Psychoneuroendocrinology (2009) **34**, 389—401

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/psyneuen

Interactive effects of sex hormones and gender stereotypes on cognitive sex differences—A psychobiosocial approach

Markus Hausmann [a,*], Daniela Schoofs [b], Harriet E.S. Rosenthal [a], Kirsten Jordan [c]

[a] *Department of Psychology, Durham University, United Kingdom*
[b] *Department of Cognitive Psychology, Ruhr-University Bochum, Germany*
[c] *Department of Medical Psychology and Medical Sociology, Georg-August University Göttingen, Germany*

Received 20 March 2008; received in revised form 25 September 2008; accepted 26 September 2008

KEYWORDS
Cognitive sex differences;
Gender stereotypes;
Mental rotation;
Perceptual speed;
Stereotype threat;
Testosterone;
Verbal fluency

Summary Biological and social factors have been shown to affect cognitive sex differences. For example, several studies have found that sex hormones have activating effects on sex-sensitive tasks. On the other hand, it has been shown that gender stereotypes can influence the cognitive performance of (gender-) stereotyped individuals. However, few studies have investigated the combined effects of both factors. The present study investigated the interaction between sex hormones and gender stereotypes within a psychobiosocial approach. One hundred and fourteen participants (59 women) performed a battery of sex-sensitive cognitive tasks, including mental rotation, verbal fluency, and perceptual speed. Saliva samples were taken immediately after cognitive testing. Levels of testosterone (T) were analysed using chemiluminescence immunoassay (LIA). To activate gender stereotypes, a questionnaire was applied to the experimental group that referred to the cognitive tasks used. The control group received an identical questionnaire but with a gender-neutral content. As expected, significant sex differences favouring males and females appeared for mental rotation and verbal fluency tasks, respectively. The results revealed no sex difference in perceptual speed. The male superiority in the Revised Vandenberg and Kuse Mental Rotations Tests (MRT-3D) was mainly driven by the stereotype-active group. No significant sex difference in MRT-3D appeared in the control group. The MRT-3D was also the task in which a strong gender-stereotype favouring males was present for both males and females. Interestingly, T levels of the stereotype-activated group were 60% higher than that of male controls. The results suggest that sex hormones mediate the effects of gender stereotypes on specific cognitive abilities.

* Corresponding author at: Department of Psychology, Durham University, South Road, Durham DH1 3LE, United Kingdom. Tel.: +44 191 3343286; fax: +44 191 3314377.
E-mail address: Markus.Hausmann@durham.ac.uk (M. Hausmann).

0306-4530/$ — see front matter.
doi:10.1016/j.psyneuen.2008.09.019

1. Introduction

Sex differences in specific cognitive abilities are well documented (Maccoby and Jacklin, 1974; Kimura, 1999; Halpern, 2000; Hines, 2004; Hausmann, 2007). Although performances of the sexes overlap to a large degree (McKeever, 1995), women tend to outperform men in specific aspects of verbal ability (McGlone, 1980; Halpern, 2000), whereas men achieve higher scores on some visuospatial tasks (Witkin et al., 1962; McGee, 1979; Hyde, 1981; Voyer et al., 1995; Halpern, 1996).

Several meta-analyses indicate that sex differences in spatial abilities exist and are robust (Tapley and Bryden, 1977; Linn and Petersen, 1985; Voyer et al., 1995; Silverman et al., 1996, see e.g. Halpern, 2000; Hines, 2004; Hausmann, 2007, for a review). Most particularly, the redrawn Vandenberg and Kuse mental rotation test (MRT-3D; Vandenberg and Kuse, 1978; Peters et al., 1995b), which uses 3D cube figures designed by Shepard and Metzler (1971), appears to produce the most reliable sex difference of all spatial paper-pencil tests (Voyer et al., 1995). To perform the MRT-3D, the individual must imagine a cube stimuli revolving in 3D space (Collins and Kimura, 1997). The male advantage in mental rotation decreases with picture plane rotations in 2D space. The meta-analysis of Voyer et al. (1995) showed that, on average, males outperform females in mental rotation by about 0.6 S.D. units but only by 0.2 S.D. units for the spatial visualization category including, for example, the Hidden Figures Test (Ekstrom et al., 1976), in which participants must find a simple figure embedded within a complex pattern (Voyer et al., 1995). Probably, however, 3D processes are not a prerequisite for large gender differences. Collins and Kimura (1997) found a similar male advantage when task difficulty was increased for a 2D mental rotation test.

Sex differences in spatial abilities, and some other cognitive tests, arise at least in part because of the influence of sex hormones. Several studies suggest that gonadal hormones affect spatial abilities during early phases of development (e.g. Williams and Meck, 1991; Christiansen, 1993). For example, it has been shown that participants with congenital adrenal hyperplasia (CAH), a genetic condition that causes an overproduction of adrenal androgens, revealed enhanced performance on some spatial tasks (Perlman, 1973; Resnick et al., 1986; Hampson et al., 1998). However, a well-designed CAH study (Hines et al., 2003) did not find that females with CAH performed better than unaffected females on mental rotation. Mental rotation performance in men with CAH was even impaired. The authors concluded that this outcome is not consistent with the hypothesis that prenatal androgen exposure enhances mental rotation performance.

Other studies suggest activating influences of sex hormones on cognitive abilities which persist throughout the whole lifetime (Gouchie and Kimura, 1991; Van Goozen et al., 1994; Kimura, 1996; Slabbekoorn et al., 1999). Both organizing and activating studies generally show a decrease in spatial abilities with increased estrogen levels, consistent with observed sex differences. Moreover, it has been found that levels of testosterone (T) within the normal adult-male range are accompanied by the male advantage on spatial tasks. However, not all studies found activating effects as a result of hormone administration (Alexander et al., 1998; Liben et al., 2002). For example, Liben et al. (2002) investigated adolescents receiving hormonal treatment for delayed puberty and found that that T and E replacements did not affect spatial performance, although spatial performance showed the typical sex difference. Moreover, inconsistent findings about the relationship between current T levels and performance on spatial tasks for males do exist. Yang et al. (2007) reviewed studies assessing the relationship between mental rotation performance and T in normal men and found evidence for positive (Gordon and Lee, 1986; Silverman et al., 1999; Hooven et al., 2004), negative (Moffat and Hampson, 1996), and no relationship (Gouchie and Kimura, 1991; Kampen and Sherwin, 1996; Alexander et al., 1998; Halari et al., 2005; Falter et al., 2006; Burkitt et al., 2007). Yang et al. (2007) point out that these inconsistencies might partly result from methodological limitations and differences between studies, such as limited accuracy in hormone assays, reliance on free saliva or plasma total T levels, different measures of spatial ability, small sample sizes, presence of female investigators, etc. Similar inconsistencies were also reported for studies including women. Aleman et al. (2004) found that a single administration of T led to an improved performance in the paper-pencil version of the MRT-3D, whereas Burkitt et al. (2007) failed to show any relationship between salivary T levels and MRT-3D performance in women (and men) but reported a positive relationship between T levels and performance on a virtual water maze in women (only).

Other studies investigated the relationship between sex hormones and cognitive abilities in normally cycling women during the menstrual cycle (Gordon et al., 1986; Hampson, 1990a,b; Gordon and Lee, 1993; Epting and Overman, 1998; Hausmann et al., 2000; Mumenthaler et al., 2001; Rosenberg and Park, 2002). However, the relationship between sex hormone levels and cognitive performance in normally cycling women also revealed conflicting results. For example, Hausmann et al. (2000) found cycle-related difference in MRT-3D, with higher scores during the menstrual phase than during the midluteal phase. In this study, MRT-3D performance was positively and negatively related to T and E levels, respectively. This finding has been supported by Maki et al. (2002) who also found higher MRT-3D scores during the menses than midluteal phase, and E levels to be negatively related to MRT-3D performance. Additionally, they found a positive relationship between E levels and verbal fluency (Maki et al., 2002). However, menstrual cycle-related fluctuations in cognitive abilities were not always found (e.g. Gordon and Lee, 1993; Epting and Overman, 1998; Rosenberg and Park, 2002), although typical cognitive sex differences have been confirmed (e.g. Epting and Overman, 1998).

Athough several inconsistencies exist, previous research suggests that levels of sex hormones, and T and E, in particular, are related to cognitive sex differences. Besides biological/hormonal explanations for cognitive sex differences, studies focusing on socio-cultural factors suggest that gender-stereotypes are also strongly related to cognitive sex differences. Of relevance here is research examining stereotype threat (Steele and Aronson, 1995). Stereotype threat can be defined as the fear of conforming to a negative stereotype associated with one's group membership, which paradoxically results in the individual behaving in line with the stereotype. While the original research examined this effect for African Americans and the stereotype of intellectual ability, the effect has been found in a number of situations, including those associated with gender stereotypes.

