Hormones and Behavior 58 (2010) 282–289

Contents lists available at ScienceDirect

Hormones and Behavior

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

Salivary testosterone does not predict mental rotation performance in men or women

David A. Puts [a,*], Rodrigo A. Cárdenas [b], Drew H. Bailey [c], Robert P. Burriss [a], Cynthia L. Jordan [b,d], S. Marc Breedlove [b,d]

[a] Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA
[b] Department of Psychology, Michigan State University, East Lansing, MI 48824, USA
[c] Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA
[d] Neuroscience Program, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 29 August 2009
Revised 3 March 2010
Accepted 4 March 2010
Available online 11 March 2010

Keywords:
Androgens
Mental rotation
Sex difference
Spatial ability
Spatial cognition
Testosterone

ABSTRACT

Multiple studies report relationships between circulating androgens and performance on sexually differentiated spatial cognitive tasks in human adults, yet other studies find no such relationships. Relatively small sample sizes are a likely source of some of these discrepancies. The present study thus tests for activational effects of testosterone (T) using a within-participants design by examining relationships between diurnal fluctuations in salivary T and performance on a male-biased spatial cognitive task (Mental Rotation Task) in the largest sample yet collected: 160 women and 177 men. T concentrations were unrelated to within-sex variation in mental rotation performance in both sexes. Further, between-session learning-related changes in performance were unrelated to T levels, and circadian changes in T were unrelated to changes in spatial performance in either sex. These results suggest that circulating T does not contribute substantially to sex differences in spatial ability in young men and women. By elimination, the contribution of androgens to sex differences in human performance on these tasks may be limited to earlier, organizational periods.

Introduction

In animal models, adult behavioral sex differences have generally been found to result from either the sex difference in circulating androgens such as testosterone (T) in adulthood, or the sex difference in exposure to androgens earlier in life, for example during the perinatal period (Morris et al., 2004). These two modes of producing behavioral sex differences have been described as "activational" and "organizational," respectively (Phoenix et al., 1959). Although there are interesting exceptions to these two modes of differentiation, and they often act through similar neural mechanisms (Arnold and Breedlove, 1985), the dichotomy nevertheless holds true in many cases.

Men and women perform similarly on tests of overall intelligence but differ on tests that measure specific cognitive abilities (Hines, 2004). The largest cognitive sex differences are found in the domain of spatial ability, with men tending to outperform women (Maccoby and Jacklin, 1974). These differences may be due partly to lasting organizational effects of prenatal or early postnatal androgens (Collaer et al., 2009; Puts et al., 2008), and/or pubertal androgens (Hier and Crowley, 1982). Some human and rodent evidence suggests a curvilinear relationship between androgen signaling and spatial ability, such that organizational effects of androgens improve performance on some spatial tasks in females and impair performance on these tasks in gonadally intact males (Puts et al., 2007).

Androgens may also have transient activational effects on spatial ability in adults, but evidence for this is equivocal. Several studies have found significant relationships between T levels and spatial ability in between-participants comparisons of adults (e.g., Christiansen, 1993; Christiansen and Knussmann, 1987; Driscoll et al., 2005; Gordon and Lee, 1986; Hausmann et al., 2009, 2000; Hooven et al., 2004; Moffat and Hampson, 1996a; Silverman et al., 1999), although others have not (e.g., Falter et al., 2006; Halari et al., 2005; Hassler et al., 1992; Janowski et al., 1998; Kampen and Sherwin, 1996; Matousek and Sherwin, 2010; McKeever and Deyo, 1990) (see Table 1 and Discussion). Moreover, correlations between adult androgen levels and spatial performance in between-participants studies leave questions about when during development androgen affects spatial ability. Testosterone production rate is highly heritable (Meikle et al., 1988), and it is therefore possible that intrasexual differences in circulating T persist throughout life. If so, associations between adult androgen levels and spatial ability may reflect prior organizational effects of hormones.

* Corresponding author. Fax: +1 814 863 1474.
E-mail address: dap27@psu.edu (D.A. Puts).

0018-506X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.yhbeh.2010.03.005

Table 1
Results of previous studies that have investigated between-subjects relationships between spatial ability and T. '0' indicates no relationship, '+' and '−' indicate positive and negative relationships, respectively. 'NL' indicates a nonlinear relationship (quadratic or third-order polynomial).

