ELSEVIER

Neuropsychologia 42 (2004) 782–790

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

The relationship of male testosterone to components of mental rotation

Carole K. Hooven[a,*], Christopher F. Chabris[b], Peter T. Ellison[a], Stephen M. Kosslyn[b]

[a] *Department of Anthropology, Harvard University, Cambridge, MA 02138, USA*
[b] *Department of Psychology, Harvard University, Cambridge, MA 02138, USA*

Received 13 February 2003; received in revised form 11 June 2003; accepted 10 November 2003

Abstract

Studies suggest that higher levels of testosterone (T) in males contribute to their advantage over females in tests of spatial ability. However, the mechanisms that underlie the effects of T on spatial ability are not understood. We investigated the relationship of salivary T in men to performance on a computerized version of the mental rotation task (MRT) developed by [Science 171 (3972) (1971) 701]. We studied whether T is associated specifically with the ability to mentally rotate objects or with other aspects of the task. We collected hormonal and cognitive data from 27 college-age men on 2 days of testing. Subjects evaluated whether two block objects presented at different orientations were the same or different. We recorded each subject's mean response time (RT) and error rate (ER) and computed the slopes and intercepts of the functions relating performance to angular disparity. T level was negatively correlated with ER and RT; these effects arose from correlations with the intercepts but not the slopes of the rotation functions. These results suggest that T may facilitate male performance on MRTs by affecting cognitive processes unrelated to changing the orientation of imagined objects; including encoding stimuli, initiating the transformation processes, making a comparison and decision, or producing a response.

Keywords: Spatial ability; Testosterone; Mental rotation; Slope; Intercept; Response time

1. Introduction

On average, human males outperform females in tests of spatial ability (Voyer, Voyer, & Bryden, 1995). A wealth of data from human and animal studies suggests that the relatively high concentration of testosterone (T) in males plays a critical role in their superior performance (Liben et al., 2002). However, the mechanisms through which T may affect spatial ability are not understood, and studies suggest that high T is not always associated with better performance among men (Moffat & Hampson, 1996). Because performing a spatial task, like any mental task, involves a series of distinct cognitive and motor processes (Sternberg, 1969), if T does modulate performance, it may do so through its relationship to any one or more of these processes. Relating T levels to relatively coarse measures of performance may not accurately reflect the relationship of T to the abilities of interest. To our knowledge, no published studies have investigated the relationship of T levels to the processes that actually transform objects in mental images per se, as distinct from other aspects of the task—such as the processes that encode the stimuli, initiate the transformation processes, make a comparison, or produce a response. To understand how T is related to variations in cognitive ability, we must analyze aspects of task performance that reflect distinct processes.

Organizational and activational effects of T influence spatial ability. In mammals, organizational effects of T occur primarily during a critical period in pre- and early post-natal development during which sexual differentiation occurs. In developing fetuses, higher levels of T and its metabolites (primarily DHT and estradiol) not only promote the development of male sexual organs, they also lead to the "masculinization" of the brain, resulting in the development of sexually-dimorphic brain structures. Studies with non-human mammals have shown that these brain structures later play a key role in the expression of male-typical behaviors, including enhanced spatial ability (Isgor & Sengelaub, 1998; Sherry, Jacobs, & Gaulin, 1992). Activational effects normally occur during and after adolescence in response to the action of circulating T. In adult males, higher T levels (typically three to ten times higher in human males than in females (Yen, Jaffe, & Barbieri, 1999)) modulate gene expression in specific brain regions (some of which have sexually differentiated patterns of androgen receptor

---

* Corresponding author.
E-mail address: hooven@fas.harvard.edu (C.K. Hooven).

0028-3932/$ – see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2003.11.012

concentration (Kruijver, Fernandez-Guasti, Fodor, Kraan, & Swaab, 2001)) to facilitate the expression of male-typed behaviors and cognitive patterns (Williams & Meck, 1991). Research on humans of both sexes who have experienced atypical levels of androgens during the organizational stage suggests that male-typical T levels lead to superior spatial ability in adulthood (Hampson, Rovet, & Altmann, 1998; Hier & Crowley, 1982).

Research on the activational effects of T on spatial ability in humans has focused on relationships between current T levels and performance on spatial tasks, or how performance varies with changes in T levels. Many researchers have reported that T level is correlated with performance on spatial tasks (Liben et al., 2002); but more than that, studies also suggest that changes in T level in adulthood cause differences in spatial abilities. Results indicate that male-typical T levels in adulthood lead to superior performance on spatial tests, but do not improve performance on non-spatial tasks, such as those measuring verbal ability (e.g., Janowsky, Oviatt, & Orwoll, 1994; Slabbekoorn, van Goozen, Megens, Gooren, & Cohen-Kettenis, 1999; Van Goozen, Cohen-Kettenis, Gooren, & Frijda, 1994).