For example, women have been found to underperform on mathematics tests when informed that the performance of men and women would be compared (Brown and Josephs, 1999; Inzlicht and Ben-Zeev, 2000; Marx and Roman, 2002; Rosenthal et al., 2007). Spencer et al. (1999) found that when a mathematics test was described as having previously shown gender differences, women underperformed compared to men. However, when told the test did not produce gender differences women performed equally to men. Stereotype threat research suggests that knowledge of a stereotype (e.g., men are better at math than women) can affect the cognitive performance of an individual within that stereotyped domain, that being reminded of a negative stereotype (be it explicitly or simply being placed in the relevant situation) can affect performance.

Stereotype threat has also been found for men when they are compared to women on stereotypically female domains (e.g. Leyens et al., 2000; Koenig and Eagly, 2005). Of relevance here is literature examining the impact of positive and negative stereotypes on men and women's performance on a mental rotation task. Positive stereotype test instructions emphasizing a female advantage (women are better at perspective taking) led to a poorer performance for men (Wraga et al., 2006, 2007). Whereas the negative stereotype (men outperform women on spatial tasks) led to poorer performance for women (Wraga et al., 2007). In addition, positively stereotyped groups can actually *improve* their performance in such situations (*stereotype lift*, Walton and Cohen, 2003; *stereotype boost*, Shih et al., 2002), although it is usually a trend which does not reach statistical significance (Walton and Cohen, 2003).

Besides possible biological (hormonal) reasons for individual differences in cognitive performance, stereotype threat and stereotype lift can also affect cognitive performance when a particular task is perceived as stereotypically male or female. Given that both sex hormones and gender stereotypes affect sex-sensitive cognitive domains, it is reasonable to assume that sex hormones and gender stereotypes can interact. To ignore the potential influence of (implicitly activated) gender stereotypes when focusing on the effects of sex hormones on cognitive sex differences (and vice versa) might contribute to some of the inconsistent results mentioned above.

In fact, recent research suggests a link between gender-stereotypes and sex hormones. For example, Josephs et al. (2003) found that females with high and low testosterone levels respond differently to gender-stereotypes. Specifically, females with high T levels significantly underperformed in the stereotype threat condition compared to the control condition, while females with low T levels did not differ on performance between conditions. This suggests that stereotype threat effects are more pronounced for individuals who are naturally higher in T. In contrast, no significant difference in performance was found between males of high and low T.

In the equivalent of a stereotype lift scenario, male participants were either given instructions that the test could only identify exceptional math ability, or poor math ability. In the exceptional condition, high T males performed better than low T males and high T males performed better in the exceptional compared to the weak abilities condition. Josephs et al. (2003) conclude that T levels, of both men and women, may moderate the association between task performance and stereotype threat. Men with naturally high T levels should view math tests as a way to maintain their high status in math, while high T women should see math tests as a threat to their status. Presumably, so too would be the case in alternative stereotyped domains.

While Josephs et al. (2003) examined T levels as a moderator of stereotype threat, they do not examine the effect of stereotype threat on T, that is, whether a stereotype threat scenario would result in an *increase* in T, compared to a control condition. The experiment presented here examines the interaction between social and biological factors that may influence performance on cognitive tasks. As previously noted, men tend to outperform women on spatial tasks, while women achieve higher scores in specific aspects of verbal ability. In addition to examining these two areas, perceptual speed is also measured—a cognitive domain which is known to reveal sex differences favoring women of small to moderate effect size (Feingold, 1992; Hedges and Nowell, 1995; Burns and Nettelbeck, 2005). These cognitive tasks are performed following either a control or experimental manipulation designed to activate gender stereotypes. In addition to examining the relationship between gender stereotypes and cognitive performance, T levels were examined.

Only very few studies have investigated the effects of gender-stereotypes on cognitive tasks such as mental rotation and verbal fluency which are known to be sensitive to sex hormones. Using a psychobiosocial approach, the aim of the present study is not only to investigate the contribution of gender stereotypes and testosterone on cognitive abilities but also to explore how these crucial factors interact.

2. Methods

2.1. Participants

One hundred and fourteen subjects (55 men, 59 women) participated in the present study. The mean age for women was 23.44 years (S.D. = 4.74), and 25.83 years (S.D. = 7.21) for men. Participants were randomly assigned to either the experimental (gender-stereotype activated) or control group. The experimental group contained 58 participants (24 men, 34 women), and the control group contained 56 participants (31 men, 25 women). All participants were tested in mixed-gender groups of four to eight participants, with an equal distribution of males and females within these groups maintained, if possible, depending on availability. Participants who had used medication affecting the central nervous system during the last six months were excluded. All participants had normal or corrected-to-normal visual acuity and were naïve to the study's hypotheses. They were recruited by announcements and paid for their participation. The study was approved by the local Research Ethics Committee at the Ruhr-University Bochum; all participants gave written informed consent prior to participation.

2.2. Procedure and materials

2.2.1. Mood

To control for mood, which might influence cognitive performance, a German mood scale (Zerssen, 1976) was administered. Mood scores ranged between 0 (euphoric) and 56 (extremely depressive). Scores were grouped into five mood categories: 0—6 (euphoric), 7—16 (well-balanced), 17—26 (depressed), 27—41 (moderately depressive), and 42—56 (extremely depressive). None of the participants fell into the pathological final category.

2.2.2. Gender-stereotypes and self-ratings

The gender-stereotype questionnaire was adopted from Halpern and Tan (2001). Participants in the experimental group were told to imagine that they were about to meet a person who they had never met before, based on each cognitive ability item (e.g., "is able to understand concepts in physics") participants were required to estimate the probability that the individual was male or female. Two columns were aligned next to each item—labeled *male* and *female*—and participants entered a number that corresponded to their probability estimate, with 100 the sum of the two estimates. Therefore a probability estimate of 50% for men and 50% for women would indicate that the participant perceived no stereotypical gender differences in this particular cognitive domain. In distinction from Halpern and Tan (2001), 16 items that were directly related to the sex-sensitive cognitive tests (mental rotation and verbal fluency) were used in the present study. All items were written in German. Items and mean probability estimates are shown in Table 2. In contrast to the gender-stereotype questionnaire, participants in the control group estimated the probability that the same selected cognitive abilities would be more or less associated with being "North American" or "European".

In line with Halpern and Tan (2001) self-ratings on the 16 items were also measured, using a seven-point scale, with 1 = *not at all descriptive of me* to 7 = *highly descriptive of me*. Self-ratings were measured to assess the possibility that participants may believe that an ability is generally associated with one sex or the other, but that as an individual, he or she is an exception to these stereotypes.

At the end of the experiment, all participants were asked to complete the gender-stereotype questionnaire, constituting the sole completion for the control group and the second completion for the experimental group. This procedure allowed measurement of gender-stereotypes in the control group without activating such stereotypes before cognitive testing. To rule out the possibility that the experiment had an impact on gender-stereotypes, the experimental group again completed the gender-stereotype questionnaire following cognitive testing.

2.2.3. Cognitive tests

2.2.3.1. Mental Rotation Test (MRT-3D). We used the *Revised Vandenberg and Kuse Mental Rotations Tests*—Version MRT-A (Peters, 1995) that involved 3D cube figures (Shepard and Metzler, 1971). The MRT-A contains two sets of 12 items. For each set, participants have a time limit of 3 min, with a 3-min break between the two sets. Each item consists of a target figure on the left and four stimulus figures on the right. Two of these stimulus figures are rotated versions of the target figure, and two of the stimulus figures cannot be matched to the target figure. One point is given if both matching stimulus figures are correctly identified. Thus, the maximum score in this test is 24 points. More information about the Revised Vandenberg and Kuse Mental Rotations Tests can be found in Peters et al. (1995a,b). All instructions for the MRT were translated into German by native German speakers who were highly proficient in English.

2.2.3.2. Mirror pictures (MP-2D). This subtest of the *WILDE-Intelligenz-Test* (Jäger and Althoff, 1994) is a 2D mental rotation test. The test contains 24 items, with five simple line drawings in each. Four of the five stimuli show the identical figure but rotated, and one stimulus shows a figure that is the mirror image which cannot be rotated to form a figure identical to the other four line drawings. The participants have three minutes to detect as many mirrored figures as possible. One point is given if the mirrored stimulus figure is identified correctly, resulting in a maximum score of 24 points.