Study	Sample size	Mean age (SD, range)	Sample method	Samples, sessions	Task	Results
Alexander et al. (1998)	33M, 10M	41.1 (20–59); 33.4 (21–44)	Blood	"Multiple," 1	MRT	0
					Surface development	0
					Paper folding	0
					Hidden patterns	0
Burkitt et al. (2007)	39F, 36M	19.9 (2.8)	Saliva	2, 1	MRT	+
					Virtual Morris water task	F:+, M:0
Christiansen and Knussmann (1987)	110M	24.1 (20–30)	Blood	1, 2	Block design	0
					Leistungsprufsystem subtest 8	+
					Leistungsprufsystem subtest 9	0
			Saliva	1, 2	Block design	0
					Leistungsprufsystem subtest 8	0
					Leistungsprufsystem subtest 9	0
Christiansen (1993)	114M	26.4 (4.7, 18–38)	Blood	1, 1	Dichaptic stimulation test	+
			Saliva	1, 1	Dichaptic stimulation test	+
Driscoll et al. (2005)	16F, 16M	20–over 60	Saliva	1, 1	Virtual Morris water task	F:0, M:+
Falter et al. (2006)	22F, 24M	(19–41)	Saliva	1, 1	MRT	0
					Targeting	0
					Perceptual discrimination	0
	34F, 35M	F:24.1(4.6); M:24.1 (3.0)			Disembedding	F:+ for 1 of 2 tasks, M:0
Gordon and Lee (1986)	32M	(18–35)	Blood	11, 2	Cognitive laterality battery	
					– Localization	0
					– Orientation	Session 1:+, Session 2:0
					– Touching blocks	0
					– Form completion	0
Gouchie and Kimura (1991)	46F, 42M	F:21.5 (18–31); M:21.0 (18–27)	Saliva	2, 1	Paper folding	0
					MRT	0
Halari et al. (2005)	42F, 41M	F:27.69, (3.96); M:28.31 (4.81)	Blood	1, 1	MRT	0
					Computerized judgment of line orientation	0
					Modified judgment of line orientation	0
Hassler et al. (1992)	25F, 26M	F:18.77 (1.42); M:19.16 (1.65)	Blood	1, 1	Spatial relations test	0
					Hidden patterns	0
					Dichaptic stimulation test	0
Hausmann et al. (2000)	12F	29.1 (4.4, 23–38)	Blood	2, 1	MRT	+
					Mirrors pictures test	0
					Hidden figures test	0
Hausmann et al. (2009)	51F, 45M	23.4 (4.7), 25.8 (7.2)	Saliva	1, 1	MRT	F:0, M:+
Hooven et al. (2004)	28M	23 (4)	Saliva	1, 2	MRT	+
Janowski et al. (1998)	17F, 29M	F:29.8 (3.2, 24–34); M:28.5 (3.1, 23–34)	Blood	2, 2	Block design task	0
					Card rotation	0
Kampen and Sherwin (1996)	32M	21.1 (18–29)	Blood	1, 1	MRT	0
Klaiber et al. (1967)	50M	21.65 (0.92)	Urine	1, 2	Block design task	−
Matousek and Sherwin (2010)	54M	68.6 (4.4)	Blood	1, 1	MRT	0
					Paper folding	0
					Water level test	0
					Block design test	0
McKeever et al. (1987)	42F, 41M	Unstated	Blood	1, 1	Stafford identical blocks test	0
McKeever and Deyo (1990)	58M	Undergraduates	Blood	4, 1	Stafford identical blocks test	0
					Minnesota Paper Forms Board	0
Moffat and Hampson (1996a)	40F, 40M	F:23.0 (4.09); M:21.8 (2.35)	Saliva	2, 1	MRT	F:0, M:NL
					Paper folding	0
Neave et al. (1999)	25F, 33M	F:28.75 (19–43); M:28.6 (18–51)	Saliva	1, 1	MRT	NL
Shute et al. (1983) 1	48F, 43M	24.5 (16–41)	Blood	1, 1	French Reference Kit for Cognitive Factors — spatial tests	F:0, M:NL
Shute et al. (1983) 2	12F, 12M	24.5 (16–41)	Blood	1, 1	Minnesota Paper Forms Board	0
					Primary mental abilities test/comprehensive ability battery space test	0
Silverman et al. (1999)	59M	22.42 (3.02)	Saliva	2, 2	MRT	+
Young et al. (2010)	26 young M, 62 old M	(25–35), (60–80)	Blood	1, 2	MRT	Young M:0, old M: +
					Figure discrimination task	0

Causal relationships are best tested by hormonal manipulation. Demonstrating that androgen treatment elicits a particular behavioral change, and that removal of treatment abolishes this effect, constitutes strong evidence for activational effects of the hormone. Several studies have reported activational effects of androgens on spatial task performance, but those that are placebo-controlled often fail to demonstrate significant effects (reviewed in Puts et al., 2007). Furthermore, these studies are frequently carried out using small and possibly unrepresentative clinical samples, such as hypogonadal males, Alzheimer patients, Turner Syndrome patients, and female-

to-male transsexuals (Aleman et al., 2004; Alexander et al., 1998; Cherrier et al., 2005; O'Connor et al., 2001; Ross et al., 2002; Van Goozen et al., 1995; van Goozen et al., 2002; Wolf et al., 2000).

Within-participants correlational studies offer an alternative means of assessing whether T has activational effects on spatial performance. These studies can be carried out on larger, non-clinical samples and have the power to show relationships between intra-individual changes in T and spatial performance. Within-participants designs produce lower error variance due to variability among individuals (i.e., differences among participants can be measured and separated from error); consequently, within-participants designs have greater statistical power and require fewer participants to show significant relationships. Correlations may not demonstrate causation, but if a causal relationship exists, then the putative dependent and independent variables should be correlated. Thus, if T has purely activational effects on spatial ability, then there should be correlations between intra-individual fluctuations in T levels and spatial performance in adults.

Using a between-participants design, Moffat and Hampson (1996a) found circadian changes in spatial performance that differed significantly by sex. Males tended to improve over the morning, whereas females exhibited the opposite trend. Because T levels also tend to decrease over the morning in both sexes, Moffat and Hampson suggested that the sex difference in performance change was the result of high morning T levels augmenting female spatial ability and impairing it in males. This hypothesis would be bolstered by evidence for within-participants correlations between changes in T levels and changes in spatial performance. However, the only previous study to examine these correlations (Silverman et al., 1999) found no significant relationship between changes in 59 men's T levels and changes in their 3D mental rotation performance over a 12-hour period.

In sum, the evidence is equivocal as to whether T has activational effects on spatial ability in humans, and whether differences in adult T levels might help explain some sex differences in spatial ability. Relatively small sample sizes may account for some of the highly varied results. We therefore sought to examine relationships in both sexes between salivary T levels and performance on a spatial task that shows one of the largest known human cognitive sex differences, the 3D Mental Rotation Task (Vandenberg and Kuse, 1978), in by far the largest sample of men and women yet recruited. If testosterone has activational effects on spatial ability, then:

1) Between-individual ("trait") differences in circulating T levels will predict differences in spatial performance.
2) Within-individual ("state") changes in T levels will predict changes in spatial performance.

Methods

Participants

Three hundred thirty-seven Michigan State University students participated in this human subject board-approved study. Participants were 160 women (20.43 yrs ± 1.53) and 177 men (20.14 yrs ± 1.71). Participants' reported ethnicities were 91.4% White, 3.6% Asian, 2.1% Hispanic or Latino, 1.2% Black or African American, 0.6% American Indian or Alaska Native, and 1.2% reported identifying with another ethnicity.

Scheduling of sessions

Participants were scheduled to participate in both a morning and an evening session, approximately one week apart (6.99 ± 0.72 days). Because the purpose of this study was in part to examine circadian changes in T levels, it was desirable to minimize menstrual cycle-related hormonal changes in women, which may affect spatial performance (Hampson, 1990a; Hampson, 1990b; Hausmann et al., 2000; Phillips and Silverman, 1997). Therefore, only women who reported currently taking hormonal contraception participated in the present study. We randomly allocated participants to attend their first session during the morning or the evening, with their second session taking place at the other time of day. Morning sessions began between 0820 h and 1000 h, and evening sessions began between 1720 h and 1900 h. An effort was made to maintain a consistent interval between morning and evening sessions of nine hours, so participants who were scheduled in the latter half of the morning testing session were also scheduled in the latter half of the evening session, and vice versa. The average time difference between the scheduled start times of the morning and evening sessions was 8.95 h (±0.55). Each session lasted approximately 1 h and included anthropometric and psychometric portions, separated by saliva collection. Anthropometric data were collected for use in other studies that are not reported here.

