ELSEVIER

Physiology & Behavior 90 (2007) 634–642

PHYSIOLOGY & BEHAVIOR

Effects of ability- and chance-determined competition outcome on testosterone

Sari M. van Anders, Neil V. Watson *

Department of Psychology, Simon Fraser University, RCB 5246, 8888 University Drive, Burnaby, BC, Canada V5A 1S6

Received 30 June 2006; received in revised form 6 November 2006; accepted 20 November 2006

Abstract

Winning competitions has been shown to lead to higher testosterone (T) relative to losing in men and males of other species. In Experiment 1, 38 women and 37 men provided a saliva sample, completed a novel computer-based vocabulary competition task at which they won or lost based on their own ability, provided feedback about the competition via questionnaire, and then produced a second saliva sample. Task outcome and performance was not sexually differentiated, and overall task performance was negatively correlated with T. Male but not female winners had lower baseline and post-competition T, and male losers had a larger decrease in T from baseline to post-competition. In Experiment 2, 31 men and 43 women completed the same as above, but were randomly assigned to win or lose. In this case, competition outcome did not affect T for men but there was an effect such that women who would have had an ability-determined loss showed a larger decrease in T than women who would have had an ability-determined win. Thus, earned wins appear to attenuate a decline in T in men, consistent with past research into the competition effect and T, and perhaps women under complex circumstances.

Keywords: Competition; Sex; Men; Women; Testosterone; Win; Androgen; Vocabulary; Challenge; Cognition; Ability; Chance; Random

1. Introduction

Research into the modulation of sex steroids like testosterone (T) by competition outcome has generally confirmed the presence of a 'Competition Effect' whereby winners of a competition exhibit either higher post-competition T than losers, higher post-competition T than pre-competition T (while losers do not), or less of a decrease in T from pre-to post-competition; these are generally thought to reflect the same physiological process, though this is open to study. The competition effect appears to occur in a wide variety of mammalian taxa, including nonhuman primates (e.g. male rhesus monkeys [1]; male mandrills [2]), although the precise function served by competition-induced T shifts in humans remains to be determined. It is noteworthy that the competition effect is conceptually related (see [5]) but differs from the Challenge Hypothesis [3], in which it is argued that T is elevated in males (and perhaps females) during times of competitive social interactions. For example, the Challenge Hypothesis does not predict differential androgen responsiveness to competition outcome, per se, and it is unclear why only winners should show higher T parameters relative to losers when both are engaging in competition. Examining what aspects of competition outcome lead to changes in T might provide clues to the functionality of the competition effect.

There are data supporting the competition effect in humans (see [4–7] for reviews), with a majority of studies using sports paradigms as models of competitive encounters (e.g. male wrestling [8]; tennis [9,10]). However, investigations looking at other sports sometimes fail to find the competition effect among male winners (basketball [11]; judo [12–14]) and have not found the effect among female winners (rugby [15]; soccer [16]). These sports differ in many ways that may be related to the different results (along with very small sample sizes in some cases), but which highlight problems associated with using sports paradigms. Because athletic competitions involve varying degrees of physical exertion, team bonding, individual efforts, and so on, they are likely complicated models. For example, physical exertion can

* Corresponding author. Tel.: +1 604 291 3550; fax: +1 604 291 3427.
E-mail address: nwatson@sfu.ca (N.V. Watson).

0031-9384/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.physbeh.2006.11.017

increase T or decrease T depending on intensity and gender (e.g. [17]), and playing in a 'home' game leads to differential T parameters than playing 'away' [18].