Although studies have consistently found that T levels within the normal adult-male range are accompanied by a sex-based advantage on spatial tasks, the literature on the relationship between current T level and performance on spatial tasks within males is less consistent. Some studies report negative relationships (e.g., Gouchie & Kimura, 1991; Moffat & Hampson, 1996), some report positive relationships (e.g., Christiansen & Knussmann, 1987; Silverman, Kastuk, Choi, & Phillips, 1999), and others have found no relationship (e.g., Alexander et al., 1998; McKeever, Rich, Deyo, & Conner, 1987). The inconsistent results might be explained by differences in any of the following factors: methods of measuring T levels (i.e., time of day when the sample is taken, assay methods, sampling serum vs. saliva), subject samples, and measures of spatial ability (Silverman et al., 1999). In this article, we focus on the relationship of salivary T to response time (RT), and error rate (ER), associated with two classes of processes on a standard test of spatial abilities, namely, mental rotation.

Tests of spatial ability have been categorized into three types, each measuring a distinct aspect of this ability. Specifically, spatial perception tests assess the ability to determine spatial relations, such as in the Rod and Frame test (Witkin & Asch, 1948); spatial visualization tests assess the processing of complex spatial information, such as in the Embedded Figures Test (Witkin, 1950) in which subjects must remember geometric forms and then pick them out from more complex forms; and mental rotation tasks (MRTs) assess the ability to rotate mental images of objects. MRTs consistently yield the largest effect sizes, of any cognitive or spatial test specifically, for sex differences in performance. Of the MRTs, the effect sizes (expressed as the number of standard deviations by which male performance is greater than female performance) are highest for the Vandenberg and Kuse MRT (herein referred to as "VK") (Vandenberg & Kuse, 1978), and range from 0.7 (Voyer et al., 1995) to 0.9 (Linn & Petersen, 1985). The magnitude of this sex difference has remained constant over time (Masters & Sanders, 1993), and is evident cross-culturally (Halpern & Tan, 2001; Oosthuizen, 1991).

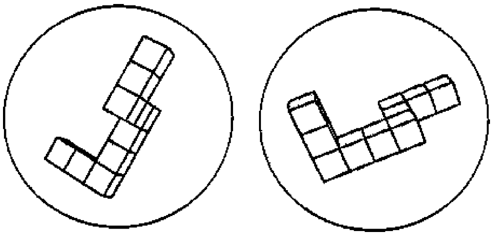

Fig. 1. Sample trial from Shepard and Metzler MRT 40° rotation—Different objects.

The VK is an adaptation of a task developed by Shepard and Metzler in 1971 (Shepard & Metzler, 1971), which was used to demonstrate that internal mental representations share spatial properties with the external objects they depict. On each trial of the Shepard and Metzler MRT (SM), subjects view a pair of two-dimensional projections of three-dimensional block objects, and the members of each pair usually are at different orientations. An example trial is presented in Fig. 1. Half the time the two objects have identical shapes, and half the time they are mirror images. Subjects must decide, as quickly and accurately as possible, whether the members of each pair are the same or different objects. Shepard and Metzler observed that in trials in which the two objects were the same, RT was a strong linear function of the degree of angular disparity between the objects. They inferred that to compare the objects in each pair and make a decision about similarity, the subjects "mentally rotated" one object into congruence with the other, so that the imagined object followed a trajectory analogous to that of a physical object rotated manually. The finding of a linear relationship between angle and RT is robust (e.g., Shepard & Judd, 1976; Wexler, Kosslyn, & Berthoz, 1998).

In 1978, Vandenberg and Kuse (1978) created a timed, paper-and-pencil MRT, which incorporated the block figures from the SM. Their version of the task includes 20 trials, set up as shown in Fig. 2. Subjects choose which two of the four target objects they believe are identical to the standard, and have six min to complete as many of the trials as possible. Points are awarded for each correct choice.

The VK represented an improvement over the SM in that it can be easily administered to large groups. Most studies on the relationship between T level and mental rotation ability use the VK, and furthermore, the bulk of the data on sex differences in mental rotation ability is derived from this test. But even on this specific test, the literature on the relationship between performance and salivary T yields contradictory results. Silverman et al. (1999) found that more T was associated with better performance on the VK

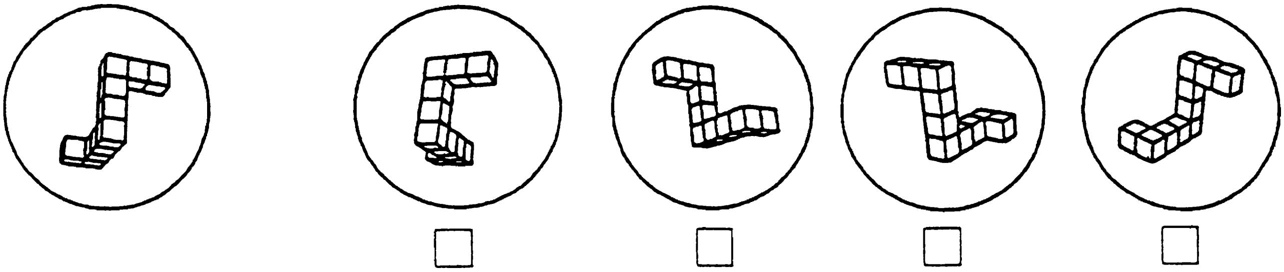

Fig. 2. Sample trial from Vandenberg and Kuse MRT (Vandenberg & Kuse, 1978).