Saliva collection

To minimize contamination of saliva samples, participants were instructed not to eat, drink (except plain water), smoke, chew gum, or brush their teeth for 1 h before their scheduled session. Women wearing lipstick were asked to remove it with a tissue. Participants rinsed their mouths with water immediately before chewing a piece of sugar-free Trident gum (inert in salivary hormone assays) to stimulate saliva flow. Each participant then collected approximately 10 ml of saliva in a sodium azide-coated polystyrene tube, after which the tube was capped and stood upright at room temperature for 18–24 h to allow mucins to settle. Each tube was then frozen at −20 °C until hormone analysis.

Testosterone assays

We obtained salivary unbound ("free") testosterone concentrations, which correlate strongly with serum concentrations (e.g., Baxendale et al., 1980; Wang et al., 1981, $r = 0.81$ and 0.94, respectively). Testosterone radioimmunoassays (RIAs) were performed by an experienced RIA technician in the Salivary Radioimmunoassay Laboratory at the University of Western Ontario. Two hundred sixty-six female saliva samples (160 from Session 1, 106 from Session 2) and 333 male samples (177 from Session 1, 156 from Session 2) were analyzed.

Following a double ether extraction, all samples were assayed in duplicate using a Coat-A-Count kit for total testosterone (Diagnostic Products, Los Angeles, CA), modified for use with saliva (for details, see Moffat and Hampson, 1996b). RIAs were performed separately for men and women in two batches for each sex. Sensitivity was 5–10 pg/ml, and the average intra-assay coefficient of variation was 6.3%. Duplicate assay concentrations were highly correlated (morning: $r(297) = 0.97, p < 0.0001$; evening: $r(301) = 0.97, p < 0.0001$; controlling for sex, $r(294) = 0.93$, $p < 0.0001$ and $r(298) = 0.91$, $p < 0.0001$, for morning and evening, respectively). Consequently, each participant's duplicates were averaged for each session. If a value was below detectable levels for one duplicate, the duplicate assay with a detectable level was used without averaging. This was the case for one session for two female participants.

Psychometric testing

Fully-automated questionnaires and tasks were administered to participants via computer. These included the Vandenberg and Kuse (1978) 3D Mental Rotation Task (MRT) and several questionnaires and tasks administered for other studies. In the MRT, participants are shown 2-dimensional line drawings of 3-dimensional block figures. For each item, a target block figure is shown on the left, followed by

four similar figures on the right. The task is to select the two figures on the right that represent the target figure rotated in space. Digital images were courtesy of Michael J. Tarr, Brown University. For example stimuli, see Shepard and Metzler (1971). One point was assigned for each item only if both correct answers were selected, because this method of scoring has been shown to capture the largest sex difference (Voyer et al., 1995). The MRT consisted of two 4-minute sections of 10 items each.

Handedness was measured using a questionnaire developed by Peters (1998), which employs 5-point scales to assess the degree to which respondents prefer the left or right hand for 22 activities (e.g., draw, comb hair, wave goodbye). Scores for each respondent are summed across all 22 activities. Sexual orientation was assessed using the sexual attraction dimension of the Kinsey Scale (Kinsey et al., 1948), which employs a 7-point scale ranging from "I am attracted to men only, never to women" to "I am attracted to women only, never to men" to identify the statement that best describes the respondent's sexual feelings at present.

Female participants were also asked which brand of hormonal contraception they were currently using. One hundred fifty-one women answered this open-ended question, reporting the use of one of 30 brands. Nine women did not answer. The brands identified by female participants were then split into 10 categories according to the active chemicals (e.g., ethinyl estradiol and desogestrel vs. ethinyl estradiol and norethindrone), for the examination of any relationships between types of hormonal contraception and testosterone concentrations or spatial performance.

Data treatment

Testosterone concentrations exhibited leftward-skewed distributions. We corrected these asymmetries by log-transforming T concentrations before statistical analysis (Drennan, 1996).

Results

Validity and reliability measures

Testosterone

Mean salivary T concentrations for women were 20.35 ± 1.02 pg/ml (morning) and 14.76 ± 0.70 pg/ml (evening). For men, T concentrations were 109.17 ± 3.52 pg/ml (morning) and 76.21 ± 2.32 pg/ml (evening) (Table 2). These values are comparable to salivary T concentrations obtained in previous studies (e.g., Dabbs, 1990; Moffat and Hampson, 1996a, b). T was higher in men than in women (mixed model ANOVA with session order (morning or evening first) and sex as between-participants factors and time of day (morning or evening) as a repeated measure: main effect of sex: $F(1,257) = 1125.8$, $p < 0.0001$) and declined over the day in both sexes (main effect of time of day: $F(1,257) = 174.6.1$, $p < 0.0001$). Controlling for session order, T levels were also highly correlated within participants across sessions ($r(261) = 0.91$, $p < 0.0001$; controlling for sex, $r(258) = 0.59$, $p < 0.0001$). That is, participants who had high T in their morning session also tended to have high T in their evening session a week apart (Fig. 1), conforming to the notion of stable individual "trait" differences in circulating T (Dabbs, 1990).

Table 2
Salivary testosterone levels were higher in the morning than in the evening and higher in men than in women.

	Women		Men	
	AM ($N = 134$)	PM ($N = 131$)	AM ($N = 163$)	PM ($N = 170$)
Mean (SE, range) T levels (pg/ml)	20.35 (1.02, 4.5–67.5)	14.76 (0.70, 3.0–34.5)	109.17 (3.52, 28.0–301.5)	76.21 (2.32, 22.0–225.5)

