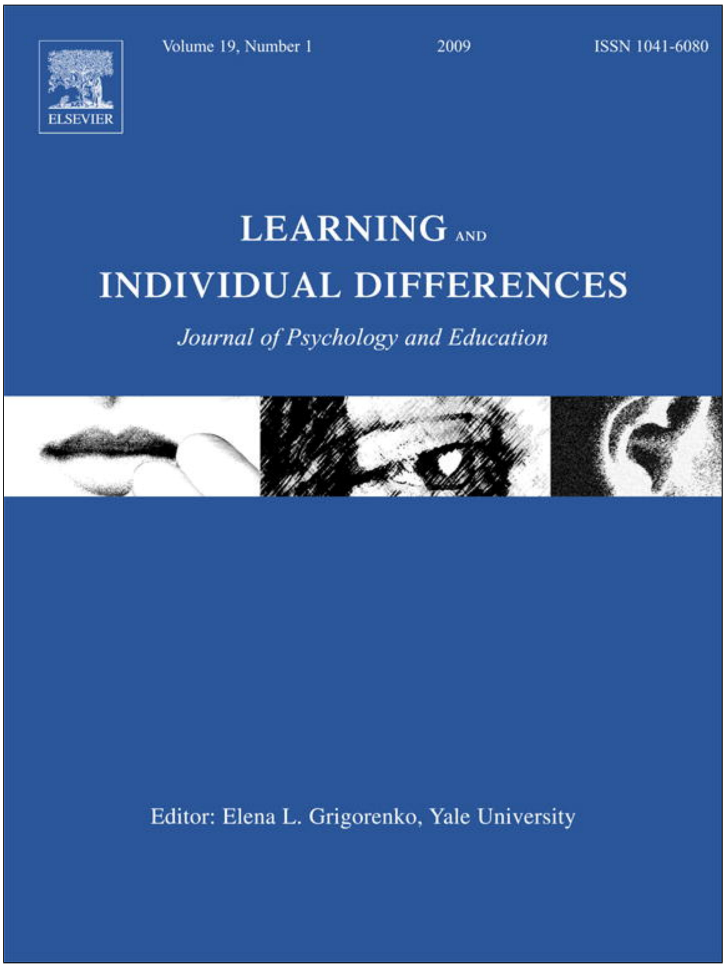
Volume 19, Number 1
2009
ISSN 1041-6080
ELSEVIER
LEARNING AND
INDIVIDUAL DIFFERENCES
Journal of Psychology and Education
Editor: Elena L. Grigorenko, Yale University

Available online at www.sciencedirect.com
ScienceDirect

Learning and Individual Differences 19 (2009) 21–27

Learning and Individual Differences

www.elsevier.com/locate/lindif

Are males always better than females in mental rotation? Exploring a gender belief explanation

Angelica Moè *

Department of General Psychology, University of Padova, Via Venezia, 8, 35131 Padova, Italy

Received 1 April 2007; received in revised form 26 January 2008; accepted 3 February 2008

Abstract

Males outperform females in the Mental Rotation Test (MRT) for biological, strategic and cultural reasons. The present research tested a motivational explanation with the hypothesis that females could do better when induced to have positive beliefs and expectations. All-female and all-male samples were divided into six groups, each having listened to different instructions: 1. men are better than women at this task; 2. women are better than men; 3. control instructions with no gender reference. Each group was further allocated to either the easy or the difficult task expectations condition. Experimental manipulation affected performance differently in relation to gender. Women's performance was affected by positive instructions about gender. Men were affected by instructions about the task difficulty. Women improved performance and reached men's scores in the MRT when they were led to believe they were better than men.

Keywords: Mental rotation; Stereotyped attitudes; Gender beliefs

1. Introduction

Mental rotation is a spatial task that involves the ability to mentally retain an object and rotate it in space. This ability is important for academic achievement given its supposed ability to predict success in topics such as geometry, mathematics, chemistry and physics; and for everyday spatial activities, such as orientation in unfamiliar places or finding a route on a map (Casey, 1996; Linn & Petersen, 1986).

The most frequently used assessment questionnaire is the mental rotation test (MRT) developed by Vandenberg and Kuse (1978). This consists of twenty graphic representations of a target three-dimensional object on the left (a 10-block figure with three angles); two rotated versions; and two distractors on the right (e.g. Fig. 1). The participants have to select in a limited time period the two correct answers, i.e. the rotations corresponding to the target object.