in a sample of 59 male undergraduates. They measured salivary T and performance on the VK for each subject at two times of the day, once when T levels were expected to be relatively high and once when they were expected to be relatively low because of diurnal decline. Accuracy on the VK was greater when mean T levels were greater, but there was no effect of T on performance on an anagrams task or a digit symbol test. In contrast, Moffat and Hampson (1996) found a negative correlation between salivary T and performance on the VK, and no relationship between T and two control tests of verbal skills, in a sample of 19 right-handed male undergraduates. Subjects were tested in two groups, at early and late times of the morning, when T levels were expected to be relatively high and relatively low, respectively. The second group had significantly lower T levels, and they performed better than the group tested earlier. This negative relationship between performance on the VK and salivary T is consistent with the findings of another study, reported by Neave, Menaged, and Weightman (1999). These researchers found that subjects who had more salivary T performed the VK more poorly than those with lower levels (they tested 34 males, 17 of whom were homosexual). Yet other researchers have found no significant relationship between performance on the VK and salivary T levels among males. For example, Gouchie and Kimura (1991) tested 42 right-handed undergraduates of both sexes, and for each sex, grouped subjects according to whether they had high or low T levels. Among the four groups (high- and low-T males, high- and low-T females), the only difference in performance was between low-T men and low-T women: there were no differences in performance between high- and low-T males.

There are no obvious systematic differences among the studies of T and the VK that explain their inconsistent results. However, one possible explanation lies in the fact that the VK is a relatively insensitive measure of cognitive processing—it generates a single score (although accuracy can be calculated from each half of the test) from performance on a complex test that involves several distinct cognitive processes. To complete each trial, subjects must (a) select two objects to compare, moving attention to the appropriate “standard” and “target” objects; (b) form a mental representation of the object to be rotated; (c) rotate the object until its orientation is the same as the standard; (d) compare the two objects; (e) decide whether the objects are the same or different; (f) produce a response (Karadi, Kallai, & Kovacs, 2001). The VK produces a composite score that reflects the operation of all of these processes. Thus, the relationship of T to performance on the VK may not accurately reflect the relationship of T to mental rotation ability per se.

In the experiment reported here, we isolate measures of mental rotation ability from other abilities involved in the task, so that we may begin to clarify the relationship of T to mental rotation ability specifically. We administered the original SM, and then examined separately the slope and intercept of the function relating performance to angular disparity, for both ER and RT. The slope of this function indexes the rotation process (step “c” above) itself, and the intercept indexes the contribution of all other processes used to perform the task. Thus, this paradigm allows us to investigate the relationship between T and mental rotation ability with greater precision than in previous studies, by assessing whether T is associated with mental rotation ability per se, or other processes typically used in MRTs.

2. Method

2.1. Participants

Twenty-eight heterosexual males volunteered to take part in two sessions, 1 week apart, for which they were paid US $15. Subjects were primarily undergraduates at Harvard University. Mean age was 23 (S.D. = 4, range 18–33). All subjects reported no drug use and no history of psychiatric illness.

2.2. Materials

2.2.1. Mental rotation task

This MRT was administered using a computerized adaptation of the three-dimensional MRT described by Shepard and Metzler (1971). Stimuli were delivered and responses were recorded by a Macintosh computer running OS 9 (Apple Computer, Cupertino, CA) with a 16 in. monitor. Error rate (the percent of trials in which the subject answered incorrectly) and RT (the number of milliseconds between stimulus onset and the subjects' keypress response) were automatically recorded by PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). We prepared 16 practice trials and 80 trials from which data were collected. Each trial consisted of pairs of circles, 3 in. in diameter, presented side-by-side. We presented a subset of the stimuli used in the original Shepard and Metzler (1971) study. As illustrated in Fig. 1, each circle (3.25 in. in diameter or 10.8° of visual angle) contained one two-dimensional representation of a three-dimensional block stimulus (approximately 2.25 in. × 1.25 in., or 7.5° × 4.2° of visual angle). An equal number of objects in each pair were presented at angles that differed by 0, 40, 80, 120, or 180°. In addition, half of the stimuli at each angle were Same and half were Different. Accordingly, there were 5 angles × 2 response types × 8 standard objects = 80 total trials.

2.3. Sample collection and hormonal measurements

Subjects provided one saliva sample at each testing session, for later analysis, following an established protocol (Lipson & Ellison, 1996). Salivary T levels are considered to reflect unbound testosterone, which correlates highly with free T levels in serum. Free T is a measure of the amount of T that is capable of exerting a biological effect[1] (Granger, Schwartz, Booth, & Arentz, 1999).