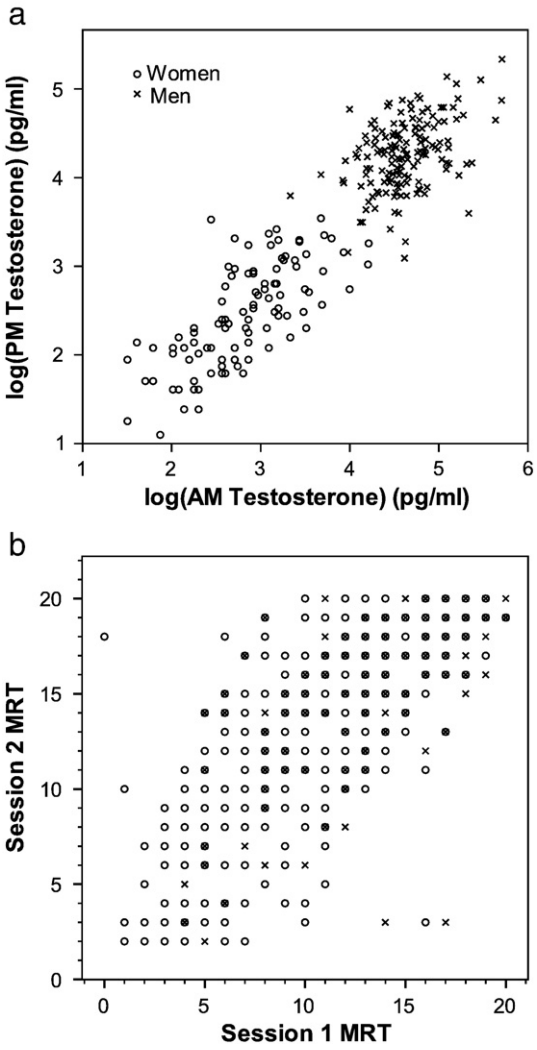

Fig. 1. Within-subject relationships between (a) morning and evening salivary testosterone levels and (b) Session 1 and Session 2 mental rotation test (MRT) performance in women and men. Testosterone levels correlated strongly across sessions, as did MRT scores.

Spatial ability

Session 1 mean MRT was 13.6 (±4.4) for men and 11.0 (±4.6) for women out of a possible score of 20. These scores are intermediate between the scores of 12.5 (men) and 10.3 (women) as reported by Burkitt et al. (2007) and those of 15.9 (men) and 11.7 (women) as reported by Burton et al. (2005), who gave 12 questions (rather than 10) per four-minute section. The decreased speed required to complete the 10 items per section in the present study should have reduced errors, raising scores, but the decreased number of possible correct answers should have had the opposite effect on scores. Thus, the MRT scores obtained in the present study are within the range expected from this participant population, given the design and scoring of the test.

Men's scores on the MRT were on average higher than women's (mixed model ANOVA with session order (morning or evening first) and sex as between-participants factors and testing session as a repeated measure: main effect of sex: $F(1,300) = 21.8$, $p < 0.0001$). MRT performance also improved from Session 1 to Session 2 (main effect of testing session: $F(1,300) = 103.1$, $p < 0.0001$; interaction with sex: $F(1,300) = 1.53$, $p = 0.22$). MRT scores did not differ between morning and evening sessions (main effect of session time (morning or evening): $F(1,301) = 0.57$, $p = 0.45$; interaction with sex: $F(1,301) = 1.74$, $p = 0.19$). Controlling for session order, MRT

scores were also highly correlated within participants across sessions ($r(301) = 0.75$, $p < 0.0001$; controlling for sex, $r(300) = 0.73$, $p < 0.0001$) (Fig. 1).

Hormonal contraception

The combination of active ingredients in the hormonal contraception used by the participants was unrelated to salivary T levels or spatial ability. There were no statistically significant effects of hormonal contraception active ingredients (10 categories) on morning or evening T levels, MRT performance in either session (one-way ANOVA with hormonal contraception type as predictor variable (10 levels) and T levels or MRT scores as outcome variables), or improvement across sessions in MRT (mixed model ANOVA with MRT scores as within-participants repeated measures outcome variable). Consequently, hormonal contraception type is not included in the following analyses.

Relationships between T and mental rotation ability

Relationships between T and mental rotation ability were tested using random-intercept multilevel models in SAS PROC MIXED (SAS Institute, 2004).

Testosterone as a time-varying covariate

A model with testing session and T as predictors yielded significant effects for testing session ($t(165) = 8.06$, $p < 0.0001$) and T ($t(165) = 3.32$, $p = 0.001$). However, because T is so closely related to sex (Fig. 1), and because sex is confounded by so many variables other than current T levels, we measured within-sex effects of T by inserting sex into this model. When sex and T are included in the same model, the effect of T can be interpreted as within-sex variation in T, and the effect of sex can be interpreted as a sex difference adjusted for the sex difference in T. The model with testing session, T, and sex as predictors yielded significant effects for testing session ($t(165) = 7.72$, $p < 0.0001$) and sex ($t(165) = 2.45$, $p = 0.02$). However, the effect for T was no longer significant ($t(165) = 0.41$, $p = 0.69$). When the T × sex interaction was added to this model, the interaction was nonsignificant ($t(164) = -0.13$, $p = 0.90$), and the effect of T remained nonsignificant ($t(164) = 0.25$, $p = 0.80$). Because these effects are in opposite directions, the within-sex effect of T was nonsignificant for both sexes (see also Fig. 2).

Within- and between-participants effects of testosterone

To test the hypothesis that between-participant differences in trait levels of T and within-participant T fluctuations have different effects on MRT performance, we partitioned T into a between-participants component (i.e., mean T across testing sessions) and a within-participants component (i.e., the difference between Session 2 T and Session 1 T). Both were treated as participant-level variables, meaning that they did not differ across testing session.

If between-participant differences in T affect mental rotation, then the between-participants variable should be significantly related to performance on the MRT. If within-participant variations in T affect spatial ability, then the within-participants variable should be significantly related to performance on the MRT, but only in the second testing session. Therefore, the within-participants component of T was allowed to interact with the testing session variable. If the within-participants variable is significantly related to performance on the MRT, but only in the second testing session, then the within-participants parameter should not reach statistical significance, but the interaction between the within-participants parameter and testing session should.

The model with sex, test session, the between-participants component of T, the within-participants component of T, and the interaction between within-participants T and testing session produced significant effects of sex ($t(165) = 2.58$, $p = 0.01$) and testing session ($t(165) = 7.66$, $p < 0.0001$). The effects of the between-participants component of T ($t(165) = -1.21$, $p = 0.23$), the within-participants component of T ($t(165) = -0.81$, $p = 0.40$), and the interaction between within-participants T and testing session ($t(165) = 0.45$, $p = 0.66$) were not statistically significant.