Other more cognitive tasks have been used to study the competition effect, producing somewhat more consistent results, although one study failed to find any effect in men or women using a simple video game [19]. Male winners of chess matches showed increased T [20], as did winners of a reaction time test, even though outcome was predetermined [21]. Similar increases have been reported for winners of a series of coin tosses, even when the participants were reminded of the chance nature of the contest [22], in contrast to an earlier chance-determined study in which a competition effect was not observed for winners of a laboratory lottery [10]. Even vicarious triumph can produce a competition effect: Male sports fans also show an increase in T when their favored team wins [23] similar to a finding from fish watching other fish fight [24]. Taken together, it appears that sustained involvement or investment is required for the competition effect to occur, but whether ability is a particular requisite is currently unclear. This issue of ability versus chance-determined outcome is key to understanding any potential functionality of the competition effect: e.g. if the competition effect influences the likelihood of future competition, a falsely (or randomly) enhanced likelihood could potentially lead an individual to compete against their interests. Matching ability to outcome should avoid T responding to inaccurate outcomes. Or, if the competition effect is solely about increased status, then a win (whether earned or not) should be sufficient to lead to increased T. As well, as Salvador notes [6], complex psychological variables likely mediate any direct effects of competition outcome on endocrine parameters, and ability versus chance-determined outcome addresses this complexity. Examining the contributions of ability and chance allows us to understand whether the competition effect is due to good performance, external recognition of good performance, or outcome.

A Competition Effect in women has not been reported, but the paradigms utilized to study women to date (simple video game [20,19]; rugby [15]; soccer [16]) have not been effective in men either. Evidence indicating that women and men are equally competitive [25], along with the finding that T levels correlate with competition in women [26], suggests that a Competition Effect could be evident in women, under appropriate experimental conditions.

Here we describe two studies with the goal of determining whether the competition effect occurs in both women and men, and whether ability is a requisite. To do this, we set out to develop a novel laboratory competition paradigm that leads to higher T in winners relative to losers, that is relevant to students, and does not show sex differences in performance to maximize the possibility of finding the effect, especially in women. The development of such a paradigm would provide more practical, reliable, and controlled ways of furthering our understanding of the effects of competition on sex hormones. We conducted two studies: In the first, participants won or lost the competition based on their own ability to meet our predetermined cut-off (explained further in Materials and procedures); In the second, participants were randomly assigned to win or lose regardless of their ability. Thus, we were able to test if winning or losing our competition affected T, as well as whether outcome must be earned (ability-determined) or not (chance-determined) to affect T.

2. Experiment 1

2.1. Goal

In experiment 1, our goal was to examine whether women and men who won the competition by their own ability would shower higher T post-test relative to those who lost by their own ability.

2.2. Method

2.2.1. Participants

Seventy-six participants were recruited from the Department of Psychology undergraduate participant pool at Simon Fraser University, and participated for course credit. There were 39 women (mean age = 19.74 years, SD = 1.92) and 37 men (mean age = 20.03 years, SD = 2.06). No participants reported using exogenous hormones in response to questions of hormone use, including hormonal contraceptives, except for one woman who was excluded from the analyses (pill-users were explicitly asked not to sign up for participation), leaving 75 participants. According to the Kinsey scales of sexual experience and fantasy [27], all participants were heterosexual, except for three who were bisexual, and two who omitted responding to one of the scales.

There were 39 participants with English as their first language and 36 participants whose first language was another language (Burmese, Cantonese, Cebuano, Farsi, Hindi, Japanese, Mandarin, Punjabi, Russian, Spanish, Swedish, Taiwanese). There were 51 participants with English as the language they felt most comfortable speaking, and 25 with other languages (Burmese, Cantonese, Farsi, Hindi, Mandarin).

2.3. Materials

2.3.1. Saliva samples and hormone assays

Participants were asked to refrain from eating, smoking, drinking, or brushing their teeth for 1 h prior to testing. Two saliva samples were collected from each participant (see Procedures) in polystyrene tubes pre-treated with sodium azide. Saliva was stimulated with Trident cherry sugar-free gum. The tubes were frozen at −20 °C until assay.

Radioimmunoassays of the samples were performed in a single batch at the Endocrine Core Lab at Yerkes National Primate Research Center, Emory University, all in duplicate, using a modified kit from Diagnostic Systems Laboratories (Webster, TX). The sensitivity was 2 pg/mL. The interassay coefficient of variation was 8.77% at 0.65 ng/mL and 6.88% at 5.06 ng/mL. The intra-assay coefficient of variance was 6.54% at 98.82 pg/mL.

2.3.2. Competition task

We adapted the Verbal Meaning Subscale of the Primary Mental Abilities battery [28] for use on the computer and to indicate outcome based on performance, i.e. win or loss. We

used this test because, in its original form, performance does not differ by sex. We also reasoned that it would be relevant to the self-concept of students, because students are routinely involved in developing, utilizing, or testing their vocabulary in classes, exams, and standardized tests. The task is to match a target word to one of five possible words that most closely matches the target's meaning. There were 60 trials.