A male superiority has been extensively demonstrated in this task, greater than that observed in other spatial tasks (for a review see Voyer, Voyer, & Bryden, 1995). It emerges very early and is stable across life (Linn & Petersen, 1985).

Many explanations have been proposed.

The first explanation is biological. Gender differences in mental rotation could depend on hormonal factors or on hemispheric specialization and brain organization. Research showed that finger-length ratios, a measure of prenatal androgen levels exposure, correlate with MRT scores (Burton, Henninger, & Hafetz, 2005), whereas the effect appears to be significant in men, but not in women (Sanders, Bereczkei, Csatho, & Manning, 2005). The levels of circulating testosterone affect the performance in spatial tasks (Driscoll, Hamilton, Yeo, Brooks, & Sutherland, 2005) following an inverted-U shape function, where a high performance correlates with high testosterone levels in women and low testosterone levels in men (e.g. Geschwind & Galaburda, 1987; McKeever & Deyo, 1990; Nyborg, 1983). Moreover, using an fMRI technique differences in activation have been found between males (Alivisatos & Petrides, 1997) and females (Richter, Ugurbil,

* Tel.: +39 049 8276689; fax: +39 049 8276600.
E-mail address: angelica.moe@unipd.it.

1041-6080/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.lindif.2008.02.002

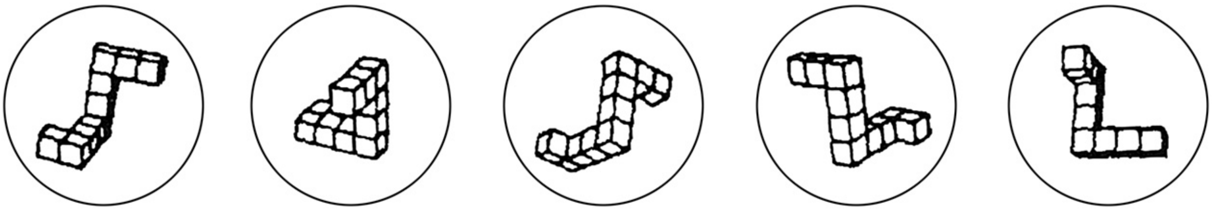

Fig. 1. A sample Mental Rotation item. Correct answers: the first and third.

Georgopoulus, & Kim, 1997). In performing mental rotation tasks there is evidence of an activation in the motor area, but women show bilateral processing both in verbal and in spatial tasks such as mental rotation (Howard, Fenwick, Brown, & Norton, 1992).

The second explanation is strategic. MRT can be performed using either holistic–spatial (e.g. rotate the target until it overlaps with the alternative stimulus) or analytic–verbal (e.g. counting blocks) strategies (Shepard & Metzler, 1971). Holistic strategies are more effective and preferred by males as demonstrated by using a selective interference paradigm (Pezaris & Casey, 1991) and more recently with an fMRI technique (Jordan, Wuestenberg, Heinze, Peters, & Jaencke, 2002). Given that females use less effective strategies than males, they attempt fewer items (Peters, 2005). When time pressure is stressed females attempt to resolve a higher number of items, but their accuracy does not improve, suggesting they are guessing (Cherney & Neff, 2004). In fact, guessing is defined by a high number of items containing a wrong response, while reluctance to guess is defined by a high proportion of blank responses (Voyer, Rodgers, & McCormick, 2004; Voyer & Saunders, 2004). Consequently, when more items are attempted but there is no increase in accuracy it is possible to argue that participants are guessing. When given unlimited time to finish the task, females perform as well as males (Scali, Brownlow, & Hicks, 2000), but this result is controversial. For instance, recently Peters (2005) found that, when the standard time allowed is doubled, females solve more problems, but the same is true for males, so the magnitude of the gender difference, measured through Cohen d, is not reduced.