We asked subjects not to eat, brush their teeth, or smoke for 1 h before coming to the laboratory. Investigators gave detailed written instructions to subjects upon arrival. They directed subjects to chew a sugarless gum (Carefree spearmint) to stimulate saliva production, and then to salivate into a 15 ml tube. Subjects began the MRT approximately 15 min later, after watching a 12 min video related to a separate part of the experiment.[2] Samples were kept in the original collection vials, which had been treated with sodium aizide to prevent destabilization of the steroid molecules. The samples were then retained at room temperature and shaded from light, until they could be frozen. Samples were thawed 24 h prior to conducting the assay. Samples were assayed in duplicate in a tritium-based radioimmunoassay (Lipson & Ellison, 1989).

2.4. Procedure

Subjects were tested individually by a female investigator.[3] In an attempt to control for differences in T levels due to the diurnal decline of T, subjects were tested in two groups: one group was tested at 10:00 am, and the other group at 1:15 p.m. We scheduled sessions for each subject on the same day of the week and time day, with appointments one week apart. At each session, subjects sat at a desk on which the monitor and keyboard were placed, completed a consent form, read the detailed instructions for the study, provided a saliva sample, and watched a 12 min video. The investigator then instructed the subject that he would be completing a short computer-based cognitive task. Specifically, investigators told subjects that they should decide whether each of the two objects in a pair had the Same shape or Different shapes, and to indicate their choice by pressing a key on the keyboard (to remember the appropriate keys, they were told to press "B" for "both the same" and "N" for "not the same"). Five hundred milliseconds after subjects pressed "B" or "N," the next stimulus pair appeared on the screen. The trials were presented in a random sequence in each session. On each trial, the computer timed the interval between stimulus presentation and the subjects' pressing a key; it also recorded which key was pressed. Subjects were told that they should "strive for speed and accuracy" in responding. Subjects completed the 16 practice trials (with different stimuli from the ones presented in the experimental trials), and questions about the procedure were answered. The investigator then left the room, and the subjects completed the 80 trials for which data were collected.

2.5. Data preparation

Consistent with T value distributions in other populations, the T value distribution in our sample was positively skewed. To normalize this distribution, values were log transformed (base 10). Men tested at 11:00 had lower ERs than men tested at 1:15. Because T levels decline throughout the day, this is a pattern we would expect if T facilitates performance on this task. This difference was significant on Day 1 only (Day 1: $t(25) = 2.13$, $P = 0.04$; Day 2 $t(25) = 1.68$, $P = 0.10$). All values were transformed to z-scores separately within each of the two groups, defined by the time of testing. T levels, ER, and RT are reported for Day 1 and Day 2. Mean RTs were computed after eliminating trials on which the subject answered incorrectly and after eliminating "outlier" trials: an outlier was defined as a time that was greater than 2.5 times the mean of the other trials in that cell of the design (i.e., a particular combination of rotation angle and response

---

[1] Testosterone circulates in the blood bound to specific (sex-hormone binding globulin) and non-specific (albumin) binding proteins. Only a small fraction of total circulating T is unbound and available to interact with intracellular androgen receptors. Methods for quantifying T levels variously measure different components of the total circulating T complement: total T; free or unbound T; and so-called "bioavailable" T, which is composed of free plus albumin-bound T. Confusion can result when comparisons are made among different T measures, or when incorrect inferences are made from any one method.

[2] To test a hypothesis about the effect of different stimuli on male T levels, subjects were randomly assigned to view one of two videos on each testing occasion: one depicting sexual activity, and one depicting dental surgery. The video type did not consistently affect T levels, nor did it affect performance on the MRT. Therefore, we combined results from both video conditions in our analysis.

[3] Because interaction with a female investigator may affect subjects' T levels (Roney et al., 2003), subject–investigator interaction was kept to a minimum until after saliva samples were collected. Conservatively-dressed investigators greeted subjects, described the basic elements of the experiment, and gave subjects detailed instructions to read on their own before beginning the experiment.

type). We removed outliers with an iterative process—when an outlier was removed from a particular cell, the calculation was then repeated to determine whether the next highest value in that cell was also an outlier. Less than one percent (0.8%) of the RTs were excluded as outliers. ERs represent the percentage of trials (out of the total in the task or in a given cell of the design) on which the subject did not answer correctly. One subject was excluded from data analysis because his pattern of performance, including exceptionally fast RTs and a mean ER of nearly 50%, indicated that he was not following instructions for the task.

3. Results

We established the reliability and validity of the MRT and the reliability of the T measures before we examined the relationship between salivary T measures and the slopes and intercepts of both RT and ER.

3.1. Validation of mental rotation task

To ensure that our implementation of the Shepard and Metzler task did in fact assess mental rotation, we performed linear contrasts on the ER and RT scores, averaged for each subject according to rotation angle. As expected, linear contrasts revealed that RT increased with increasing angle ($t(26) = 7.76$, $P < 0.0001$), and ER increased with increasing angle ($t(26) = 10.63$, $P < 0.0001$). To test the reliability of this task, we correlated performance measures from Day 1 with those on Day 2, and found that with the exception of the slope for Different trials, measures of individual differences in performance remained consistent over the two test days (see Table 1, below).