The experiment was conducted using Presentation® software (Version 0.71) [29]. The title image noted "Competition Test," and instructions as per the original test [28] followed, except that the word 'test' was replaced with the word 'competition'. Participants were given three practice trials, and were informed that winning depended on correctly answering as many questions as possible within 4 min; participants were told to try their best to win. Questions were presented with the target word centered above the five options. Participants indicated their response by pressing the number key that corresponded with the answer option [1–5]. Each trial remained on the screen until participants made a selection.

We programmed the test such that participants who answered 20 or more questions correctly 'won,' while those who answered fewer 'lost.' This cut-off was based on a median split of performance from previous research [30], and was chosen to provide roughly equal numbers of winners and losers. The test ended at 4 min or 60 questions answered, whichever came first. When the test ended, losers saw the words "You Lose," and the letters dropped to the bottom of the screen in sequence, followed by an "unsmiley face" (frown instead of smile) that grew larger in the middle. Winners saw the words "You win!!", and the words changed colors over a large "first-place"-type ribbon. We attempted to keep the stimuli somewhat similar, using changes and motion in each. The stimuli were intended to reinforce the win or loss.

Reaction time and answers were recorded by Presentation. We calculated average reaction time (RT), average reaction time for questions answered correctly (Correct RT), average reaction time for questions answered incorrectly (Incorrect RT), number of questions answered (Total Questions), number of questions answered correctly (Total Correct), percentage of questions answered correctly (Percent Correct), and number of questions answered incorrectly (Total Incorrect).

2.3.3. Paper-and-pencil questionnaires

Participants completed the Post-Competition Questionnaire, labeled "Winner's Questionnaire" or "Loser's Questionnaire" (to reinforce the win or loss), which contained the following questions and 5-point response scales. [1] "How did you feel about doing the test?" with responses from "I liked doing it very much" to "I disliked doing it very much." [2] Winners: "How did you feel about winning the test?" with responses from "Very pleased" to "Neutral." Losers: "How did you feel about losing the test?" with responses from "Very disappointed" to "neutral." [3] "How would you describe the task?" with responses from "Very challenging" to "Not challenging at all." [4] "What factor do you feel is most likely involved with your win/loss?" with responses from "All luck" to "All ability." [5] "When you are in a competition, are you:" with responses from "Very competitive" to "Not competitive at all." [6] "When you were doing the test, how much did you care about winning?" with responses from "I cared a lot" to "I didn't care at all."

Participants also completed a Health and Background Questionnaire, and the Profile of Mood States (POMS: [31]) to ascertain post-competition mood. The POMS provides one total mood disturbance (TMD) score and six subscale measures: tension-anxiety, depression-dejection, vigor, anger-hostility, fatigue, and confusion. Sexual orientation was determined via the Kinsey questions of sexual orientation [32]) where participants indicate their degree of same-versus opposite-sex sexual fantasy and sexual experience.

2.4. Procedure

All procedures received prior approval from the Simon Fraser University Research Ethics Board. Participants were tested individually in a small room by one of two white female experimenters in their mid-20′s. Participants were not aware that the study involved a competition prior to arriving and only knew that they would be completing a task on the computer, and questionnaires, and providing saliva samples. Each session took no longer than 30 min, and all sessions took place between 1400 and 1800 h to control for diurnal T rhythms.

Participants completed informed consent (which informed participants of the competitive nature of the study), and then immediately produced the first saliva sample (Baseline T). Participants were then told they would be doing a competition on the computer and to try their best to win. Participants then completed the competition task on the computer, and the experimenter sat watching participants while they engaged in the task. At task completion, the experimenter told the losers "Too bad, you lost" and the winners "Congratulations, you won." Then, participants completed the Post-Competition Questionnaire (Winner's or Loser's version), then the POMS, and then the Health and Demographics Questionnaire. Participants provided the second saliva sample (PostComp T) either after all questionnaires were completed, or at 25 min from the session's start (whichever came first). Participants were then given debriefing forms and their participation credit.