The third explanation lies in the spatial experience. The role of prior exposure to spatial tasks (computer, video-games, and some sports) is important (Cherney & Neff, 2004). Prior performance of spatial tasks may have increased women's self-confidence and the knowledge of effective strategies for mental rotation (Casey, Nuttall, & Pezaris, 1997). Ginn and Pickens (2005) found that experience with spatial activities (e.g. participating in basket-ball, volleyball or being music performance majors or engaged in artistic activities) increases the mental rotation performance. Richardson (1994) found that gender differences in mental rotation performance can be reduced by educational experience. Casey, Nuttall, and Pezaris (1999) proposed a biological environmental interaction model following which only girls with a biological aptitude for spatial thinking, given by an inherited right-shift factor rs + −, i.e. right-handed with at least one first-degree relative left-handed or ambidextrous (Annett, 1995), improve their spatial abilities with experience.

Recently, a motivational explanation based on a stereotype threat effect has been put forward (Moè & Pazzaglia, 2006). The stereotype threat is the fear of confirming a stereotype about the group to which one belongs (Steele, 1997; Steele & Aronson, 1995). When a negative gender belief is aroused by the test instructions or by presenting the test as diagnostic of specific abilities, participants tend to under perform. This can be due to fear of failure (Steele, 1997), disengagement, intrusive thoughts (Cadinu, Maass, Rosabianca, & Kiesner, 2005), anxiety (Osborne, 2001), negative expectations (Cadinu, Maass, Frigerio, Impigliazzo, & Latinozzi, 2002), reduced working memory capacity (Schmader & Johns, 2003), increased mental load (Croizet et al., 2004) or heightened arousal (Ben-Zeev, Fein, & Inzlicht, 2005).

Research has found a number of mediators of the stereotype threat effect (for a review see Maass & Cadinu, 2003). Among these, in the present research, expectations about the difficulty of the task will be considered. A task presented as difficult can create a challenging situation that can motivate, or be a threat to one's abilities, thereby decreasing motivation and performance depending on perceived abilities, goals, and achievement motivation (e.g. Atkinson, 1964; Dweck, 1999). In the stereotype context it is possible that a task presented as difficult can create an additional pressure and hence produce a decrement in performance or, alternatively, following an attributional perspective, encourage the subject ('It isn't my fault: the task is difficult') (Weiner, 1985).

Stereotype threat effects on performance can be reduced by shaping an incremental theory of intelligence (Aronson, Fried, & Good, 2002) or through self-affirmation in an unrelated domain (Martens, Johns, Greenberg, & Schimel, 2006). Self-affirmation consists in affirming a valued characteristic that is not under threat before taking the test (Steele & Liu, 1983). Providing that the stereotype threat comes from a threat to self-integrity, self-affirmation can reduce the stereotype threat through an increase in self-esteem, boosting the sense of competence, integrity and self-worth and by reducing the fear of failure (Koole, Smeets, van Knippenberg, & Dijkesterhuis, 1999).

Positive effects due to stereotyping, known as stereotype lift (Walton & Cohen, 2003) or stereotype susceptibility (Shih, Ambady, Richeson, Fujita, & Gray, 2002; Shih, Pittinsky, & Ambady, 1999) has also been found. Stereotype lift occurs when an out-group is explicitly negatively stereotyped (e.g. men are told than women do worse) causing an enhancement in performance, probably due to an increase in self-efficacy (Bandura, 1997). It focuses on non-stereotyped groups (e.g. men for spatial abilities). Stereotype susceptibility is the performance boost caused by activation of a positive in-group stereotype (e.g. men are instructed men do better). The magnitude of the stereotype lift, measured through Cohen d, is half that of the stereotype susceptibility, probably because stereotype lift works by encouraging social comparison, while stereotype susceptibility works through self-oriented mechanisms, such as boosting self-esteem and a sense of self-integrity, which are more pervasive. Mainly identified and prejudiced people make use of social comparison, while a positive self-esteem is sought after by a large number of people (Steele, 1997). The magnitude of the increase due to an activation of a positive stereotype (i.e. caused by stereotype susceptibility) is equivalent to the decrease due to stereotype threat.

Sometimes, a positive stereotype can create apprehension and consequently create a decrease in performance, an effect called 'choking under pressure' (Cheryan & Bodenhausen, 2000). This occurs when a subject is expected to have a high performance and is worrying about confirming the high expectations. This does not generally occur when a subject is confident of performing well.