Paired *t*-tests revealed that the effects of angle differed for Same and Different trial types. For both RT and ER, slopes were higher for Same trials: RT ($t(26) = 8.47$, $P < 0.0001$); ER ($t(26) = 6.17$, $P < 0.0001$).

3.2. Testosterone assay results

The sample means of the untransformed Day 1 and Day 2 T values were 418 (S.D. = 232) and 387 (S.D. = 135) pmol/l. The sample means of the log-transformed T values for Day 1 and Day 2 were 2.56 (S.D. = 0.23) and 2.55 (S.D. = 0.17), respectively. The correlation between the Day 1 and Day 2 values was significant ($r = 0.67$; $P < 0.001$). Samples were assayed in two groups. Intra-assay reliability, as measured by the correlation between the duplicates, was $r = 0.75$ for group 1 and $r = 0.86$ for group 2. Mean T values were higher for the AM sample (476 pmol/l, S.D. 214) than for the PM sample (332, S.D. 117) ($t(25) = 2.27$, $P = 0.03$). This difference is consistent with the expected diurnal decline in T levels. The correlation between T level and age was not significant ($r = 0.19$, $P = 0.34$), which was expected, given the restricted age range of the subjects.

3.3. MRT performance measures

To test the hypothesis that T levels are related to mental rotation processes rather than non-rotation processes that also contribute to task performance, we computed the slopes and intercepts of the ER and RT rotation functions. The slope of these regressions indicates the average change, in either milliseconds (RT) or percent incorrect (ER), per additional degree of rotation from the upright (0° rotation). That is, the slope indicates the cost, in terms of time or accuracy, associated with rotating the target object one additional degree. The intercept, in contrast, is a measure of the contributions to performance of all non-rotation processes. The primary results of interest were correlations between these measures of MRT performance and the measures of salivary T. To test whether the relationship between T and task performance was affected by trial type, ERs and RTs were calculated separately for Same and Different trials at each degree of angular disparity. Because any of these measures

Table 1
Descriptive statistics for performance on mental rotation test

Measure	Error rate			Response time (ms)		
Trial type (# trials)	Day 1	Day 2	Reliability	Day 1	Day 2	Reliability
Composite						
All (80)	17 (2)	13 (2)	0.76	6232 (495)	4702 (434)	0.83
Same (40)	13 (1)	12 (1)	0.66	5089 (416)	3880 (386)	0.81
Different (40)	21 (4)	14 (3)	0.71	7374 (601)	5523 (509)	0.78
Slope (per degree of rotation)						
All (80)	0.13 (0.01)	0.11 (0.01)	0.12	18 (2)	17 (2)	0.58
Same (40)	0.19 (0.02)	0.20 (0.02)	0.50	28 (3)	25 (3)	0.52
Different (40)	0.08 (3.5)	0.03 (0.01)	0.08	7 (2)	9 (3)	0.02
Intercept						
All (80)	6 (1.8)	4 (1.5)	0.46	4755 (424)	3253 (302)	0.76
Same (40)	−3 (0.89)	−4 (0.70)	0.44	2699 (250)	1768 (229)	0.87
Different (40)	15 (3.4)	12 (3.0)	0.49	6811 (658)	4738 (416)	0.65

Standard errors are given in parentheses.

Table 2
Correlations between T level and performance on mental rotation test (disattenuated correlations in parentheses)

Measure	Error rate		Response time (ms)	
Trial type (# trials)	Day 1	Day 2	Day 1	Day 2
Composite				
All (80)	−0.41* (−0.47)	−0.45* (−0.51)	−0.57** (−0.62)	−0.16 (−0.17)
Same (40)	−0.03 (−0.03)	−0.18 (−0.22)	−0.56** (−0.62)	−0.14 (−0.15)
Different (80)	−0.45* (−0.03)	−0.50** (−0.59)	−0.54** (−0.61)	−0.17 (−0.19)
Slope				
All (80)	0.06 (0.13)	−0.21 (−0.60)	−0.04 (−0.05)	0.03 (0.03)
Same (40)	0.01 (0.01)	−0.18 (−0.25)	−0.22 (−0.30)	−0.11 (−0.15)
Different (40)	−0.08 (−0.28)	−0.13 (−0.46)	0.39* (2.75)	0.18 (1.27)
Intercept				
All (80)	−0.38 (−0.56)	−0.38* (−0.56)	−0.66*** (−0.75)	−0.19 (−0.21)
Same (40)	−0.11 (−0.16)	0.12 (0.18)	−0.67*** (−0.71)	−0.14 (−0.15)
Different (40)	−0.39* (−0.55)	−0.40* (−0.57)	−0.59*** (−0.73)	−0.20 (−0.24)

* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.

may be affected by practice, all measures were also reported separately for Day 1 and Day 2. Table 1 presents descriptive statistics for ER and RT on the MRT, and reliability scores for each measure (as referred to in Section 3.1).