2.2.3.3. Four-word sentences (4W). The four-word sentences are a subtest of the *Verbaler—Kreativitäts-Test* (Schoppe, 1975). The paper-pencil test contains two items with four letters each (T-G-F-U and B-H-K-N). Participants have 150 s per item to build as many four-word sentences as possible, each word starting with one of the given letters. The sentences do not need to be highly meaningful but have to be grammatically correct. One point is given for each correct sentence.

2.2.3.4. Word fluency (WF). Word fluency is a subtest of the *Leistungsprüfsystem* (LPS; Horn, 1962). Participants successively receive two letters (L and P) and have one minute per letter to generate as many words (excluding names) as possible in a written format. One point is given for each correct word.

2.2.3.5. Perceptual speed (PS). Perceptual speed is a subscale of the *WILDE Intelligenz Test* (Jäger and Althoff, 1994). This paper-pencil test consists of 42 items, each containing three drawings of outlined faces. Two of these faces match, and one face differs from the other two in only one detail, e.g., one eye is missing. One point is given if the odd face is identified correctly. The maximum score possible is 42 points. The participants have three minutes to identify as many of the odd faces as possible.

2.2.4. Testosterone assays

Saliva samples were taken from each participant immediately after test session, about 60—80 min after participants were confronted with the gender-stereotype questionnaire. Saliva samples were stored at $-22\,^{\circ}\mathrm{C}$ until all participants had completed the experiment. Free T levels were determined with Chemiluminescence Immunoassay (LIA) by an independent professional hormone laboratory, with commercially available LIA kits. For T, intra-assay coefficients of variations (CVs) were 3.0% and 1.5%, inter-assay CVs were 4.0% and 7.0% for high and low T samples, respectively. From participants who agreed to saliva samples, only samples not contaminated with blood and with the required ph-level were analysed by a professional hormone laboratory. Free T levels were available from 96 participants (45 men, 51 women).

3. Results

3.1. Effects of mood

Mood scores were subjected to a 2×2 ANOVA with sex and condition (stereotype activated, stereotype not activated) as between-participant factors. However, neither the main effects of sex and condition, nor the interaction approached significance, all $F(1,110) < 2.09$, ns, suggesting that mood

Table 1 Mood scores and cognitive performance (means, standard deviations, 95%-confidence intervals (in parantheses), maximal score) in mental rotation (MRT-3D and MP-2D), verbal fluency (4W and WF), and perceptual performance (PS) according to sex (males, females) and condition (experimental group, controls).

Tasks	Experimental group (stereotypes activated)		Controls (stereotypes not activated)		Max.
	Males	Females	Males	Females	
MRT-3D[a]	13.21 ± 4.64 (11.51–14.90)	8.68 ± 4.04 (7.25–10.10)	11.03 ± 4.16 (9.54–12.52)	9.72 ± 3.96 (8.06–11.38)	24
MP-2D[a]	16.83 ± 5.95 (14.58–19.09)	14.65 ± 5.15 (12.75–16.54)	14.27 ± 5.73 (12.25–16.28)	12.60 ± 5.58 (10.39–14.81)	24
4W[b]	7.71 ± 2.56 (6.44–8.98)	8.94 ± 3.75 (7.87–10.01)	6.42 ± 3.15 (5.30–7.54)	8.96 ± 2.69 (7.72–10.21)	—
WF[b]	21.75 ± 5.67 (19.43–24.07)	23.94 ± 6.52 (22.00–25.89)	20.03 ± 5.34 (17.99–22.07)	23.52 ± 5.04 (21.25–25.79)	—
PS	16.75 ± 4.96 (15.01–18.49)	17.00 ± 4.00 (15.54–18.46)	15.81 ± 4.83 (14.28–17.34)	16.44 ± 3.16 (14.74–18.14)	42
Mood	13.96 ± 6.08 (10.76–17.16)	14.50 ± 8.39 (11.81–17.19)	12.45 ± 7.44 (9.64–15.27)	16.24 ± 9.26 (13.10–19.38)	0–56

[a] Tasks favoring men.
[b] Tasks favoring women. Mood scores can range between euphoric (0) and extremely depressive (56).

did not differ between groups. Means and standard deviations are shown in Table 1.

3.2. Gender-stereotypes and self-ratings

For the experimental group, mean probability estimates of the gender-stereotype questionnaire, measured before and after cognitive testing, were analyzed separately. For the control group, mean probability estimates were only measured after cognitive testing. Before cognitive testing, the control group received a gender-neutral, nationality-stereotype questionnaire. To avoid alpha-error inflation, the significance level for the subsequent analyses was set to 1%. The results of the gender-stereotype questionnaire and self-ratings are shown in Table 2. This table also includes the results based on all participants, including the control group. The analyses of gender-stereotypes are based on probability estimates of being male.

To analyse whether sex differences in the perception of gender-stereotypes exist, unpaired t-tests were performed between the sexes to compare probability estimates (of the individual being male for each item). The analyses revealed no sex differences for any of the 16 items, all $t(55) < -1.15$, ns, indicating that implicit gender-stereotypes were very similar for males and females in the experimental group before cognitive testing. When analysing sex differences in gender-stereotypes for all participants after cognitive testing, only item 10 (*... can imagine abstract objects and rotate them mentally in all directions*) was significant, $t(112) = -3.08$, $p = .003$. Although this gender-stereotype exists for males, 60.6% (S.D. = 12.3), it was even more pronounced for females, 67.9% (S.D. = 13.1).

Analyses of sex differences in self-ratings revealed few significant effects for the experimental group. Self-rankings of females were significantly better than those of males for item 3 (*I can easily remember the names of guests at a party*; $t(55) = -3.04$, $p = .004$), and item 12 (*I can generate many words beginning with the same letter in one minute*; $t(55) = -3.30$, $p = .002$). Males showed slightly better self-rankings than females for item 5 (*I can draw a map of the area where I'm living*), and item 10 (*I can imagine abstract objects and rotate them mentally in all directions*), although, sex differences for these items only approached significance, $t(55) = 2.12$, $p = .038$, and $t(55) = 2.50$, $p = .015$, respectively. No other sex differences in self-rankings approached significance, all $t(55) < 2.50$, ns.

To further investigate whether males and females in the experimental group had pronounced gender-stereotypes, one-sample t-tests were calculated separately for males and females. In these analyses, probability estimates were compared to a probability estimate of 50%. In both males and females, five items revealed probability estimates which differed significantly from 50%: item 3 (*... can easily remember the names of guests at a party*; males: 38.3% (S.D. = 12.4), $t(23) = -4.61$, $p < .001$, females: 39.1% (S.D. = 11.4), $t(33) = -5.58$, $p < .001$), item 5 (*... can draw a map of the area where he/she lives*; males: 64.4% (S.D. = 13.2), $t(23) = 5.33$, $p < .001$, females: 63.1% (S.D. = 11.3), $t(33) = 6.72$, $p < .001$), item 9 (*... can speak three different languages*; males: 37.1% (S.D. = 10.5), $t(23) = -6.02$, $p < .001$, females: 34.7% (S.D. = 12.3), $t(33) = -7.25$, $p < .001$), item 10 (*... can imagine abstract objects and rotate them mentally in all*

Table 2 Gender-stereotype questionnaire items and mean probability estimates for the experimental group (before cognitive testing), and for all participants (after cognitive testing; in brackets).