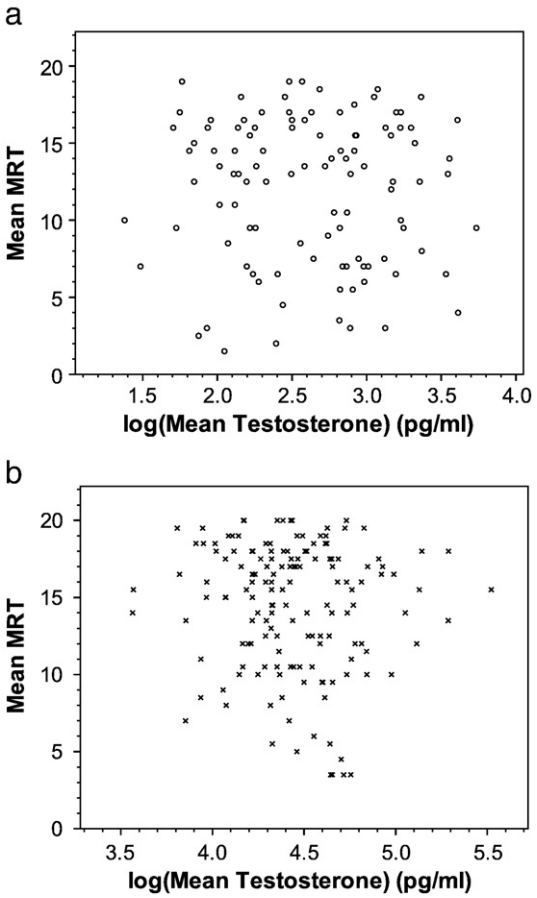

Fig. 2. Salivary testosterone levels and mental rotation test (MRT) performance scores were related in neither (a) women nor (b) men.

Testosterone and spatial learning

To test the hypothesis that higher levels of T during the first testing session led to an increased proficiency on the MRT in the second testing session, Session 1 T was transformed into an individual level variable. If higher levels of T during the first testing session affect subsequent, but not current, performance on the MRT, then Session 1 T should interact significantly and positively with testing session.

The model with Session 1 T, sex, testing session, and the interaction between Session 1 T and testing session yielded a significant effect for testing session ($t(193) = 8.70$, $p < 0.0001$) but no significant effects for sex ($t(193) = 1.74$, $p = 0.08$), Session 1 T ($t(193) = 1.20$, $p = 0.23$), or the interaction between Session 1 T and testing session ($t(193) = -1.94$, $p = 0.054$). While the interaction falls near the cutoff for statistical significance, it is in the opposite direction of what we would predict based on the hypothesis that higher levels of T during the first testing session led to an increased proficiency on the MRT in the second testing session.

To examine whether collinearity between testosterone and testing session made it impossible to simultaneously observe the effects of testosterone and session on MRT, we calculated a model with sex and testosterone as the only predictors of MRT performance. The effect of sex remained significant ($t(166) = 3.28$, $p = 0.001$), and the effect of testosterone remained nonsignificant ($t(166) = -0.53$, $p = 0.59$).

Other potential relationships between testosterone and mental rotation ability

It is possible that T has a more complex relationship with mental rotation ability than tested in the models above. Consequently, we tested 24 additional models, including additional scorings and scalings of the MRT, additional covariates, quadratic effects of T, and higher-order interactions. All of the models above containing sex as a predictor were rerun using an alternate scoring of the MRT (scoring correct answers without requiring both correct answers for each item, for a possible 40 points) as the dependent variable (see, e.g., Moffat and Hampson, 1996a). Secondly, all of the models were rerun using the squares of the first and second scorings of the MRT as dependent variables. Lastly, all of the models using both scorings and scalings of the MRT were rerun with T squared as an additional predictor, testing interactions between this variable and all variables with which T was tested to interact. In none of these models were T, its quadratic component, or any interactions containing T found to significantly predict MRT score, except for one parameter. After controlling for the quadratic component of Session 1 T and its interaction with session, a significant negative effect for the interaction between Session 1 T and session was observed ($t(191) = -2.08$, $p = 0.04$). However, this finding – which falls on the border of statistical significance – is only observed using this particular scoring and scaling of the MRT and after controlling for the quadratic component of Session 1 T and its interaction with session, and is inconsistent with the prediction that higher Session 1 T improves spatial learning.

Entering handedness and sexual orientation as control variables into analyses did not alter any of the above results.

Discussion

The present study is the first to examine within-participant diurnal changes in T and explore their possible relationship to spatial performance in women, and only the second to do so in men. This study also offers by far the largest sample to examine relationships between T concentrations and spatial ability within- or between-participants. T levels were unrelated to performance on the Vandenberg and Kuse 3D Mental Rotation Task, a male-biased spatial task showing one of the largest known human cognitive sex differences. Individual differences in T levels did not predict individual differences in mental rotation performance in either sex. Furthermore, the between-session improvement in performance was unrelated to T levels, and circadian changes in T were unrelated to changes in performance in either sex.

There are multiple reasons to believe that these findings are valid. First, T levels and mental rotation performance showed the normative sex differences (Dabbs, 1990; Vandenberg and Kuse, 1978; Voyer et al., 1995), and T levels exhibited expected diurnal changes and were highly correlated within participants across sessions (Dabbs, 1990). Second, the present study employed a far larger sample than those of previous studies, both in terms of numbers of participants and total T samples analyzed. Third, although power calculations are inappropriate in the interpretation of results (Hoenig and Heisey, 2001), test statistics for relationships between T levels and spatial performance were generally near zero and did not approach threshold for rejecting the null hypothesis. Finally, to maximize the chance of finding significant relationships, we did not correct for multiple statistical tests when exploring additional statistical models to our main analysis. Had we done so, relationships between T levels and mental rotation performance would have been even further from achieving statistical significance.

We attempted to minimize the influence of menstrual cycle phase on spatial performance and sex hormone levels by examining only female participants who were taking hormonal contraception. Although we view this as an advantage of our methods, this approach may be viewed as a limitation because it supports conclusions only about women taking hormonal contraception. Furthermore, although we found no effects of the type of hormonal contraception that our participants used on T levels or MRT performance, sex steroids may have fluctuated in women depending on whether they were in the "on" versus "off" phase of hormonal contraception use, introducing noise. However, such noise seems unlikely to have contributed significantly to our failure to find significant relationships between T levels and MRT performance, given that relationships between salivary T and mental rotation ability were found in neither sex. Moreover, although hormonal contraception and other factors such as diurnal fluctuations in cortisol and glucose levels could have introduced noise, reducing our ability to detect relationships between T and MRT performance, the influence of this noise should have been reduced by our large sample size and our use of both between- and within-participants designs.

We also analyzed possible confounding effects of handedness and sexual orientation and found no significant relationships between T and MRT performance or learning after controlling for these variables and their interactions. However, it remains possible that T could affect mental rotation in interaction with other hormones and non-hormonal factors that we did not examine.