3.4. *T levels and performance*

We correlated T level (on Day 1 and on Day 2) with the four measures of task performance (slope and intercept for RT and ER, with correlations including only data collected on the same day). Table 2 presents Pearson's correlation coefficients relating T levels to MRT results for the two days and two types of trials, with disattenuated correlations presented in parentheses. Several results are noteworthy. First, the correlation coefficients relating T to ER are negative and significant on Day 1 and Day 2. When we examined the T/performance correlations by trial type, we found that they were significant only for the Different trials. Moreover, these significant correlations involved the intercept of the rotation function, and not the slope. We also correlated mean ER with T for each time group separately and, consistent with the direction of the correlation of T with ER across time groups, we obtained negative correlations (for the subjects tested at 11:00, $r = -0.38$, and for subjects tested at 1:15, $r = -0.55$).

We found that the correlations between T and RT were significantly negative on Day 1, but did not find a significant correlation between the two variables on Day 2. The Day 1 correlations between T and RT were significant in both Same and Different trials. Consistent with the relationship between T and ER, significant correlations emerged between T and the RT intercept. With one exception (a positive correlation between T and the RT slope for Different trials), the correlations with the RT slope were not significant (for Day 1 or Day 2) as shown in Table 2.

To address the possibility that the T/performance correlations may have been affected by low reliabilities of some of the cognitive measures, we corrected these correlations for attenuation (Muchinsky, 1996) (presented in Table 2). These disattenuated correlations suggest that the reliability differences between slopes and intercepts do not account for the meaningful differences between T correlations with rotation performance.[4]

3.5. *Difficulty*

The correlations between T and ER were high and significant for Different trials, and low and non-significant for Same trials on both testing days. We found that the subjects were less accurate in general on the Different trials than on the Same trials, although the effect was only marginally significant ($t(26) = 2.03$, $P = 0.051$), and that they required more time to respond on the Different trials than on the Same trials ($t(26) = 7.74$, $P < 0.0001$). In comparing performance on Day 1 and Day 2, we found that subjects were more accurate ($t(26) = 3.26$, $P = 0.001$), and faster on both Different ($t(26) = 5.39$, $P < 0.0001$) and Same trials ($t(26) = 4.54$, $P < 0.0001$) on Day 2. These findings of longer RT and higher ER in Different trials are consistent with the results of the original Shepard and Metzler study (Shepard & Metzler, 1971). Given that the Different trials were more difficult, we asked whether T might be correlated with ER in the Different, but not the Same trials,

---

[4] The reliabilities for the intercept measurements on Different trials are fairly substantial, and comparable with those of the slope measurement for Same trials—our most precise measure of rotational performance per se. The disattenuated correlations of slope on Same trials with T are minimal, whereas they are high for T and intercept on Different trials. The data provide no support for the idea that rotational performance per se is associated with T, even when differences in reliability are taken into account.

because of a relationship between T and the level of difficulty per se, rather than to a processing demand imposed on subjects that differed for the two types of trials. To explore this question, we recalculated the ER results while controlling for difficulty. Because trials with greater angular disparities are more difficult than those with lesser angular disparities, we recalculated the ER results so that they included only the three greatest angular disparities (and most difficult trials) for the Same trials, and only the three smallest angular disparities (and hence the easiest trials) for the Different trials. In other words, we excluded ERs from the 16 trials at 120 and 180° in the Different trials, and from the 16 trials at 0 and 40° from the Same trials. Across all subjects and testing sessions, these ERs were comparable for Same and Different trials ($t(26) = 0.53$, $P = 0.53$).

We then recalculated the correlations between the subjects' mean T levels and the revised ERs, and found that the original results were preserved. The correlation between T and ER in the Different trials remained significant ($r = -0.47$, $P = 0.01$), whereas the correlation with ER in the Same trials remained low and non-significant ($r = 0.06$, $P = 0.80$). Also consistent with the original results, the correlation between T and the ER intercept in the Different trials was significant, whereas the correlation with ER intercept in the Same trials was non-significant (Different trials: $r = -0.40$, $P = 0.04$; Same trials: $r = 0.09$, $P = 0.60$). The correlations between T and the slopes of the rotation functions remained low and non-significant (Different slope: $r = 0.00$, $P = 0.99$; Same slope: $r = 0.04$, $P = 0.84$).

4. Discussion

We investigated the relationship of male T to different processes used in a MRT, and found that higher T levels are associated with lower error rates and faster responses. Interestingly, for both ER and RT, T was correlated not with the slopes of the rotation functions, but with the intercepts. Our results provide no evidence that the efficacy of the rotation process is correlated with T; rather, T appears to facilitate processes related to other aspects of the task, which may or may not be spatial in nature.