Item	Probability male		Self-ratings (0–7)	
	Male	Female	Male	Female
"You are going to meet a person whom you have never met before. What is the probability that this person is male of female given that this person ..."				
1. ... has problems recognizing a complicated drawing when he/she sees it upside-down	41.3 (41.3*)	38.7* (45.1)	3.61 (3.43*)	4.03 (3.93)
2. ... can imagine common objects from different perspectives	55.8 (55.9*)	52.7 (55.3)	5.17* (5.28*)	5.26* (4.92*)
3. ... can easily remember the names of guests at a party[b]	38.3* (36.7*)	39.1* (36.7*)	3.22 (3.39*)	4.62 (4.93*)
4. ... often makes spelling mistakes	53.1 (52.1)	56.8* (55.5*)	3.22 (3.65)	2.44* (3.08*)
5. ... can draw a map of the area where he/she lives	64.4* (62.0*)	63.1* (61.0*)	6.00* (5.89*)	5.24* (5.03*)
6. ... is bad at reading street maps	44.6 (43.7)	39.9* (41.4*)	2.96 (2.98*)	3.41 (3.39)
7. ... has problems to summarize a book or movie in a short and clear manner	47.3 (49.7)	50.0 (50.4)	3.70 (3.34)	3.18 (3.37*)
8. ... does not use landmarks for orientation	50.6 (48.3)	52.8 (47.9)	3.13 (3.04*)	3.26 (3.39)
9. ... can speak three different languages fluently	37.1* (38.1*)	34.7* (36.8*)	3.91 (3.59)	4.62 (4.63*)
10. ... can imaging abstract objects and rotate them mentally in all directions[a]	60.8* (60.6*)	65.3* (67.9*)	5.09* (4.91*)	3.91 (3.86)
11. ... often forgets where common objects like keys were put	49.4 (50.9)	51.9 (48.8)	4.13 (4.07)	3.38 (3.58)
12. ... can generate many words beginning with the same letter in one minute[b]	42.3* (42.4*)	41.9* (40.9*)	3.61 (3.72)	4.94* (4.78*)
13. ... finds it difficult to imagine common objects and rotate them mentally	41.7* (45.0)	44.6 (46.6)	3.17 (3.19*)	3.44 (3.61)
14. ... remembers the way based on left-right turnoffs	56.9 (56.3*)	58.4 (58.1*)	3.87 (4.06)	4.12 (4.19)
15. ... finds many synonyms to a specific term	52.7 (52.5)	54.4 (56.3*)	4.13 (3.96)	3.32 (3.37*)
16. ... can easily summarize the essentials from a newspaper article	48.3 (49.8)	49.9 (48.1)	4.83* (4.87*)	5.38* (5.14*)

Self-ratings for the experimental group and for all participants (in brackets), collected immediately after cognitive testing. Results are shown separately for male and female participants.
Probability estimates marked with asterisks indicate significant differences from 50% (*$p < .01$).
Bold self-ratings indicate significant differences from test score 4 (*$p < .01$).
[a] Significant sex difference in gender-stereotypes when all participants are included ($p = .003$).
[b] Significant sex differences in self-ratings $p < .01$.

directions; males: 60.8% (S.D. = 16.0), $t(23) = 3.32$, $p = .003$, females: 65.3% (S.D. = 14.5), $t(33) = 6.17$, $p < .001$), and item 12 (... *can generate many words beginning with the same letter in one minute*; males: 42.3% (S.D. = 10.5), $t(23) = -3.59$, $p = .002$, females: 41.9% (S.D. = 10.3), $t(33) = -4.58$, $p < .001$). Additionally, female probability estimates of item 1 (... *has problems recognizing a complicated drawing when he/she sees it upside down*; 38.7% (S.D. = 14.4), $t(33) = -4.63$, $p < .001$), item 4 (... *often makes spelling mistakes*; 56.8%, $t(33) = 3.78$, $p = .001$), and item 6 (... *is bad at reading maps*; 39.9% (S.D. = 17.2), $t(33) = -3.44$, $p = .002$) differed significantly from 50%. Male participants showed a further significant probability estimate in item 13 (... *finds it difficult to imagine common objects and rotate them mentally*; 41.7% (S.D. = 14.4), $t(23) = -2.83$, $p = .009$). No further probability estimates of males and females differed significantly from 50%, all $t < 2.72$, ns. Probability estimates which differed significantly from 50% when all participants were included are shown in Table 2.

Additionally, one-sample t-tests were performed to analyze whether self-ratings in the experimental group differed significantly from test-score 4, indicating that a participant self-rated his/her performance as superior or inferior from the average. The analyses revealed that males and females self-ranked their performance as superior for item 2 (*I can imagine common objects from different perspectives*; males: 5.17 (S.D. = 1.47), $t(23) = 3.84$, $p = .001$, females: 5.26 (S.D. = 1.46), $t(33) = 5.04$, $p < .001$), item 5 (*I can draw a map of the area where I'm living*; males: 6.00 (S.D. = 1.24), $t(23) = 7.72$, $t(23) = 7.72$, $p < .001$, females: 5.24 (S.D. = 1.39), $t(33) = 5.17$, $p < .001$), and item 16 (*I can easily summarize the essentials from a newspaper article*; males: 4.83 (S.D. = 1.34), $t(23) = 2.96$, $p = .007$, females: 5.38 (S.D. = 1.28), $t(33) = 6.30$, $p < .001$). Moreover, male participants rated their performance as superior on item 10 (*I can imagine abstract objects and rotate them mentally in all directions*; 5.09 (S.D. = 1.08); $t(23) = 4.81$, $p < .001$). No self-rating inferior to the average was given by males in the experimental group. Female participants additionally self-rated their performance as inferior to the average for item 4 (*I often make spelling mistakes*; 2.44 (S.D. = 1.56); $t(33) = -5.82$, $p < .001$). No other self-ratings of males and females differed significantly from test-score 4 (average performance), all $t < -2.68$, ns.

Product-moment correlations between-gender stereotypes and according self-ratings revealed only few significant relationships. Male participants in the experimental group showed significant *positive* relationships between gender-stereotypes and self-ranking for item 1 (*... has problems recognizing a complicated drawing when he/she sees it upside down*), $R = .57$, $p = .005$, and item 11 (*... often forgets where common objects like keys were put*), $R = .64$, $p = .001$. Item 1 and item 13 (*... finds it difficult to imagine common objects and rotate them mentally*) were significant when all male participants were included in the analyses, $R = .40$, $p = .003$, and $R = .63$, $p < .001$, respectively. Female participants revealed significant *negative* correlations between gender-stereotypes and self-ranking for item 10 (*... can imagine objects and rotate them mentally in all directions*), $R = -.44$, $p = .01$, and item 14 (*... remembers the way based on left/right turnoffs*), $R = -.46$, $p = .006$. Item 10 and item 14 were also significant when all female participants were included in the analyses, $R = -.41$, $p = .001$, and $R = -.48$, $p < .001$. No further significant correlations were found, neither in experimental groups, nor when control participants were also included in the analysis, all $R < \pm.33$, ns.

In sum, the main results regarding gender-stereotypes and self-ratings suggest that (a) gender-stereotypes exist for several cognitive domains, with better spatial abilities in males than females (in particular, item 5 and item 10), and better verbal abilities in females than males (in particular, item 3, item 9, and item 12). (b) Gender-stereotypes did not differ between males and females but were almost identical. The only exception of this rule is item 10 which refers directly to the MRT-3D. For this item, female participants had an even larger bias in favour of males and mental rotation, than male participants. (c) Self-ratings for cognitive abilities were relatively independent of implicit gender-stereotypes. For example, both male and female participants ranked their own performance as superior from the average, e.g. in item 5 (*I can draw a map of the area where I'm living*), although participants of both genders agreed that this is rather a male cognitive domain. Only few significant relationships between self-ratings and gender-stereotypes occurred, and when they did, they were even negatively related for female participants.

Finally, it should be noted that gender stereotypes of the experimental group were almost identical before and after cognitive testing. Paired t-tests revealed a significant difference only for item 9 (*... can speak three different languages*), $t(57) = -2.25$, $p = .028$. Repeated measures did not reveal any significant differences for any of the other 15 items, all $t(57) < 1.76$, ns, indicating that cognitive testing did not significantly affect gender stereotypes. Moreover, independent t-tests for each item of the gender-stereotype questionnaire (after cognitive testing) revealed no significant differences between groups, all $t(112) < -1.50$, ns, indicating almost identical gender stereotypes for the experimental and control group.

3.3. Cognitive sex differences

A 2 × 2 ANOVA with sex and condition (experimental group with gender-stereotypes activated, control group with gender-stereotypes not activated) as between-participant factors was performed for each task. The results are shown in Table 1. For better comparisons across tasks, z-scores of cognitive performances are shown in Fig. 1. Overall, the cognitive tests revealed the expected sex differences (Table 1). Males outperformed females in MRT-3D (mental rotation), $F(1,110) = 13.57$, $p < .0001$, $\eta^2 = .11$, and MP-2D (mirror pictures), $F(1,110) = 3.31$, $p = .036$ (one-tailed), although the effect size was small, $\eta^2 = .03$. In contrast, females outperformed males in 4W (four word sentences), $F(1,110) = 10.07$, $p = .002$, $\eta^2 = .08$, and WF (word fluency), $F(1,110) = 6.68$, $p = .01$, $\eta^2 = .06$. No significant sex difference was found for PS (perceptual speed), $F(1,110) = 0.30$, ns.