It is also possible that a different male-biased spatial test from the Vandenberg and Kuse mental rotation test, or an alternative method of administering this test, would produce significant relationships between performance and salivary T concentrations. For example, perhaps the easier task implemented in the present study (10 questions and 4 min per section versus other methods, e.g., 12 questions and 3 min per section) might lead to scores that are less susceptible to changes in T levels. While we cannot rule out this possibility, the task used in the present study showed a large, highly robust sex difference, and the mental rotations test is the most frequently used test in studies of sex differences in spatial ability (e.g., Table 1 in Voyer et al., 1995). Scores on this test also correlate with performance on other male-biased spatial tasks, for example real world (Silverman et al., 2007) and virtual (Driscoll et al., 2005) navigation. Thus, this test constitutes one of the best candidates for showing a relationship between spatial performance and salivary T. Although we employed only one method of task administration (10 questions and 4 min per section), we examined two methods of scoring (see above) and found relationships between salivary T and performance using neither scoring method.

We collected two saliva samples per participant and one sample per session. Because T secretion is pulsatile, this method will have limited our ability to detect individual differences and within-individual diurnal variation in T production. However, morning and evening T levels were highly correlated across participants, even within sexes, suggesting that we captured a substantial proportion of the between-individual variation in T production. Furthermore, T levels showed a highly significant drop from morning to evening sessions, indicating that we were able to capture a substantial proportion of the diurnal variation in T production.

Our final caveats regard the representativeness of our sample. We examined only a restricted age range of young adults, and so it is possible that different results might have been obtained had we examined other ages (see below). In addition, there was notable attrition from the first to second sessions, especially in women (54 women versus 21 men, see Methods). This sex difference may be due to the greater difficulty reported by women in collecting the nearly 10 ml of saliva required. However, we can think of no reason why the relationship between T and spatial performance would differ between those who continued and those who dropped out of our study.

Overall, our results suggest that any within-individual effects of changing T levels would have to occur on a time scale longer than the diurnal intervals examined in this study and that of Silverman et al. (1999). Although Moffat and Hampson (1996a) found circadian changes in mental rotation performance – as did Sanders et al. (2002),

though only in men, and in only one of two samples – neither Silverman et al. (1999) nor the present study replicated these results. Furthermore, none of these studies found relationships between diurnal changes in T and changes in spatial performance.

Spatial ability could change as a result of longer-term variations in T levels, such as those due to menstrual cycle phase or season. Hausmann et al. (2000) found a relationship between mental rotation performance and the T levels of 12 women sampled at several times in their cycle. Kimura and Hampson (1994) found seasonal changes in spatial performance in young men, which the authors suggested may have resulted from fluctuating T levels. These changes may be analogous to those observed in meadow voles (*Microtus pennsylvanicus*) and deer mice (*Peromyscus maniculatus*), in which males expand their home ranges during the breeding season in order to increase access to mates (Galea et al., 1996, 1995; Gaulin and FitzGerald, 1989). In both of these species, males outperform females in laboratory spatial tasks only during the breeding season (Galea et al., 1996; Gaulin and FitzGerald, 1989) when T levels are elevated (Galea and McEwen, 1999).

However, other cyclically fluctuating factors could cause menstrual cycle and seasonal variation in human spatial ability, and the aggregate of previous findings suggests a weak, if any, causal relationship between levels of circulating T and spatial ability in adults. For example, Halari et al. (2005) found no between-individual relationships between T and mental rotation performance, and while Moffat and Hampson (1996a) found a significant negative linear relationship in men, Silverman et al. (1999) found a significant positive relationship. Driscoll et al. (2005) also found a positive correlation between salivary T and virtual water maze performance in men but not in women. This relationship was found only with both young and elderly men included in the analysis (not in either group separately), so it is possible that declining T levels were tracking the decline of some other function or functions associated with spatial cognition. This interpretation is supported by the finding of Young et al. (2010) that mental rotation performance correlated with T levels in older men but not in younger men. Moreover, in the present study, by far the largest of its kind, spatial performance was not associated with T levels in between-participants comparisons. It is possible that previously reported associations between spatial performance and adult T concentrations reflect prior organizational effects of T and a tendency for adult levels to correlate with earlier ones.

Further, we suggest that testosterone treatment studies generally find negligible effects on spatial ability. Although some studies have claimed to show effects of T treatment on spatial performance (e.g., Van Goozen et al., 1994), those that used placebo-treated controls often find no improvement beyond normal learning effects. For example, Alexander et al. (1998) found no effect of six weeks of T treatment on spatial performance in 33 hypogonadal and 10 eugonadal men, ages 21–59. Young et al. (2010) found no effect of six weeks of T treatment on spatial performance in 32 older and 13 younger men. O'Connor et al. (2001) found no effect of eight weeks of T treatment on spatial performance in seven hypogonadal men (ages 23–40), although performance in 15 eugonadal men failed to exhibit normal practice effects after T treatment. Similarly, Aleman et al. (2004) found improvement (likely due to practice) in 12 women treated with placebo on their first testing session and T on their second, but no improvement in 14 women treated with T then placebo. Van Goozen et al. (2002) found no effects of T treatment in 19 female-to-male transsexuals on several male-biased spatial tasks, including mental rotation. Likewise, Ross et al. (2003) observed no improvement in spatial abilities in 26 androgen-treated Turner Syndrome (TS) patients relative to placebo-treated TS controls. Wolf et al. (2000) found no effect on spatial ability of a single T injection relative to placebo in 30 elderly men. Although several studies have found cognitive improvements following T treatment in elderly men (Cherrier et al., 2001; Janowsky et al., 2000, 1994), including Alzheimer patients (Cherrier et al., 2005), these effects were not restricted to spatial tasks or to tasks showing a male advantage and seemed only to restore practice effects present in younger men. Thus, these treatment effects do not provide evidence that circulating T contributes to spatial cognitive sex differences in younger adults.

Finally, putative activational effects of testosterone on spatial ability may be questioned on theoretical grounds (Puts et al., 2007). Spatial behaviors and their underlying neural systems should remain susceptible to hormonal fluctuations only if the cost of maintaining such plasticity was counterbalanced by fitness benefits over the evolutionary history of the organism. This is likely to occur if spatial demands change significantly and repeatedly (e.g., seasonally). In some rodent species, males expand their home ranges during the breeding season in order to increase access to mates, and males outperform females in laboratory spatial tasks only during the breeding season when T levels are elevated (see above). In relatively non-seasonal species, such as laboratory rats, spatial ability appears to be comparatively unresponsive to T after certain early critical periods (Commins, 1932). The aggregate of human research seems to suggest that, at least within the normal range of circulating levels in young adults, T has little activational effect on performance on male-biased spatial tasks. Given that humans exhibit very low breeding seasonality, these findings might be expected.