All these effects have been widely studied with mathematics, which is stereotypically considered as a male aptitude (Spencer, Steele, & Quinn, 1999), however spatial ability may also be relevant, as, in this case, gender differences partially stem from biological explanations. Following the stereotype threat theory, women can do worse than men in certain spatial tasks such as the MRT because they consider themselves less capable of performing this type of task than their male counterparts. However, if they are lead to believe they are capable of successfully solving the task, their performance can improve.

The present research explored two kinds of beliefs which could lead females to perform better in the MRT. The first concerned the 'able-in-the-task stereotype'. The second concerned the expectations aroused by presenting the task as easy or difficult. The belief that own gender is better or worse than the opposite in performing the task, and the perceived task difficulty (easier vs. more difficult than the previous one), were experimentally manipulated by feedback administration between presentation of the first and second part of the MRT. Participants were divided into six groups; each group was told either that men are more able than women in the performance of the MRT, or that women are more able than men, or that there is no gender difference in performing the task (control). Furthermore each group received instructions on the level of difficulty of the task (easy vs. difficult).

Past research indicates that females are more susceptible to gender stereotypes in the presence of males (Inzlicht & Ben-Zeev, 2000) in that to a greater number of males corresponds a greater decrease in the females' performance, whereas males are unaffected by the gender stereotype, even when they are in the minority. Hence, to avoid effects due to context, participants were tested in same-gender groups.

The hypotheses were that:

1) females' performance in the MRT improves if they are induced to think they are better than males (stereotype susceptibility and stereotype lift) given that instructions stress both in-group superiority and out-group inferiority; whereas, males' performance would improve when told they are better than their female counterpart;
2) in the threatening situation, that is when a gender stereotype is aroused, prior knowledge of task difficulty may add pressure to that already felt by a need to confirm the stereotype. This can affect performance negatively. Alternatively, it may decrease the effect of stereotyped instructions because high task difficulty may provide an alternative to which to attribute ones' failure;
3) for the control group, the expectation of a difficult task should facilitate performance for two reasons: it creates a challenging situation and it allows failure to be attributed to task characteristics.

2. Method

2.1. Participants

Participants were 71 women and 81 men, age range 15 to 22 years (mean 17.97 yrs, SD = 1.45), attending the third or fourth year of an Italian high school. They were divided into six groups according to the instructions given (see Procedure).

2.2. Materials

The original version of the MRT (Vandenberg & Kuse, 1978) is composed of 20 items of increasing difficulty. In the present research a version composed of two equal difficulty halves of the MRT was used to administer one before and the other after the experimental manipulation. The two equally complex parts were obtained by submitting the original version of the MRT to a group of 20 high school pupils and then matching the items with an equal percentage of correct answers (Moè & Pazzaglia, 2006).

2.3. Design

A mixed design was applied: 3 gender instructions (male superiority, female superiority, control) × 2 task instructions (easier vs. more difficult) × 2 (gender: males vs. females) × 2 time (before vs. after experimental manipulation), with the first three factors between-subjects and the last within-subjects.

2.4. Procedure

Participants were randomly assigned to one of the six groups.

In collective sessions, they were presented with the three practice items of the MRT. They were then told to complete the first half of the test in 4 min and that no additional time would be allowed. A 'start' signal by the experimenter followed. After a 'stop' at the end of the assigned time, the experimenter read aloud the instructions written on a sheet that the participants could also read.

The first group was told: 'This test measures spatial abilities. Research has shown that men perform better than women in this test, probably for genetic reasons. This means that women score lower than men. Now you will be presented with the second part of the test. Note that it is easier than the first' (23 participants). The second group was told: 'This test measures spatial abilities. Research has shown that men perform better than women in this test, probably for genetic reasons. This means that women score lower than men. Now you will be presented the second part of the test. Note that it is more difficult than the first' (21 participants). The third group was told: 'This test measures spatial abilities. Research has shown that women perform better than men in this test, probably for genetic reasons. This means that men score lower than women. Now you will be presented the second part of the test. Note that it is easier than the first' (26 participants)'. The fourth group was instructed as the third except for the final words which read '..it is more difficult than the first' (27 participants). The fifth and the sixth groups received a general instruction: 'This test measures spatial abilities. Research has shown that spatial ability is very important in everyday life, e.g. to find a route or describe a pathway to someone. This test has been used in the USA and over the last few years also in Europe, in particular Italy, confirming the results obtained with the American samples'. The fifth was then told 'Now you will be presented the second part of the test. Note that it is easier than the first' (25 participants), the sixth '..it is more difficult than the first' (30 participants).