Higher T levels were associated with lower ERs only in trials where the two objects were different. A similar result was reported by Kerkman, Wise, and Hardwood (2000), who also used a MRT that required subjects to distinguish between Same and Different objects. The male advantage in performance was evident only in the trials in which the objects were different—there was no sex difference in accuracy in trials in which objects were the same. We also showed that the emergence of significant correlations between T and ER for Different and not Same trials was not an artifact of differences in the difficulty of the trial types. Other researchers have found that difficulty in MRTs is not related to the effect size of the sex difference in performance (e.g., Collins & Kimura, 1997), supporting the possibility that the effects of T levels might not be moderated by task difficulty. An unexpected result was the effect of test day on the relationship between RT and T: T level was highly, negatively correlated with RT on Day 1, but was not correlated on Day 2. The RTs on Day 2 were sufficiently long to suggest that a simple ceiling effect cannot explain the lack of correlation. High T may facilitate quick reactions, risk-taking (Gerra et al., 1999), and the exploration of novel stimuli (Cornwell-Jones & Kovanic, 1981)—suggesting the possibility that subjects' reaction is highly responsive to T level when novel stimuli are involved, but not with more familiar stimuli.

Because males outperform females on MRTs, and this advantage is related to higher T levels in males, a reasonable hypothesis is that male-typical T levels enhance the ability to mentally rotate. Although the results presented here do not contradict this, they suggest an alternative hypothesis: higher T in males relative to females enhances abilities associated with non-rotation processes drawn upon in MRTs. Our results also suggest that these abilities may be particularly important in solving trials in which the objects are different. It is possible that these same abilities, that give high T men an advantage on the task reported on here, contribute significantly to the male advantage on tests such as the Vandenberg and Kuse. More research is needed to determine if this is indeed the case; if so, then the magnitude of the sex-difference in performance on MRTs (accounting for about 20% of the variance in performance) does not accurately reflect the true size of the sex difference in the ability to mentally rotate. Instead, the large sex difference may be at least partially a result of male superiority in abilities related to non-rotation task components, particularly those related to discriminating different objects.

A study by Karadi et al. (2001) found evidence in support of the hypothesis that performance on MRTs may be significantly influenced by abilities unrelated to ER and RT slopes. In addition to completing an MRT, subjects completed tasks that measure abilities based on Kosslyn's (1994) clustering of the cognitive variables involved in mental rotation (focused attention, visual scanning, perceptual decision and visual memory). Subjects who scored high on MRTs scored higher on the perceptual decision and focused attention tests, but there was no difference between high and low MRT scorers on the visual scanning or visual memory tasks. This finding is intriguing because the former two processes are not involved in rotation per se, whereas the second two may be (see Kosslyn, 1994).

We suggest that among males, T may facilitate performance on MRTs because of its relationship to cognitive processes that are separate from the slope-related rotation component. The mechanisms that relate T to performance may differ between the sexes, and additional studies are necessary to determine whether results from males can be generalized to the females, or whether these mechanisms will be consistent across the sexes. Additional research should address how accuracy on MRTs relates to specific abilities tapped by Same and Different trials, respectively, and how these abilities affect performance on tests such as the Vandenberg and Kuse.

Our results suggest one possible explanation for the male advantage on MRTs: high T may facilitate accuracy primarily because it influences abilities related to the encoding, comparison, initiation and/or decision processes, not the rotation process itself. In order to gain a clearer picture of the factors affecting performance on MRTs, and the true nature of the relationship of T to ability, these candidate processes could be measured separately and related to T levels. In addition, to measure mental rotation ability most precisely, investigators should consider using tasks that record the slope of the function relating RT and ER to angle, in trials in which the objects can actually be rotated into congruity.

Acknowledgements

We are grateful to the NIH (award 5 R01 MH60734) and the NSF (ROLE award REC-0106760) for funding for this research. Christopher F. Chabris was supported by a Director of Central Intelligence postdoctoral fellowship. We thank Christine Monta and Mary O'Rourke for assistance, and Richard Wrangham, Susan Lipson, Jennifer Shephard and Steve Gangestad for insightful comments. We also wish to thank the anonymous reviewers for their helpful comments on an earlier draft of this paper.

References

Alexander, G. M., Swerdloff, R. S., Wang, C., Davidson, T., McDonald, V., & Steiner, B. et al., (1998). Androgen-behavior correlations in hypogonadal men and eugonadal men: II. Cognitive abilities. *Hormones Behavior*, *33*(2), 85–94.

Christiansen, K., & Knussmann, R. (1987). Sex hormones and cognitive functioning in men. *Neuropsychobiology*, *18*(1), 27–36.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments & Computers*, *25*(2), 257–271.

Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behavioral Neuroscience*, *111*(4), 845–849.

Cornwell-Jones, C. A., & Kovanic, K. (1981). Testosterone reduces olfactory neophobia in male golden hamsters. *Physiology & Behavior*, *26*(6), 973–977.

Gerra, G., Avanzini, P., Zaimovic, A., Sartori, R., Bocchi, C., & Timpano, M. et al., (1999). Neurotransmitters, neuroendocrine correlates of sensation-seeking temperament in normal humans. *Neuropsychobiology*, *39*(4), 207–213.

Gouchie, C., & Kimura, D. (1991). The relationship between testosterone levels and cognitive ability patterns. *Psychoneuroendocrinology*, *16*(4), 323–334.

Granger, D. A., Schwartz, E. B., Booth, A., & Arentz, M. (1999). Salivary testosterone determination in studies of child health and development. *Hormones Behavior*, *35*(1), 18–27.