If, as our data indicate, circulating T has little effect on spatial cognitive performance in young men and women, then the well established sex differences in such performance cannot be attributed to the sex difference in circulating T. By elimination, this would suggest that any role for androgens in establishing sex differences in spatial ability must occur earlier in life, either perinatally or pubertally, when reliable sex differences in androgen secretion could have lasting, organizational effects.

Acknowledgments

We thank Bradly Alicea, Michael Burla, Melina Durhal, Rebecca Frysinger, Sana Khan, Mallory Leinenger, Erin MacCourtney, Heather Malinowski, Ernestine Mitchell, Joe Morehouse, Sara Sutherland, Lisa Vroman, Tyesha Washington, and Molly Zolianbawi for their conscientious assistance in study preparation and data collection, and Elizabeth Hampson and Bavani Rajakumar for their assistance with hormone assays.

References

Aleman, A., Bronk, E., Kessels, R.P., Koppeschaar, H.P., 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. II. Cognitive abilities. Horm. Behav. 33, 85–94.

Arnold, A.P., Breedlove, S.M., 1985. Organizational and activational effects of sex steroids on brain and behavior: a reanalysis. Horm. Behav. 19, 469–498.

Baxendale, P.M., Reed, M.J., James, V.H.T., 1980. Testosterone in saliva of normal men and its relationship with unbound and total testosterone levels in plasma. J. Endocrinol. 187, 46–47.

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.

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. Dev. Neuropsychol. 28, 493–505.

Cherrier, M.M., Asthana, S., Plymate, S., Baker, L., Matsumoto, A.M., Peskind, E., Raskind, M.A., Brodkin, K., Bremner, W., Petrova, A., La Tendresse, S., Craft, S., 2001. Testosterone supplementation improves spatial and verbal memory in healthy older men. Neurology 57, 80–88.

Cherrier, M.M., Matsumoto, A.M., Amory, J.K., Asthana, S., Bremner, W., Peskind, E.R., Raskind, M.A., Craft, S., 2005. Testosterone improves spatial memory in men with Alzheimer disease and mild cognitive impairment. Neurology 64, 2063–2068.

