stimulus familiarity on mental rotation performance could, like the sex difference itself, be mediated by confidence. That is, practice and familiarity could increase confidence and hence improve performance on spatial tasks.

Biological models of the sex difference in spatial ability do not account for the present results. Kimura (1999) argued that differential cortical lateralization and/or differential exposure to sex hormones may explain the sex difference in spatial ability. Casey (1996) also argued that the sexual dimorphism in spatial ability is partially attributable to differential lateralization, though she additionally posited that spatial experience contributes to spatial ability. The present results do not exclude these biological explanations. But if differential lateralization and/or differential hormone exposure do contribute to the sex difference in mental rotation, their influence is mediated by confidence.

Possible Mechanisms

These experiments clearly demonstrate that confidence mediates performance on the MRT, but they do not indicate how it does so. One plausible account is that confidence affects the selection or deployment of strategies for mental rotation. Researchers have noted differential strategy use among participants (Schultz, 1991; Smith & Dror, 2001; Tomasino & Rumiati, 2004), with a

holistic strategy (i.e., rotating the object as a single entity) producing better performance than an analytic strategy (i.e., rotating the object part-by-part; Bethel-Fox & Shepard, 1988; Geiser, Lehmann, & Eid, 2006; Kail et al., 1979; Moe et al., 2009). And indeed, females are more likely than males to rotate analytically (Cochran & Wheatley, 1989; Geiser et al. 2006; Heil & Jansen-Osmann, 2008). For instance, women rotate complex objects more slowly than simple objects, thus suggesting an analytic strategy. In contrast, men rotate complex and simple objects equally fast, suggesting a holistic strategy (Heil & Jansen-Osmann, 2008). It may be that high confidence promotes holistic rotation, whereas low confidence induces analytic rotation.

Other potential cognitive mechanisms are spatial attention and memory. In their integrative model of stereotype threat effects, Schmader et al. (2008) argued that physiological stress, performance monitoring, and suppression of negative stereotypic thoughts collectively deplete working memory, which subsequently hinders performance on cognitive tasks. For instance, when women are the target of a negative stereotype, their working memory capacity is significantly reduced, and this decrement in working memory capacity mediates the decline in mathematical performance (Beilock et al., 2007; Johns et al., 2008; Rydell et al., 2009; Schmader et al., 2009; Schmader & Johns, 2003) and logical reasoning (Regner et al., 2010). In the present case, confidence might reduce stress or alleviate the need to suppress negative thoughts, thereby liberating working memory capacity for use on mental rotations. More directly relevant is a recent study by Kaufman (2007), who showed that mental rotation performance was predicted by spatial working memory in particular. Relatedly, Feng et al. (2007) found that males were better able to distribute attention across space, and they concluded that the sex difference in mental rotation performance was partially mediated by spatial attention. Thus, some evidence indicates that sex differences in spatial attention (Feng et al., 2007) and spatial working memory (Kaufman, 2007) contribute to the sex difference in mental rotation. It is currently unclear whether participants' confidence is related to either or both of these mechanisms.

Conclusions

In sum, we found that confidence predicted performance both between and within sexes (Experiment 1), that rendering confidence irrelevant to the task reliably eliminated the sex difference in performance (Experiments 2 and 3), and that manipulating confidence significantly affected performance (Experiment 4). Given that mental rotation exhibits the largest sex difference of any cognitive task (Halpern, 2000; Hines, 2004; Linn & Petersen, 1985; Maccoby & Jacklin, 1974), it is striking that this effect was reliably evoked and eliminated in each of the four experiments by such simple controls and manipulations. Confidence at least partially explains the variability in mental

rotation performance within each sex, as well as the difference between sexes. Thus, the sex difference in mental rotation appears to be a difference of performance rather than ability. An important endeavor for future research is to investigate how confidence relates to other potential mediators and mechanisms of this sex difference, such as sex stereotypes, spatial experience, rotation strategies, working memory, and spatial attention.

References

- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*, 113–125.
- Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*, 327–343.
- Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences*, *7*, 363–379.
- Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 929–945.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256–276.
- Bethel-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.
- Beyer, S., & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, *23*, 157–172.
- Blanton, H., Christie, C., & Dye, M. (2002). Social identity versus reference frame comparisons: The moderating role of stereotype endorsement. *Journal of Experimental Social Psychology*, *38*, 253–267.
- Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, *16*, 572–578.
- Casey, M. B. (1996). Understanding individual differences in spatial ability within females: A nature/nurture interactionist framework. *Developmental Review*, *16*, 241–260.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, *33*, 669–680.
- Cochran, K. F., & Wheatley, G. H. (1989). Ability and sex-related differences in cognitive strategies on spatial tasks. *Journal of General Psychology*, *116*, 43–55.
- Collaer, M. L., & Hines, M. (1995). Human behavioral sex differences: A role for gonadal hormones during early development? *Psychological Bulletin*, *118*, 55–107.
- Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the Mental Rotations Test. *Learning and Individual Differences*, *17*, 181–186.
- Estes, Z. (2004). Confidence and gradedness in semantic categorization: Definitely somewhat artificial, maybe absolutely natural. *Psychonomic Bulletin & Review*, *11*, 1041–1047.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.
- Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*, 850–855.
- 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.
- Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28*, 659–670.
- Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., et al. (2005). Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins. *Behavioral Neuroscience*, *119*, 104–117.
- Halpern, D. (2000). *Sex differences in cognitive abilities* (3rd ed.). Hillsdale, NJ: Erlbaum.
- 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.
- Hines, M. (2004). *Brain gender*. New York: Oxford University Press.
- Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: Targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology*, *28*, 1010–1026.
- Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology*, *46*, 208–212.
- Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*, 691–705.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle—Teaching stereotype threat as a means of improving women’s math performance. *Psychological Science*, *16*, 175–179.
- Kail, R., Carter, P., & Pellegrino, J. W. (1979). The locus of sex differences in spatial ability. *Perception & Psychophysics*, *26*, 182–186.
- Kail, R., & Park, Y. (1990). Impact of practice on speed of mental rotation. *Journal of Experimental Child Psychology*, *49*, 227–244.
- 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.
- Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.
- Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, *16*, 841–845.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of gender differences in spatial abilities: A meta-analysis. *Child Development*, *56*, 1479–1498.
- Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior*, *39*, 990–997.
- Lizarraga, M. L. S., & Ganuza, J. M. G. (2003). Improvement of mental rotation in girls and boys. *Sex Roles*, *49*, 277–286.
- Lohman, D. F. (1986). The effect of speed-accuracy tradeoff on sex differences in mental rotation. *Perception & Psychophysics*, *39*, 427–436.