The interaction between sex and condition was significant for MRT-3D, $F(1,110) = 4.12$, $p = .045$, $\eta^2 = .04$. Due to this significant interaction, alpha-adjusted post hoc t-tests were performed. The analyses revealed a strong male advantage in MRT-3D for the experimental group, $t(56) = 3.95$, $p = .0002$, $d = .98$. Although males also showed better performance than females in the control group, this effect did not approach significance, $t(54) = 1.20$, ns, $d = .31$. Moreover, posthoc t-tests revealed a slightly better performance for the gender-stereotype activated males than control males, $t(53) = 1.83$, $p = .07$, $d = .47$. Although females in the experimental group revealed numerically lower scores than female controls, this difference did not approach significance either, $t(57) = -0.99$, ns, $d = -.26$.

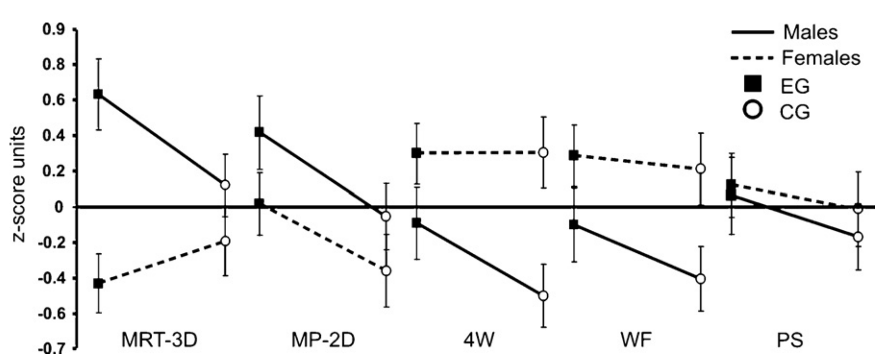

Fig. 1 Mean performances (z-scores and standard error means) in 3D- and 2D- mental rotation (MRT, SP), verbal fluency (4W, WF) and perceptual speed (PS). Dotted lines represent women and solid lines men. Black squares represent participants in the gender-stereotype-active group, open circles control participants. Positive and negative z-scores represent cognitive performance above or below average, respectively.

The interaction between sex and condition did not approach significance for the other cognitive tasks, all $F(1,110) < 1.21$, ns, $\eta^2 < .011$. The main effect of condition was only significant for MP-2D, $F(1,110) = 4.74$, $p = .032$, $\eta^2 = .042$, indicating higher performance for the experimental than control group. For the other cognitive tasks, the main effect of condition was not significant, all $F(1,110) < 1.14$, ns, $\eta^2 < .01$. None of the reported effects changed significantly, when mood scores were included as a covariate.

3.4. Hormone assay

Free T levels were based on saliva samples collected immediately after test session and were available from 96 participants (45 men, 51 women). T levels were subjected to a 2×2 ANOVA with sex and condition (stereotype activated, stereotype not activated) as between-participant factors. The ANOVA revealed a significant sex difference in saliva T levels (Fig. 2), $F(1,91) = 93.89$, $p < .001$, $\eta^2 = .51$, with higher T levels in males, 90.56 pg/ml (S.D. = 49.28), than in females, 24.15 pg/ml (S.D. = 20.87). Moreover, the main effect of condition was significant, $F(1,91) = 7.29$, $p < .01$, $\eta^2 = .07$. The mean T level was higher in the experimental group, 57.47 pg/ml (S.D. = 58.73), than in the control group, 51.90 pg/ml (S.D. = 37.14). Finally, the interaction between sex and condition was significant, $F(1,91) = 11.22$, $p = .001$, $\eta^2 = .11$. Alpha-adjusted posthoc t-tests revealed that the interaction was mainly driven by males in the experimental group, who showed significantly higher T levels, 115.04 pg/ml (S.D. = 56.88), than males in the control group, 71.96 pg/ml (S.D. = 33.01), $t(42) = 3.16$, $p < .003$, $d = 0.93$, and females in the experimental group, 22.19 pg/ml (S.D. = 17.54), $t(48) = 8.50$, $p < .001$, $d = 2.21$. Mean T levels for the male control group also differed from the female control group, 26.82 pg/ml (S.D. = 25.03), $t(43) = 5.06$, $p = .001$, $d = 1.54$. No differences in T levels were found between female groups, $t(49) = -0.78$, ns.

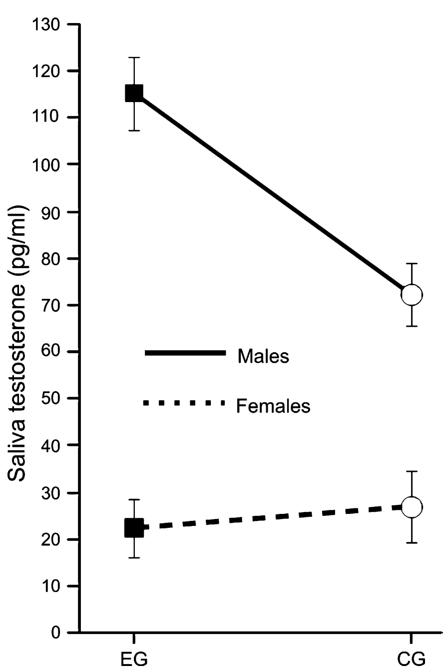

Fig. 2 Mean concentration and standard error means of testosterone of the gender-stereotype-active experimental (EG) and control group (CG). Solid and dotted lines indicate testosterone levels for males and females, respectively.

3.5. Sex hormones/behavior relationships

Multiple linear regressions were carried out, with cognitive performance for each test (MRT-3D, MP-2D, 4W, WF, and PS) acting as the dependent variable and T levels, gender-stereotypes, and mood scores acting as predictors. The regressions were performed for all participants, for those of whom hormone levels were available and for each sex separately. To form a measure of gender-stereotypes as a predictor, two items from the questionnaire were used. Probability estimates of being male, measured after cognitive testing, for item 10 (... *can imagine abstract objects and rotate them mentally in all directions*) were used when analysing MRT-3D and MP-2D, and probability estimates for item 12 (... *can generate many words beginning with the same letter in one minute*) were used when analyzing 4W and WF. The decision to use these gender-stereotype items was based on the finding that (a) males and females of the experimental group agreed on the probability estimates of these items, suggesting true implicit gender-stereotypes, and (b) both items refer directly to the cognitive test used as the dependent variable in the regression. No gender-stereotype predictor was used for PS, because none of the gender-stereotype items referred directly to perceptual speed.

Multiple regressions were based on gender stereotypes of both the experimental and control groups. The gender-stereotypes of the control group were tested only once after cognitive testing (see above), and thus this variable was used to predict past cognitive behavior. However, it should be noted that gender-stereotypes neither differed before and after cognitive testing for the experimental group, nor differed between the experimental and the control group after cognitive testing. This indicates that gender-stereotypes are relatively stable. The results of the multiple regressions are summarized in Table 3.

For males, multiple regression revealed a significant model for MRT-3D, $F(3,40) = 3.82$, $p = .017$, accounting for 22% of variance. T levels, $\beta = .34$, $p = .026$, and mood, $\beta = -.29$, $p = .045$, contributed significantly to the regression equation. Gender-stereotype (probability estimates of item 10), $\beta = -.05$, was not a significant predictor. Multiple regression in males revealed similar, but not significant results for MP-2D, $F(3,39) = 2.47$, $p = .077$, accounting for 16% of variance. The contribution of T levels and mood to the regression approached significance, $\beta = .28$, $p = .072$, and $\beta = .27$, $p = .075$, respectively. Again, the gender-stereotype (probability estimates of item 10) did not contribute significantly to the model, $\beta = -.02$, ns. Regression analyses for 4W, WF, and PS revealed no significant model, all $F < 2.11$, ns, and none of the predictors were significant, all $\beta < \pm.27$, ns.

In contrast to males, multiple regressions revealed no significant model for females, all $F < 1.75$, ns. However, for MRT-3D performance, regression revealed the gender-stereotype (probability estimates of item 10), to be a significant predictor, $\beta = -.30$, $p = .038$. The lower the performance in MRT-3D, the more the gender-stereotype of item 10, in favour of men, was pronounced. Additionally, regression revealed mood to be a marginal predictor for PS,

Table 3 Multiple linear regressions (standardized beta coefficients) for testosterone levels (T-levels), gender-stereotypes, and mood as predictors of cognitive performance in males (M) and females (F).