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

Christiansen, K., Knussmann, R., 1987. Sex hormones and cognitive functioning in men. Neuropsychobiology 18, 27–36.
Collaer, M.L., Brook, C.G., Conway, G.S., Hindmarsh, P.C., Hines, M., 2009. Motor development in individuals with congenital adrenal hyperplasia: strength, targeting, and fine motor skill. Psychoneuroendocrinology 34, 249–258.
Commins, W.D., 1932. The effect of castration at various ages upon learning ability of male albino rats. J. Comp. Psychol. 14, 29–53.
Dabbs Jr., J.M., 1990. Salivary testosterone measurements: reliability across hours, days, and weeks. Physiol. Behav. 48, 83–86.
Drennan, R.D., 1996. Statistics for Archeologists: A Commonsense Approach. Plenum, New York.
Driscoll, I., Hamilton, D.A., Yeo, R.A., Brooks, W.M., Sutherland, R.J., 2005. Virtual navigation in humans: the impact of age, sex, and hormones on place learning. Horm. Behav. 47, 326–335.
Falter, C.M., Arroyo, M., Davis, G.J., 2006. Testosterone: activation or organization of spatial cognition? Biol. Psychol. 73, 132–140.
Galea, L.A., McEwen, B.S., 1999. Sex and seasonal differences in the rate of cell proliferation in the dentate gyrus of adult wild meadow voles. Neuroscience 89, 955–964.
Galea, L.A., Kavaliers, M., Ossenkopp, K.P., Hampson, E., 1995. Gonadal hormone levels and spatial learning performance in the Morris water maze in male and female meadow voles, *Microtus pennsylvanicus*. Horm. Behav. 29, 106–125.
Galea, L.A., Kavaliers, M., Ossenkopp, K.P., 1996. Sexually dimorphic spatial learning in meadow voles *Microtus pennsylvanicus* and deer mice *Peromyscus maniculatus*. J. Exp. Biol. 199, 195–200.
Gaulin, S.J.C., FitzGerald, R.W., 1989. Sexual selection for spatial-learning ability. Anim. Behav. 37, 322–331.
Gordon, H.W., Lee, P.A., 1986. A relationship between gonadotropins and visuospatial function. Neuropsychologia 24, 563–576.
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.
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.
Hassler, M., Gupta, D., Wollmann, H., 1992. Testosterone, estradiol, ACTH and musical, spatial and verbal performance. Int. J. Neurosci. 65, 45–60.
Hausmann, M., Slabbekoorn, D., Van Goozen, S.H., Cohen-Kettenis, P.T., Gunturkun, O., 2000. Sex hormones affect spatial abilities during the menstrual cycle. Behav. Neurosci. 114, 1245–1250.
Hausmann, M., Schoofs, D., Rosenthal, H.E., Jordan, K., 2009. Interactive effects of sex hormones and gender stereotypes on cognitive sex differences—a psychobiosocial approach. Psychoneuroendocrinology 34, 389–401.
Hier, D.B., Crowley Jr., W.F., 1982. Spatial ability in androgen-deficient men. N Engl J. Med. 306, 1202–1205.
Hines, M., 2004. Brain Gender. Oxford University Press, New York.
Hoenig, J.M., Heisey, D.M., 2001. The abuse of power: the pervasive fallacy of power calculations for data analysis. Am. Stat. 55, 19–24.
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.
Janowski, J.S., Chavez, B., Zamboni, B.D., 1998. The cognitive neuropsychology of sex hormones in men and women. Dev. Neuropsychol. 14, 421–440.
Janowsky, J.S., Oviatt, S.K., Orwoll, E.S., 1994. Testosterone influences spatial cognition in older men. Behav. Neurosci. 108, 325–332.
Janowsky, J.S., Chavez, B., Orwoll, E., 2000. Sex steroids modify working memory. J. Cogn. Neurosci. 12, 407–414.
Kampen, D.L., Sherwin, B.B., 1996. Estradiol is related to visual memory in healthy young men. Behav. Neurosci. 110, 613–617.
Kimura, D., Hampson, E., 1994. Cognitive pattern in men and women is influenced by fluctuations in sex hormones. Curr. Dir. Psychol. Sci. 3, 57–61.
Kinsey, A.C., Pomeroy, W.B., Martin, C.E., 1948. Sexual Behavior in the Human Male. W. B. Saunders, Philadelphia.
Klaiber, E.L., Broverman, D.M., Kobayashi, Y., 1967. The automatization cognitive style, androgens, and monoamine oxidase. Psychopharmacologia. 11, 320–336.
Maccoby, E.E., Jacklin, C.N., 1974. The Psychology of Sex Differences. Stanford University Press, Stanford, CA.
Matousek, R.H., Sherwin, B.B., 2010. Sex steroid hormones and cognitive functioning in healthy, older men. Horm. Behav.
McKeever, W.F., Deyo, R.A., 1990. Testosterone, dihydrotestosterone, and spatial task performance of males. Bull. Psychonometric Soc. 28, 305–308.
McKeever, W.F., Rich, D.A., Deyo, R.A., Conner, R.I., 1987. Androgens and spatial ability: failure to find a relationship between testosterone and ability measures. Bull. Psychonomic Soc. 25, 438–440.
Meikle, A.W., Stringham, J.D., Bishop, D.T., West, D.W., 1988. Quantitating genetic and nongenetic factors influencing androgen production and clearance rates in men. J. Clin. Endocrinol. Metab. 67, 104–109.
Moffat, S.D., Hampson, E., 1996a. A curvilinear relationship between testosterone and spatial cognition in humans: possible influence of hand preference. Psychoneuroendocrinology 21, 323–337.
Moffat, S.D., Hampson, E., 1996b. Salivary testosterone levels in left- and right-handed adults. Neuropsychologia 34, 225–233.
Morris, J.A., Jordan, C.L., Breedlove, S.M., 2004. Sexual differentiation of the vertebrate nervous system. Nat. Neurosci. 7, 1034–1039.
Neave, N., Menaged, M., Weightman, D.R., 1999. Sex differences in cognition: the role of testosterone and sexual orientation. Brain Cogn. 41, 245–262.
O'Connor, D.B., Archer, J., Hair, W.M., Wu, F.C., 2001. Activational effects of testosterone on cognitive function in men. Neuropsychologia 39, 1385–1394.
Peters, M., 1998. Description and validation of a flexible and broadly usable handedness questionnaire. Laterality Asymmetries Body Brain Cogn. 3, 77–96.
Phillips, K., Silverman, I., 1997. Differences in the relationship of menstrual cycle phase to spatial performance on two- and three-dimensional tasks. Horm. Behav. 32, 167–175.
Phoenix, C.H., Goy, R.W., Gerall, A.A., Young, W.C., 1959. Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. Endocrinology 65, 369–382.
Puts, D.A., Gaulin, S.J., Breedlove, S.M., et al., 2007. Sex differences in spatial ability: evolution, hormones and the brain. In: Platek, S.M. (Ed.), Evolutionary Cognitive Neuroscience. MIT Press, Cambridge, Massachusetts, pp. 329–379.
Puts, D.A., McDaniel, M.A., Jordan, C.L., Breedlove, S.M., 2008. Spatial ability and prenatal androgens: meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies. Arch. Sex. Behav. 37, 100–111.
Ross, J.L., Stefanatos, G.A., Kushner, H., Zinn, A., Bondy, C., Roeltgen, D., 2002. Persistent cognitive deficits in adult women with Turner syndrome. Neurology 58, 218–225.
Ross, J.L., Roeltgen, D., Stefanatos, G.A., Feuillan, P., Kushner, H., Bondy, C., Cutler Jr., G.B., 2003. Androgen-responsive aspects of cognition in girls with Turner syndrome. J. Clin. Endocrinol. Metab. 88, 292–296.
Sanders, G., Sjodin, M., de Chastelaine, M., 2002. On the elusive nature of sex differences in cognition: hormonal influences contributing to within-sex variation. Arch. Sex. Behav. 31, 145–152.
SAS, I., Statistical analysis system 7.5. Author, Cary, NC, 2004.
Shepard, R.N., Metzler, J., 1971. Mental rotation of three dimensional objects. Science 171, 701–703.
Shute, V.J., Pellegrino, J.W., Hubert, L., Reynolds, R.W., 1983. The relationship between androgen levels and human spatial abilities. Bull. Psychonomic Soc. 21, 465–468.
Silverman, I., Kastuk, D., Choi, J., Phillips, K., 1999. Testosterone levels and spatial ability in men. Psychoneuroendocrinology 24, 813–822.
Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., Olshansky, E., 2007. Evolved mechanisms underlying wayfinding: further studies on the hunter–gatherer theory of spatial sex differences. Evol. Hum. Behav. 21, 201–213.
Van Goozen, S.H., Cohen-Kettenis, P.T., Gooren, L.J., 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.
Van Goozen, S.H., Cohen-Kettenis, P.T., Gooren, L.J., Frijda, N.H., Van de Poll, N.E., 1995. Gender differences in behaviour: activating effects of cross-sex hormones. Psychoneuroendocrinology 20, 343–363.
van Goozen, S.H., Slabbekoorn, D., Gooren, L.J., Sanders, G., Cohen-Kettenis, P.T., 2002. Organizing and activating effects of sex hormones in homosexual transsexuals. Behav. Neurosci. 116, 982–988.
Vandenberg, S.G., Kuse, A.R., 1978. Mental rotations: a group test of three-dimensional spatial visualization. Percept. Mot. Skills 47, 599–604.
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.
Wang, C., Plymate, S., Nieschlag, E., Paulsen, C.A., 1981. Salivary testosterone in men: further evidence of a direct correlation with free serum testosterone. J. Clin. Endocrinol. Metab. 53, 1021–1024.
Wolf, O.T., Preut, R., Hellhammer, D.H., Kudielka, B.M., Schurmeyer, T.H., Kirschbaum, C., 2000. Testosterone and cognition in elderly men: a single testosterone injection blocks the practice effect in verbal fluency, but has no effect on spatial or verbal memory. Biol. Psychiatry 47, 650–654.
Young, L.A., Neiss, M.B., Samuels, M.H., Roselli, C.E., Janowsky, J.S., 2010. Cognition is not modified by large but temporary changes in sex hormones in men. J. Clin. Endocrinol. Metab. 95, 280–288.