The participants were then asked to complete the second half of the test in 4 min, beginning after the 'start' signal and ending at the 'stop' announced by the experimenter. All the sheets were then collected and the participants were debriefed about the actual aims of the research.

2.5. Data scoring

One point was assigned for each correct response. Since there were two correct responses for each item, the maximum theoretical score for each half of the MRT was 20. The maximum score obtained was 18 for the first half and 19 for the second half. The minimum was 5 for the first half and 4 for the second half. The correlation between the first and the second part of the MRT was $r=.529$, $p<.001$.

3. Results

A 3 gender instructions (men superiority, women superiority, control) × 2 difficulty instructions (easier vs. more difficult) × 2 (gender: males vs. females) × 2 time (before vs. after experimental manipulation) MANOVA was run on the mean accuracy score (total correct answers). The first three factors were between-subject, the forth within-subject. Moreover, in order to evaluate the effect size and the strength of the relationships (Kirk, 1996; Thompson, 2002), both the partial eta squared and the Cohen (1988) d indexes were calculated. The partial eta squared is a measure of explained variance, while the Cohen d is the difference between means in standard deviations, i.e. a Cohen d of .50 means that two groups differ by half standard deviation. There is some debate in the literature about the appropriate levels of both partial eta squared and the Cohen d (Fern & Monroe, 1996; Olejnik & Algina, 2000). However, a Cohen d less than .20 can be considered as a small effect, .20 < Cohen d < .50 as a medium effect and > .50 as a large effect (Cohen, 1988). As to eta squared, it depends on the design and on the number and kind of factors (Olejnik & Algina, 2003).

Results showed a significant main effect due to gender, $F(1, 140)=18.66$, $p<0.001$, $\eta^2=0.12$ (females $M=10.13$, SD=2.64, males $M=12.06$, SD=2.77, Cohen $d=0.67$), due to time, $F(1, 140)=17.42$, $p<0.001$, $\eta^2=0.11$ (before $M=10.59$, SD=3.15, after $M=11.73$, SD=3.42, Cohen $d=0.40$) and an interaction gender × gender instructions × difficulty instructions × time, $F(2, 140)=4.50$, $p=0.013$, $\eta^2=0.06$, Tukey $a=1.74$), see Table 1.

The difference between before and after instructions was significant for males being told that males are better and the task is easy (before $M=11.08$, SD=3.96, after $M=13.00$, SD=3.16, Cohen $d=0.61$), for females told that females are better and the task is easy (before $M=9.67$, SD=1.92, after $M=11.08$, SD=3.06, Cohen $d=0.72$) or is difficult (before $M=9.46$, SD=3.33, after $M=12.85$, SD=3.31, Cohen $d=1.09$) and for males in the control group told that the task is difficult (before $M=11.16$, SD=3.73, after $M=13.68$, SD=3.75, Cohen $d=0.75$). All Cohen d were greater than .50. This indicates a large effect. As to the partial eta squared, in a design with one or two factors a medium effect is given by partial eta squared > .15 and a large effect by eta squared > .30. In the present research there are four factors. As a result, a smaller than .15 effect may also have a practical significance, because it explains a portion of variance.

As the interaction gender × time was non-significant, $F(1, 140)=0.07$, $p>.05$, and there was no other interaction involving time, data were analyzed after instructions and along a males vs. females comparison.

The overall MANOVA 3 gender instructions (men superiority, women superiority, control) × 2 difficulty instructions (easier vs. more difficult) × 2 (gender: males vs. females) showed a significant main effect due to gender, $F(1, 140) = 13.83$, $p<0.001$, $\eta^2=0.09$ (females $M=10.13$, SD=2.64, males $M=12.06$, SD=2.77, Cohen $d=0.67$), and an interaction gender × gender instructions, $F(2, 140)=3.23$, $p=0.04$, $\eta^2=0.04$. The Tukey critical value was 1.81. Comparisons showed that males outperformed females when told that males are

Table 1
Mean scores in accuracy (total number of correct responses) in the MRT before and after instructions