Task	Sex	T-levels	Stereotype	Mood	R^2	p
MRT	M	.34*	.05[a]	−.29*	.22	.02
	F	−.00	−.30*, [a]	−.15	.10	ns
MP	M	.28+	−.02[a]	−.27+	.16	.08
	F	−.01	.06[a]	−.16	.03	ns
4W	M	.15	−.10[b]	−.04	.04	ns
	F	−.04	−.07[b]	−.10	.02	ns
WF	M	.27	.04[b]	−.22	.13	ns
	F	−.18	.02[b]	−.06	.04	ns
PS	M	.27+	n/a	−.12	.09	ns
	F	.07	n/a	.25+	.07	ns

Determination coefficients (R^2) and significances (p) indicate the goodness-of-fit for the regression model.
[a] Gender-stereotypes are based on probability estimates of item 10.
[b] Gender-stereotypes are based on probability estimates of item 12.
* $p < .05$.
+ $p < .10$, ns (not significant), n/a (not available).

$\beta = -.25$, $p = .08$. No further significant predictors on cognitive performance were present, all $\beta < \pm.18$, ns.

In sum, multiple regressions showed that MRT-3D performance for men was positively related to T levels and negatively related to mood scores. The same effect approached significance for MP-2D. For females, the gender stereotype (item 10) was negatively related to MRT-3D performance. Cognitive sex differences in both verbal fluency tasks (WF and 4W) were neither related to T levels, mood nor the probability estimates of the gender-stereotype questionnaire (item 12). For PS, men revealed a marginal relationship to T levels, whereas mood was the best predictor for females.

4. Discussion

Overall, the present study revealed expected cognitive sex differences. Males outperformed females on two mental-rotation tasks, whereas females outperformed males on two measures of verbal creativity, i.e. verbal fluency. No significant sex differences were found for perceptual speed. These cognitive sex differences are in agreement with previous studies, and effect sizes lie within the range of mean effect sizes of previous meta-analyses (Linn and Petersen, 1985; Hyde and Linn, 1988; Voyer et al., 1995, see also Kimura, 1999; Halpern, 2000; Hines, 2004; Hausmann, 2007, for review). For instance, Peters et al. (2006) found an effect size between $d = 0.62$ and d1.1 for the MRT-3D in four cross-cultural samples. Interestingly, the sex difference in the MRT-3D (Peters, 1995) was particularly pronounced when gender-stereotypes were activated before cognitive testing.

In line with previous studies focusing on stereotype threat and stereotype lift, men improved their performance and women's performance deteriorated on the MRT-3D, following activation of the stereotype. Overall, men performed better on this test than women. However, in the control condition, no significant differences between the sexes were found. The effect appears to result from differences observed in the experimental condition, with women underperforming compared to men (in line with stereotype threat). This effect does not extend to the other cognitive tasks. Similar stereotype-threat/stereotype-lift effects were not observed for the MP-2D, or either of the verbal fluency tasks.

The difference in stereotype effects between the MRT-3D and the MP-2D is surprising, given that both tasks consist of similar mental rotation processes. However, both tasks also differ in various aspects. For example, it is clear from the number of correctly solved items (Table 1), that participants obtain much higher scores in the MP-2D than MRT-3D, suggesting that the MRT-3D is a much more difficult task. This difference is especially apparent when one considers that the amount of time participants have to perform the MP-2D is double that of the MRT-3D. Task difficulty might be of particular relevance here, as previous studies have shown that stereotype threat affects only tasks which are very difficult to perform (Spencer et al., 1999; Blascovich et al., 2001; Keller, 2007).

A possible explanation for the finding that stereotype threat/stereotype-lift occurred only for the MRT-3D can be found with the results for the gender-stereotype questionnaire. The questionnaire revealed the expected gender stereotypes, men were perceived as superior at spatial tasks and women were perceived as superior in verbal aspects, which was the case for both male and female respondents. The only item that did receive gender differences in response asked "*... can imagine objects and rotate them mentally in all directions*" (item 10), which directly refers to the MRT-3D task. For this item, female participants had an even larger bias in favour of males and mental rotation, than male participants. However, what is relevant here is not only the finding of sex difference in probability estimates for item 10 but also that both sexes perceive the stereotype. This could be seen as evidence that the strength of the stereotype is important for the occurrence of stereotype threat. That is, the stronger the stereotype, the more likely that stereotype threat/lift will occur.

The activation of stereotypes might explain why the magnitude of sex differences in sex-sensitive cognitive task varies across studies, depending on whether participants' gender-stereotypes are activated or not. It is likely that gender-stereotypes are implicitly active in most studies focusing on cognitive sex differences, because test instructions and announcements to recruit participants are not formulated neutrally in terms of gender. Gender-stereotype research indicates that implicit activation of a gender identity can facilitate as well as impede cognitive performance (e.g. Shih et al., 1999). This research also suggests that when a particular stereotype was made salient at an implicit level, performance was altered in the direction predicted by the stereotype associated with being male or female.

However, the same argument did not apply to verbal fluency. One reason might be that the gender-stereotype referring to mental rotation, i.e. item 10, is more pronounced, and consequently more salient than the stereotypes referring to verbal fluency, i.e. item 12. This explanation is rather speculative but the results indicate that the interaction between gender stereotypes and sex hormones are rather complex and can differ between cognitive domains.

Although sex differences in MRT-3D are known to be very consistent (Voyer et al., 1995; Yang et al., 2007), the control group revealed no significant sex difference in this test, albeit the effect size of $d = .31$ still suggests a slight sex difference favoring men. This may appear surprising, however, it should be noted that the control group received a stereotype questionnaire about nationality. This manipulation might have resulted in moving the attention away from underlying implicit gender stereotypes towards nationality stereotypes. This is of particular interest because previous research suggests that only one stereotype is active at a time (e.g. Shih et al., 1999), which suggests that the gender stereotype was not activated in the control group. Interestingly, the stereotype that Europeans outperform North Americans was present in control men (item 10: probability estimate of 53.23%, $p < .001$). Women of the control group revealed a probability estimate (item 10) of 50%, suggesting that female controls were not affected by stereotype threat. This might at least partly explain the reduced sex difference in MRT-3D performance in the control group. Moreover, it suggests that a gender-neutral control intervention might also affect cognitive performance.

There might be also an alternative and perhaps more provocative explanation. Only very few studies have controlled for gender-stereotype priming, which may suggest that some other studies might have implicitly activated gender stereotypes. If this is true, the well known sex differences in MRT-3D would have been overestimated in the past. That does not of course mean that sex differences in MRT-3D performance are simply a product of stereotype activation. As already mentioned above, several social and biological performance factors have been shown to affect the size of sex differences in MRT-3D performance. However, the present study suggests that implicit stereotype priming can have an important influence.

In the present study, gender stereotypes were very similar for males and females. Only one single item on the stereotype questionnaire (item 10) revealed a clear sex difference when all participants were included. Females were even more convinced than males that a person is more likely to be male given that this person can imagine abstract objects and rotate them mentally in all directions. This stereotype also existed for males but to a lower extent. The fact that this gender-stereotype was particularly strong for females might be the reason why this item was the best predictor for MRT-3D performance in females. In general, however, the magnitude of a specific gender-stereotype seems to be only marginally related to its corresponding cognitive test performance. Although pronounced gender-stereotypes of better verbal abilities in females exist, explicit activation of those did not affect sex differences in verbal fluency.

In contrast to females, levels of sex hormones were the best predictors for cognitive performance in males. Specifically, free T was positively related to spatial abilities. This was particularly the case for mental rotation, as measured by the Revised Vandenberg and Kuse Mental Rotations Tests (Peters, 1995). Although inconsistent findings about the relationship between current T levels and performance on spatial tasks for males do exist (see introduction for details), a positive linear relationship between these factors shown in the present study is in line with several previous reports (e.g. Gordon and Lee, 1986; Silverman et al., 1999; Hooven et al., 2004). It is noteworthy that males in the experimental and control group did not only differ in mental rotation (MRT-3D) performance, but also differed significantly in T levels. Free T levels were higher in males in the experimental group. This finding might suggest that an activation of gender stereotypes can lead to an increase in T levels which then modulates performance of sex-sensitive cognitive tests which are known to be sensitive to free T levels.