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain Cognition*, *45*(3), 392–414.

Hampson, E., Rovet, J. F., & Altmann, D. (1998). Spatial reasoning in children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Developmental Neuropsychology*, *14*(2–3), 299–320.

Hier, D. B., & Crowley Jr., W. F. (1982). Spatial ability in androgen-deficient men. *New England Journal of Medicine*, *306*(20), 1202–1205.

Isgor, C., & Sengelaub, D. R. (1998). Prenatal gonadal steroids affect adult spatial behavior. *Hormones Behavior*, *34*(2), 183–198.

Janowsky, J. S., Oviatt, S. K., & Orwoll, E. S. (1994). Testosterone influences spatial cognition in older men. *Behavioral Neuroscience*, *108*(2), 325–332.

Karadi, K., Kallai, J., & Kovacs, B. (2001). Cognitive subprocesses of mental rotation: Why is a good rotator better than a poor one? *Perceptual Motor Skills*, *93*(2), 333–337.

Kerkman, D. D., Wise, J. C., & Hardwood, E. A. (2000). Impossible "mental rotation" problems. A mismeasure of women's spatial abilities? *Learning & Individual Differences*, *12*(3), 253–269.

Kosslyn, S. M. (1994). Image and brain: The resolution of the imagery debate. Cambridge, MA; MIT Press.

Kruijver, F. P., Fernandez-Guasti, A., Fodor, M., Kraan, E. M., & Swaab, D. F. (2001). Sex differences in androgen receptors of the human mamillary bodies are related to endocrine status rather than to sexual orientation or transsexuality. *Journal of Clinical Endocrinology & Metabolism*, *86*(2), 818–827.

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. *Developmental Psychology*, *38*(2), 236–253.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*(6), 1479–1498.

Lipson, S., & Ellison, P. (1989). Development of protocols for the application of salivary steroid analyses to field conditions. *American Journal of Human Biology*, *1*, 249–255.

Lipson, S. F., & Ellison, P. T. (1996). Comparison of salivary steroid profiles in naturally occurring conception and non-conception cycles. *Human Reproduction*, *11*(10), 2090–2096.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, *23*(4), 337–341.

McKeever, W. F., Rich, D. A., Deyo, R. A., & Conner, R. L. (1987). Androgens and spatial ability: Failure to find a relationship between testosterone and ability measures. *Bulletin of the Psychonomic Society*, *25*(6), 438–440.

Moffat, S. D., & Hampson, E. (1996). A curvilinear relationship between testosterone and spatial cognition in humans: Possible influence of hand preference. *Psychoneuroendocrinology*, *21*(3), 323–337.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational & Psychological Measurement*, *56*(1), 63–75.

Neave, N., Menaged, M., & Weightman, D. R. (1999). Sex differences in cognition: The role of testosterone and sexual orientation. *Brain & Cognition*, *41*(3), 245–262.

Oosthuizen, S. (1991). Sex-related differences in spatial ability in a group of South African students. *Perceptual & Motor Skills*, *73*(1), 51–54.

Roney, J. R., Mahler, S. V., & Maestripieri, D. (2003). Behavioral and hormonal responses of men to brief interactions with women. *Evolution and human Behavior*, *24*, 365–375.

Shepard, R. N., & Judd, S. A. (1976). Perceptual illusion of rotation of three-dimensional objects. *Science*, *191*(4230), 952–954.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Sherry, D. F., Jacobs, L. F., & Gaulin, S. J. (1992). Spatial memory and adaptive specialization of the hippocampus. *Trends in Neurosciences*, *15*(8), 298–303.

Silverman, I., Kastuk, D., Choi, J., & Phillips, K. (1999). Testosterone levels and spatial ability in men. *Psychoneuroendocrinology*, *24*(8), 813–822.

Slabbekoorn, D., van Goozen, S. H., Megens, J., Gooren, L. J., & 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*(4), 423–447.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, Amsterdam*, *30*, 276–315.

Van Goozen, S. H., Cohen-Kettenis, P. T., Gooren, L. J., & Frijda, N. H. et al., (1994). Activating effects of androgens on cognitive performance: Causal evidence in a group of female-to-male transsexuals. *Neuropsychologia*, *32*(10), 1153–1157.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual* & *Motor Skills*, *47*(2), 599–604.

Voyer, D., Voyer, S., & Bryden, M. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68*(1), 77–94.

Williams, C. L., & Meck, W. H. (1991). The organizational effects of gonadal steroids on sexually dimorphic spatial ability. *Psychoneuroendocrinology*, *16*(1–3), 155–176.

Witkin, H. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality*, *19*, 1–15.

Witkin, H., & Asch, S. (1948). Studies in space orientation. III. Perception of the upright in the absence of a visual field. *Journal of Experimental Psychology*, *38*, 603–614.

Yen, S. S. C., Jaffe, R. B., & Barbieri, R. L. (1999). *Reproductive endocrinology: Physiology, pathophysiology, and clinical management*. Philadelphia: Saunders.