Instructions		n	Before		After	
Gender	Difficulty level		M	SD	M	SD
Men						
Men are better	Easy [a]	12	11.08	3.96	13.00	3.16
	Difficult	10	11.50	2.63	12.80	2.66
Women are better	Easy	14	11.43	2.21	12.57	3.11
	Difficult	14	11.29	3.45	11.57	3.84
No gender instruction control	Easy	12	12.58	2.64	11.92	2.23
	Difficult [a]	19	11.16	3.73	13.68	3.76
Women						
Men are better	Easy	11	8.73	3.32	9.00	3.41
	Difficult	11	10.18	2.23	11.18	3.16
Women are better	Easy [a]	12	9.67	1.92	11.08	3.06
	Difficult [a]	13	9.46	3.33	12.85	3.31
No gender instruction control	Easy	13	9.61	3.71	10.15	2.41
	Difficult	11	9.82	2.18	9.45	3.70

Note. Maximum score = 20.
[a] Significant difference at $p<0.05$ level.

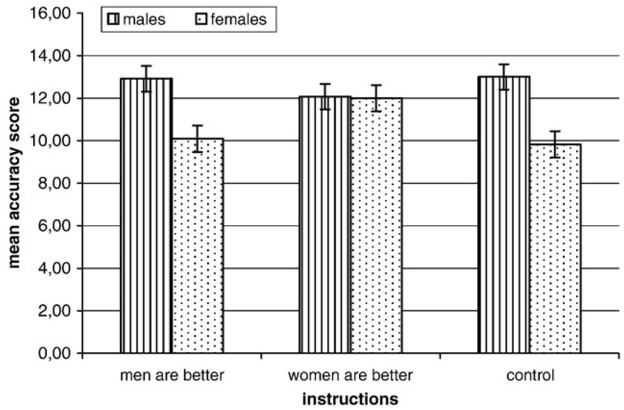

Fig. 2. Mean MRT accuracy score in males and females after instructions.

better (females M=10.09, SD=3.39, males M=12.91, SD=2.88, Cohen d=0.82), and in the control condition (females M=9.83, SD=3.02, males M=13.0, SD=3.33, Cohen d=0.89), but not when they were told that females are better than males (females M=12.00, SD=3.25, males M=12.07, SD=3.46, see Fig. 2). In these analyses also, Cohen d values showed a large effect.

As can be seen in Fig. 2, after instructions, females performed as well as males in MRT if induced to believe they do better than the opposite gender, but performed worse than males both when told that men are better and in the control condition.

4. Discussion

Instructions affected the females' performance when the gender of the participants was indicated as performing better than the opposite, no matter if the task was presented as easy or difficult. Females were unaffected by the expected difficulty, their level of engagement appearing to be tied more to expectations based on the self than on those concerning the task.

Differently from females, the males' performance was affected by both gender and difficulty instructions. Males improved their performance when induced to believe themselves to be more able than females and when expecting an easy task or, in the control condition, if expecting a difficult task. The expectation to face a difficult task added pressure to the apprehension to confirm the stereotype and removed the effect of improvement caused by the positive stereotype. When nothing was said about gender, the expectation of a difficult task caused an increase in performance.

No effect on performance was found for the less favourable conditions, i.e. when males were told that women do better than men or females were told they do worse than males. Two explanations could be put forward. The first is that, differently from mathematics — the typical task used in research on stereotype threat (Maass & Cadinu, 2003; Steele, 1997) — participants cannot hold the view that, in this task, males do better than females; because they may be unaware they are undertaking a spatial task, where research has consistently found a male superiority. In fact, a deficit can occur only in negatively stereotyped domains (Steele & Aronson, 1995) and when people endorse the stereotype and value the performance domain highly (Aronson et al., 1999). In this situation, the difference appears to be in the way the task is presented. If participants are led to know they are good at the task they believe in what is said, and performance is affected accordingly. The second explanation is that, providing a general increase in the performance from the first to the second part of the MRT, the lack of increase can be assumed as a small negative effect.

In comparing males and females it emerged that the gender difference disappeared when participants were told that females perform better than males.

5. Conclusions

Following a socio-cognitive perspective, beliefs about one's abilities or task characteristics are supposed to affect cognitive performance (Dweck, 1999).