Although stereotype activation seems to be related to T levels, the direction of this relationship cannot be established. While it is possible that the T levels increase as a consequence of stereotype activation, we cannot rule out that the differences in the hormonal levels results in an increase in susceptibility to stereotype activation. However, it might also be that gender-stereotype challenge leads to better performance which leads to higher testosterone. Such a relationship might also contribute to 60% higher T levels in men of the experimental group. Unfortunately, this study cannot directly show that T levels increased as a consequence of gender-stereotype activation. To determine an increase in T levels, measuring initial T levels before and after cognitive testing would have been required. Although the present study measured hormone levels only once after cognitive testing, it seems unlikely that group differences in T levels between stereotype-active males and males in the control group are simply due to a selection bias. Firstly, all participants were randomly assigned to either the experimental (gender-stereotype activated) or control group. Secondly, the T levels of gender-stereotype activated males are 60% higher than that of the control group which is difficult to explain with random effects. Thirdly, the sample sizes are sufficiently large and suggest differences in mean hormone levels were not driven by only a few male participants.

It is also important to bear in mind that participants were tested in groups of mixed gender. This, in combination with the fact that gender stereotypes were made salient might have established a competitive situation between males and females. Previous studies have shown that fluctuations of T levels can be induced by competition in men and women (e.g., Bateup et al., 2002; Archer, 2006), and suggest that in competitive situations, changes in T levels are not exclusively a direct response to the outcome, but rather a response to the contribution the individual makes to the outcome and to the causes the person attributes (Gonzalez-Bono et al., 1999; Archer, 2006; Mehta and Josephs, 2006). It is interesting to note that men get (numerically) better on all tasks in response to stereotype activation. This might indicate that activation of stereotypes (and possibly increase in T) generally increases men's effort on all tasks. However, only spatial tasks were significantly affected by stereotype activation. Further studies should explore if the T increase found in this study is caused by gender-stereotype activation and/or the competition situation.

To our knowledge this is the first study demonstrating an interaction between sex hormones, gender stereotypes and cognitive abilities. The combined influence of T and gender stereotypes was particularly pronounced in one of the most sex-sensitive paper-pencil test, i.e. the Revised Vandenberg and Kuse Mental Rotations Tests (Peters, 1995). In line with previous reports, mental rotation performance was positively related to T levels, a finding which was particularly pronounced in men. However, sex differences in MRT-3D occurred only if participants' gender stereotypes were activated experimentally. The present study suggests that an activation of gender stereotypes can lead to an increase in T levels in men which then modulate performance in a sex-sensitive cognitive test which is known to be sensitive to free testosterone levels. Whether this gender-stereotype induced effect is indeed mediated by sex hormones needs to be directly proved in future research. However, the results of the present study clearly indicate that the understanding of the origins of cognitive sex differences does not simply involve determining the proportion of nature and nurture but rather requires understanding how these entities interact within a psychobiosocial approach.

Role of the funding sources

Funding for this study was provided by Grant HA 3285/4-1 of the Deutsche Forschungsgemeinschaft (DFG). The DFG had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Conflicts of interest

None declared.

Acknowledgement

The authors thank all participants for their cooperation. This work is supported by grant HA 3285/4-1 of the Deutsche Forschungsgemeinschaft (DFG).

References

Aleman, A., Bronk, E., Kessels, R.P.C., Koppeschaar, H.P.F., van Honk, J., 2004. A single administration of testosterone improves visuospatial ability in young women. Psychoneuroendocrinology 29, 612—617.

Alexander, G.M., Swerdloff, R.S., Wang, C., Davidson, T., McDonald, V., Steiner, B., Hines, M., 1998. Androgen-behavior correlations in hypogonadal men and eugonadal men. Horm. Behav. 33, 85—94.

Archer, J., 2006. Testosterone and human aggression: an evaluation of the challenge hypothesis. Neurosci. Biobehav. Rev. 30, 319—345.

Bateup, H.S., Booth, A., Shirtcliff, E.A., Granger, D.A., 2002. Testosterone, cortisol, and women's competition. Evol. Hum. Behav. 23, 181—192.

Blascovich, J., Spencer, S.J., Quinn, D., Steele, C., 2001. African Americans and high blood pressure: the role of stereotype threat. Psychol. Sci. 12, 225—229.

Brown, R.P., Josephs, R.A., 1999. A burden of proof: Stereotype relevance and gender differences in math performance. J. Pers. Soc. Psychol. 76, 246—257.

Burkitt, J., Widman, D., Saucier, D.M., 2007. Evidence for the influence of testosterone in the performance of spatial navigation in a virtual water maze in women but not in men. Horm. Behav. 51, 649—654.

Burns, N.R., Nettelbeck, T., 2005. Inspection time and speed of processing: sex differences on perceptual speed but not IT. Pers. Ind. Diff. 39, 439—446.

Christiansen, K., 1993. Sex hormone-related variations of cognitive performance in !Kung San hunter-gatherers of Namibia Neuropsychobiology 27, 97—107.

Collins, D.W., Kimura, D., 1997. A large sex difference on a two-dimensional mental rotation task. Behav. Neurosci. 111, 845—849.

Ekstrom, R.B., French, J.W., Harman, H.H., 1976. Manual For Kit of Factor-Referenced Cognitive Tests. Educational Testing Service, Princeton, NJ.

Epting, L.K., Overman, W.H., 1998. Sex-sensitive tasks in men and women: a search of performance fluctuations across the menstrual cycle. Behav. Neurosci. 112, 1304—1317.

Falter, C.M., Arroyo, M., Davis, G.J., 2006. Testosterone: activation or organization of spatial cognition? Biol. Psychol. 73, 132—140.

Feingold, A., 1992. Sex differences in variability in intellectual abilities: a new look at an old controversy. Rev. Educ. Res. 62, 61—84.

Gonzalez-Bono, E., Salvador, A., Serrano, M.A., Ricarte, J., 1999. Testosterone, cortisol, and mood in a sports team competition. Horm. Behav. 35, 55—62.

Gordon, H.W., Lee, P.L., 1986. A relationship between gonadotropins and visuospatial function. Neuropsychologia 24, 563—576.

Gordon, H.W., Lee, P.L., 1993. No differences in cognitive performance between phases of the menstrual cycle. Psychoneuroendocrinology 18, 521—531.

Gordon, H.W., Corbin, E.D., Lee, P.A., 1986. Changes in specialized cognitive function following changes in hormone levels. Cortex 22, 399—415.

Gouchie, C., Kimura, D., 1991. The relationship between testosterone levels and cognitive ability patterns. Psychoneuroendocrinology 16, 323—334.

Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., Sharma, T., 2005. Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins. Behav. Neurosci. 119, 104—117.

Halpern, D.F., 1996. Sex, brains, hands, and spatial cognition. Dev. Rev. 16, 261—270.

Halpern, D.F., 2000. Sex Differences in Cognitive Abilities. Lawrence Erlbaum Associates, Publishers, Mahawah, NJ.

Halpern, D.F., Tan, U., 2001. Stereotypes and steroids: using a psychobiosocial model to understand cognitive sex differences. Brain Cogn. 45, 392—414.

Hampson, E., 1990a. Estrogen-related variations in human spatial and articulatory-motor skills. Psychoneuroendocrinology 15, 97—111.

Hampson, E., 1990b. Variations in sex related cognitive abilities across the menstrual cycle. Brain Cogn. 14, 26—43.

Hampson, E., Rovet, J.F., Altmann, D., 1998. Spatial reasoning in children with congenital adrenal hyperplasia due to 21-hydrooxylase deficiency. Dev. Neuropsychol. 14, 299—320.

Hausmann, M., 2007. Kognitive Geschlechtsunterschiede. In: Lautenbacher, S., Hausmann, M., Güntürkün, O. (Eds.), Gehirn und Geschlecht—Neurowissenschaft des kleinen Unterschieds zwischen Frau und Mann. Springer Science, Heidelberg, pp. 105—124.

Hausmann, M., Slabbekoorn, D., Van Goozen, S.H.M., Cohen-Kettenis, P.T., Güntürkün, O., 2000. Sex hormones affect spatial abilities during the menstrual cycle. Behav. Neurosci. 114, 1245—1250.

Hedges, L.V., Nowell, A., 1995. Sex differences in mental test scores, variability, and number of high scoring individuals. Science 269, 41—45.

Hines, M., 2004. Brain Gender. University Press, Oxford.

Hines, M., Fane, B.A., Pasterski, V.L., Mathews, G.A., Conway, G.S., Brooke, C., 2003. Spatial abilities following prenatal androgen abnormality: targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. Psychoneuroendocrinology 28, 1010—1026.

Hooven, C.K., Chabris, C.F., Ellison, P.T., Kosslyn, S.M., 2004. The relationship of male testosterone to components of mental rotation. Neuropsychologia 42, 782—790.