- Lohman, D. F., & Nichols, P. D. (1990). Training spatial abilities: Effects of practice on rotation and synthesis tasks. *Learning and Individual Differences*, 2, 67–93.
- Maass, A., & Cadinu, M. (2003). Stereotype threat: When minority members underperform. *European Review of Social Psychology*, 14, 243–275.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, 42, 236–243.
- Massa, L. J., Mayer, R. E., & Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences*, 15, 99–111.
- Masters, M. S. (1998). The gender difference on the Mental Rotations Test is not due to performance factors. *Memory & Cognition*, 26, 444–448.
- Maylor, E. A., Reimers, S., Choi, J., Collaer, M. L., Peters, M., & Silverman, I. (2007). Gender and sexual orientation differences in cognition across adulthood: Age is kinder to women than to men regardless of sexual orientation. *Archives of Sexual Behavior*, 36, 235–249.
- McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27, 486–493.
- Moe, A., Meneghetti, C., & Cadinu, M. (2009). Women and mental rotation: Incremental theory and spatial strategy use enhance performance. *Personality and Individual Differences*, 46, 187–191.
- Moe, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, 16, 369–377.
- Newcombe, N., Bandura, M. M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. *Sex Roles*, 9, 377–386.
- Ortner, T. M., & Sieverding, M. (2008). Where are the gender differences? Male priming boosts spatial skills in women. *Sex Roles*, 59, 274–281.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles*, 48, 265–276.
- Pasterski, V., Zwierzyńska, K., & Estes, Z. (2011). Sex differences in semantic categorization. *Archives of Sexual Behavior*, 40, 1183–1187.
- Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, 57, 176–184.
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition*, 28, 39–58.
- Peters, M., Manning, J. T., & Reimers, S. (2007). The effects of sex, sexual orientation, and digit ratio (2D:4D) on mental rotation performance. *Archives of Sexual Behavior*, 36, 251–260.
- Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review*, 10, 177–183.
- Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. *Personality and Individual Differences*, 40, 609–619.
- Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Biosocial factors, sexual orientation and neurocognitive functioning. *Psychoneuroendocrinology*, 29, 867–881.
- Regner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J.-M., & Huguët, P. (2010). Individual differences in working memory moderate stereotype threat effects. *Psychological Science*, 21, 1646–1648.
- Resnick, S. M. (1993). Sex differences in mental rotations: An effect of time limits? *Brain and Cognition*, 21, 71–79.
- Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology*, 96, 949–996.
- Saucier, D. M., McCreary, D. R., & Saxberg, J. K. J. (2002). Does gender role socialization mediate sex difference in mental rotations? *Personality and Individual Differences*, 32, 1101–1111.
- Scali, R. M., Brownlow, S., & Hicks, J. L. (2000). Gender differences in spatial task performance as a function of speed or accuracy orientation. *Sex Roles*, 43, 359–376.
- Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin*, 35, 584–596.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356.
- Schultz, K. (1991). The contribution of solution strategy to spatial performance. *Canadian Journal of Experimental Psychology*, 45, 474–491.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science*, 171, 701–703.
- Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology*, 83, 638–647.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10, 81–84.
- Signorella, M. L., Jamison, W., & Krupa, M. H. (1989). Predicting spatial performance from gender stereotyping in activity preferences and in self-concept. *Developmental Psychology*, 25, 89–95.
- Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology*, 16, 97–115.
- Smith, W., & Dror, I. E. (2001). The role of meaning and familiarity in mental transformations. *Psychonomic Bulletin & Review*, 8, 732–741.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Stericker, A., & LeVesconte, S. (1982). Effect of brief training on sex-related differences in visual-spatial skill. *Journal of Personality and Social Psychology*, 43, 1018–1029.
- Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, 53, 433–441.
- Thurstone, L. L., & Thurstone, T. G. (1949). *Manual for the SRA primary mental abilities*. Chicago: Science Research Associates.
- Tomasino, B., & Rumiati, R. I. (2004). Effects of strategies on mental rotation and hemispheric lateralization: Neuropsychological evidence. *Journal of Cognitive Neuroscience*, 16, 878–888.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599–604.

- 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., Rodgers, M. A., & McCormick, P. A. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. *Memory & Cognition, 32*, 72–82.
- Voyer, D., & Saunders, K. A. (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.
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456–467.
- Witkin, H. A., & Asch, S. E. (1948). Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields. *Journal of Experimental Psychology, 38*, 762–782.
- Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13*, 813–819.