Research conducted following the stereotype threat, the stereotype lift and the stereotype susceptibility paradigms show that the belief to be or not to be a stigmatised group for that specific ability can be of detriment or boost performance following the direction of the activated expectations. These effects were studied with the MRT, a test where robust data show male superiority, with the general hypothesis that females could do better if induced to hold positive beliefs about personal abilities and the task.

Data showed that females increased performance after having been instructed that they do better than males. Encouraging the expectation to succeed by stressing in-group skills in the specific task had a self-enhancing effect, even in a task such as mental rotation where there are biological issues that demonstrate out-group superiority.

Males engaged differently. They were more affected than females by instructions about the characteristics (easier vs. more difficult) of the task. They improved performance in the no-threatening situations, that is when nothing was said about the gender and failure could be explained by the difficulty of the task and when they believed themselves to be more able than females and the task was presented as easy. When the fear of failure due to the difficulty of the task was aroused the enhancing effect due to the gender superiority (stressed by the instructions and true in reality) disappeared.

Taken together results suggest that, regardless of gender, a subject increases performance when gender superiority is suggested by the given instructions. Difficulty or ease expectations affect the males' performance, but not the females'.

Probably gender instructions and instructions on task characteristics work through different mechanisms. Instructions about ones' own gender superiority can motivate to confirm held beliefs (i.e. to perform according to the stereotype) giving support through self-integrity and self-esteem, because focused on the self. Differently, expectations and instructions about difficulty levels are focused on the task and may provide an external excuse for failure, thus protecting or enhancing self-esteem.

To better understand these mechanisms future research could include a measure of self-reported attributions for success and failure. Moreover, the perceived difficulty of the task could be assessed. Perhaps the instructions about the ease or difficulty of the task differently affect subjects who have perceived the first part as easy or as difficult.

Finally, further research could better explore some mediators. One that appears interesting is the prevention vs. promotion-oriented regulatory focus, i.e. the tendency in face of uncertainty to answer anyway by guessing or by being more cautious (Higgins et al., 2001). A greater tendency to caution in the more threatening situations may be expected. Research with chess suggests that women tend to be less promotion-oriented than men and that the more promotion-oriented they are the better their performance (Maass, D'Ettole, & Cadinu, in press).

Results suggest that females' performance in MRT could increase and reach men's scores when positive beliefs about self are given. This in turn can have a positive impact on female achievement and career choices in topics such as geometry and mathematics, and for some sports and everyday orienteering.

Positive beliefs about personal abilities to perform the task and the task's difficulty can explain performance, predict success and direct choices particularly for subjects generally considered to be less capable in this domain.

Acknowledgements

Thank to Drs. Anna Pregnolato and Agnese Carpaci for their assistance in data collection.

I am particularly grateful to Francesca Pazzaglia for her valuable suggestions in earlier versions of this paper.

References

Alivisatos, B., & Petrides, M. (1997). Functional activation of the human brain during mental rotation. *Neuropsychologia*, *35*, 111−118.

Annett, M. (1995). The right shift theory of a genetic balanced polymorphism for cerebral dominance and cognitive processing. *Cahiers de Psychologie Cognitive*, *14*, 427−480.

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.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29−46.

Atkinson, J. W. (1964). *An introduction to motivation.* Princeton, N.J.: Van Nostrad.

Bandura, A. (1997). *Self-efficacy: the exercise of control.* New York: Freeman.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, *41*, 174−181.

Burton, L. A., Henninger, D., & Hafetz, J. (2005). Gender differences in relations of mental rotation, verbal fluency, and SAT scores to finger length ratios as hormonal indexes. *Developmental Neuropsychology*, *28*(1), 493−505.

Cadinu, M., Maass, A., Frigerio, S., Impigliazzo, L., & Latinozzi, S. (2002). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, *33*, 267−285.

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*(7), 572−578.

Casey, M. B. (1996). Understanding individual differences in spatial ability with 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 test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, *33*(4), 669−680.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1999). Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills. *Developmental Psychology*, *35*, 1237−1247.

Cherney, I. D., & Neff, N. L. (2004). Role of strategies and prior exposure in mental rotation. *Perceptual and Motor Skills*, *98*, 1269−1282.

Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of "model minority" status. *Psychological Science*, *11*(5), 399−402.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*, 2nd ed. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Croizet, J. C., Després, G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering disruptive mental load. *Personality and Social Psychology Bulletin*, *30*(6), 721−731.