Horn, W., 1962. Leistungsprüfsystem (LPS). Hogrefe, Göttingen.

Hyde, J.S., 1981. How large are cognitive gender differences? Am. Psychol. 36, 892—901.

Hyde, J.S., Linn, M.C., 1988. Gender differences in verbal abilities: a meta-analysis. Psychol. Bull. 104, 53—69.

Inzlicht, M., Ben-Zeev, T., 2000. A threatening intellectual environment: why females are susceptible to experiencing problem-solving deficits in the presence of males. Psychol. Sci. 11, 365—371.

Jäger, A.O., Althoff, K., 1994. Der WILDE-Intelligenz-Test (WIT). Hogrefe & Huber, Göttingen.

Josephs, R.A., Newman, M.L., Brown, R.P., Beer, J.M., 2003. Status, testosterone, and human intellectual performance: stereotype threat at status concern. Psychol. Sci. 14, 158—163.

Kampen, D.L., Sherwin, B.B., 1996. Estradiol is related to visual memory in healthy young men. Behav. Neurosci. 110, 613—617.

Keller, J., 2007. Stereotype threat in classroom settings: the interactive effect of domain identification, task difficulty and stereotype threat on female students' math performance. Br. J. Educ. Psychol. 77, 323—338.

Kimura, D., 1996. Sex, sexual orientation and sex hormones influence human cognitive function. Curr. Opin. Neurobiol. 6, 259—263.

Kimura, D., 1999. Sex and Cognition. MIT Press, Cambridge.

Koenig, A.M., Eagly, A.H., 2005. Stereotype threat in men on a test of social sensitivity. Sex Roles 52, 489—496.

Leyens, J.-P., Désert, M., Croizet, J.-C., Darcis, C., 2000. Stereotype threat: are lower status and history of stigmatization preconditions of stereotype threat? Pers. Soc. Psychol. Bull. 26, 1189—1199.

Liben, L.S., Susman, E.J., Finkelstein, J.W., Chinchilli, V.M., Kunselman, S., Schwab, J., et al., 2002. The effects of sex steroids on spatial performance: a review and an experimental clinical investigation. Dev. Psychol. 35, 236—253.

Linn, M.C., Petersen, A.C., 1985. Emergence and characterization of sex differences in spatial ability: a meta-analysis. Child Dev. 56, 1479—1498.

Maki, P.M., Rich, J.B., Rosenbaum, R.S., 2002. Implicit memory varies across the menstrual cycle: estrogen effects in young women. Neuropsychologia 40, 518—529.

Maccoby, E.E., Jacklin, C.N., 1974. The Psychology of Sex Differences. Stanford University Press, Stanford, CA.

Marx, D.M., Roman, J.S., 2002. Female role models: protecting women's math test performance. Pers. Soc. Psychol. Bull. 28, 1183—1193.

McGee, M.G., 1979. Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychol. Bull. 86, 889—918.

McGlone, J., 1980. Sex differences in human brain asymmetry: a critical survey. Behav. Brain Sci. 3, 215—263.

McKeever, W.F., 1995. Hormone and hemisphericity hypotheses regarding cognitive sex differences: possible future explanatory power, but current empirical chaos. Learn. Individ. Differ. 7, 323—340.

Mehta, H.P., Josephs, R.A., 2006. Testosterone changes after losing predicts the decision to compete again. Horm. Behav. 50, 684—692.

Moffat, S.D., Hampson, E., 1996. A curvelinear relationship between testosterone and spatial cognition in humans: possible influence of hand preference. Psychoneuroendocrinology 21, 323—337.

Mumenthaler, M.S., O'Hara, R., Taylor, J.L., Friedman, L., Yesavage, J.A., 2001. Relationship between variations in estradiol and progesterone levels across the menstrual cycle and human performance. Psychopharmacology (Berl.) 155, 198—203.

Perlman, S.M., 1973. Cognitive abilities of children with hormone abnormalities: screening by psychoeducational tests. J. Learn. Disabil. 6, 21—29.

Peters, M., 1995. Revised Vandenberg and Kuse Mental Rotations Tests: forms MRT-A to MRT-D. University of Guelph, Department of Psychology, Guelph, ON.

Peters, M., Chrisholm, P., Laeng, B., 1995a. Spatial ability, student gender and academic performance. J. Eng. Educ. 84, 69—73.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., Richardson, C., 1995b. A Redrawn Vandenberg and Kuse Mental Rotation Test: different versions and factors that affect performance. Brain Cogn. 28, 39—58.

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., Jordan, K., 2006. Mental rotation test performance in four cross-cultural samples (N = 3367): overall sex differences and the role of academic program in performance. Cortex 42, 1005—1014.

Resnick, S.M., Berenbaum, S.A., Gottesman, I.I., Bouchard, T., 1986. Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. Dev. Psychol. 22, 191—198.

Rosenberg, L., Park, S., 2002. Verbal and spatial functions across the menstrual cycle in healthy young women. Psychoneuroendocrinology 27, 835—841.

Rosenthal, H.E.S., Crisp, R.J., Suen, M.-W., 2007. Improving performance expectancies in stereotypic domains: task relevance and the reduction of stereotype threat. Eur. J. Soc. Psychol. 37, 586—597.

Schoppe, K., 1975. Verbaler-Kreativitäts-Test (VKT). Hogrefe, Göttingen.

Shepard, R.N., Metzler, J., 1971. Mental rotation of three-dimensional objects. Science 171, 701—703.

Shih, M., Pittinsky, T.L., Ambady, N., 1999. Stereotype susceptibility: identity salience and shifts in quantitative performance. Psychol. Sci. 10, 80—83.

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. J. Pers. Soc. Psychol. 38, 638—647.

Silverman, I., Phillips, K., Silverman, L.K., 1996. Homogeneity of effect sizes for sex across spatial tests and cultures: implications for hormonal theories. Brain Cogn. 31, 90—94.

Silverman, I., Kastuk, D., Choi, J., Phillips, K., 1999. Testosterone levels and spatial ability in men. Psychoneuroendocrinology 24, 813—822.

Slabbekoorn, D., Van Goozen, S.H.M., Megens, J., Gooren, L.J.G., Cohen-Kettenis, P.T., 1999. Activating effects of cross-sex hormones on cognitive functioning: a study of short-term and long-term hormone effects in transsexuals. Psychoneuroendocrinology 24, 423—447.

Spencer, S.J., Steele, C.M., Quinn, D.M., 1999. Stereotype threat and women's math performance. J. Exp. Soc. Psychol. 35, 4—28.

Steele, C.M., Aronson, J., 1995. Stereotype threat and the intellectual test performance of African Americans. J. Pers. Soc. Psychol. 69, 797—811.

Tapley, S.M., Bryden, M.P., 1977. An investigation of sex differences in spatial ability: mental rotation of three-dimensional objects. Can. J. Psychol. 31, 122—130.

Vandenberg, S.G., Kuse, A.R., 1978. Mental rotations: a group test of three-dimensional spatial visualization. Percept. Mot. Skills 47, 599—604.

Van Goozen, S.H.M., Cohen-Kettenis, P.T., Gooren, L.J.G., Frijda, N.H., Van de Poll, N.E., 1994. Activating effects of androgens on cognitive performance: causal evidence in a group of female-to-male transsexuals. Neuropsychologia 32, 1153—1157.

Voyer, D., Voyer, S., Bryden, M.P., 1995. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychol. Bull. 117, 250—270.

Walton, G.M., Cohen, G.L., 2003. Stereotype lift. J. Exp. Soc. Psychol. 39, 456—467.

Williams, C.L., Meck, W.H., 1991. The organizational effects of gonadal steroids on sexually dimorphic spatial ability. Psychoneuroendocrinology 16, 155—176.
Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., Karps, S.A., 1962. Psychological Differentiation. Wiley, New York.
Wraga, M., Duncan, L., Jacobs, E.C., Helt, M., Church, J., 2006. Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. Psychon. Bull. Rev. 13, 813—819.
Wraga, M., Helt, M., Jacobs, E., Sullivan, K., 2007. Neural basis of stereotype-induced shifts in women's mental rotation performance. Soc. Cogn. Affect. Neurosci. 2, 12—19.
Yang, C.-F.J., Hooven, C.K., Boynes, M., Gray, P.B., Pope, H.G., 2007. Testosterone levels and mental rotation performance in Chinese men. Horm. Behav. 51, 373—378.
Zerssen, D., 1976. Die Befindlichkeits-Skala. Beltz-Verlag, Weinheim.