Driscoll, I., Hamilton, D. E., Yeo, R. A., Brooks, W. M., & Sutherland, R. J. (2005). Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Hormones and Behavior*, *47*, 326−335.

Dweck, C. S. (1999). *Self-theories: their role in motivation, personality, and development.* Ann Arbor, M. I.: Psychology Press.

Fern, F. E., & Monroe, K. L. B. (1996). Effect-size estimates: Issues and problems in interpretation. *Journal of Consumer Research*, *23*, 89−105.

Geschwind, N., & Galaburda, A. (1987). *Cerebral lateralization: Biological mechanisms, associations, and pathology.* Cambridge, MA: MIT Press.

Ginn, S. R., & Pickens, S. J. (2005). Relationships between spatial activities and scores on the mental rotation test as a function of sex. *Perceptual and Motor Skills*, *100*, 877−881.

Higgins, T. E., Friedman, R. S., Harlow, R. E., Chen Idson, L., Ayduk, O. N., & Taylor, A. (2001). Achievement orientation from subjective histories of success: Promotion pride and prevention pride. *European Journal of Social Psychology*, *31*, 3−23.

Howard, R., Fenwick, P., Brown, D., & Norton, R. (1992). Relationship between CNV asymmetries and individual differences in cognitive performance, personality and gender. *International Journal of Psychophysiology*, *13*, 191−197.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365−371.

Jordan, K., Wuestenberg, T., Heinze, H. J., Peters, M., & Jaencke, L. (2002). Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia*, *40*, 2397−2408.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746−759.

Koole, S. L., Smeets, K., van Knippenberg, A., & Dijkesterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology*, *77*, 111−125.

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

Linn, M. C., & Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In S. Hyde & M.C. Linn (Eds.), *The psychology of gender* (pp. 67−101). Baltimore: Johns Hopkins Univ. Press.

Maass, A., & Cadinu, M. (2003). Stereotype threat: When minority members underperform. *European Review of Social Psychology*, *14*, 243−275.

Maass, A., D'Ettole, C., & Cadinu, M. (in press). Checkmate ? The role of gender stereotypes in the ultimate intellectual sport. European Journal of Social Psychology (www.interscience.wiley.com). doi:10.1002/ejsp.440

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.

McKeever, W. F., & Deyo, R. A. (1990). Testosterone, dihydrotestosterone, and spatial task performances of males. *Bulletin of the Psychonomic Society*, *28*, 305−308.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, *16*, 369−377.

Nyborg, H. (1983). Spatial ability in men and women: Review and theory. *Advances in Behavior Research and Therapy*, *5*, 89−140.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations and limitations. *Contemporary Educational Psychology*, *25*, 241−286.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*(4), 434−447.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, *26*, 291−310.

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, *57*, 176−184.

Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition*, *17*, 1−22.

Richardson, J. T. (1994). Gender differences in mental rotation. *Perceptual and Motor Skills*, *78*, 435−448.

Richter, W., Ugurbil, K., Georgopoulus, A., & Kim, S. (1997). Time-resolved fMRI of mental rotation. *NeuroReport*, *8*, 3697−3702.

Sanders, G., Bereczkei, T., Csatho, A., & Manning, J. (2005). The ratio of the 2nd to 4th finger length predicts spatial ability in men but not in women. *Cortex*, *41*, 789−795.

Scali, R., Brownlow, S., & Hicks, J. L. (2000). Gender differences in spatial task performance as a function of speed of accuracy orientation. *Sex Roles*, *43*, 359−376.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440−452.

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*(1), 80−83.

Spencer, S., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4−28.

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 performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797−811.

Steele, C. M., & Liu, T. J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology*, *45*, 5−19.

Thompson, B. (2002). "Statistical", "practical", and "clinical": How many kinds of significance do counsellors need to consider? *Journal of Counseling & Development*, *76*, 436−441.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations: A group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599−604.

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*(1), 72−82.

Voyer, D., & Saunders, K. A. (2004). Gender differences on the mental rotations test: A factor analysis. *Acta Psychologica*, *117*, 74−94.

Voyer, D., Voyer, S. D., & 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*(5), 456−467.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548−573.