2.5. Statistics

All analyses were conducted using the Statistical Package for the Social Sciences (SPSS Inc., Chicago IL), ver. 11.0.1. To analyze the difference between baseline and post-competition T levels, a new variable was computed (Pre–Post T) in which post-competition T was subtracted from baseline competition T (i.e. Baseline-Postcomp). One Postcomp sample was accidentally lost, leaving 74 Postcomp samples for T assay. Chi-square analyses and independent samples *t*-tests are used to assess group differences. Paired *t*-tests are used to assess within-participant differences over time. Pearson Product Moment Correlations are used to assess associations between variables. All analyses are two-tailed. Corrections for multiple tests were not undertaken because of the novel nature of our paradigm and our desire to understand dependent variables of use in characterizing our paradigm but unrelated to our hypotheses.

3. Results

3.1. Sex similarities on competition parameters

There was no significant difference between women and men in the following, confirming that our task was not gender-specific: competition outcome (i.e. win/loss), χ^2 (1)=0.19, p=0.665, caring about winning, $t(73)=1.01$, p=0.316, feelings about the test (i.e. like/dislike), $t(73)=0.44$, p=0.659, level of challenge, $t(73)=1.08$, p=0.284, factor involved in win (i.e. luck, ability), $t(73)=-0.17$, p=0.864, Total Questions, $t(73)=0.31$, p=0.759, Total Correct, $t(73)=1.36$, p=0.178, Total Incorrect, $t(73)=-1.49$, p=0.140, Percent Correct, $t(73)=1.56$, p=0.123, Average RT, $t(73)=-0.51$, p=0.613, Average Correct RT, $t(73)=-.49$, p=0.627, and Average Incorrect RT, $t(73)=-0.21$, p=0.836. Women and men did not differ significantly in how competitive they reported being in competitions, $t(73)=-0.44$, p=0.662.

3.2. Competition outcomes, post-competition measures, and language

There were no significant differences between winners and losers in any of the POMS subscales. For men, there was a significant correlation between Pre–Post T and fatigue, $r(34)=0.43$, p=0.009, and trends towards significance between Pre–Post T and Tension-Anxiety, $r(34)=0.31$, p=0.064, and TMD, $r(33)=0.33$, p=0.054. For women, there were no significant (or trends for) correlations between Pre–Post T and any POMS measures. There were no significant correlations between any of the T measures and any of the Post-Competition Questionnaire measures, for women or men. Not surprisingly, those with English as a first language were more likely to win than those with English not as a first language, $t(52)=4.67$, $p<0.001$, as were those who were more comfortable with English, $t(35)=5.55$, $p<0.001$.

3.3. Task parameters

Because our task is novel, we entered the performance indicators (Total Questions, Total Correct, Total Incorrect, Percent Correct, RT, Correct RT, Incorrect RT) into a correlation to see how the variables were interrelated; all were highly correlated. Please see Table 1 for correlations and p-values. We also analyzed how winners and losers differed: winners had higher.

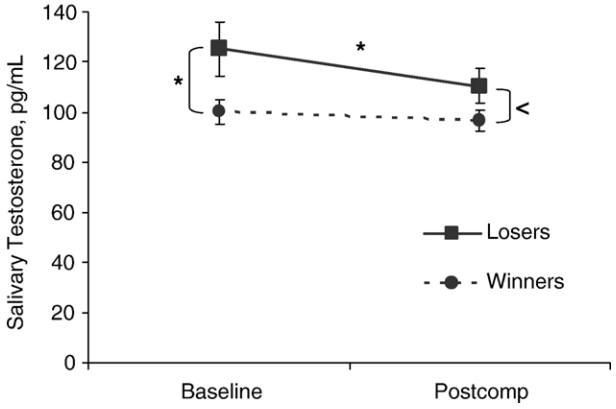

Fig. 1. Mean testosterone for men at baseline*, post-competition<, and the change* from baseline to post-competition. * indicates $p<0.05$. '<' indicates $p<0.05$.

Total Questions, $t(73)=7.02$, $p<0.001$, Total Correct, $t(65)=13.48$, $p<0.001$, Percent Correct, $t(73)=7.86$, $p<0.001$, faster RT, $t(24)=-5.41$, $p<0.001$, faster Correct RT, $r(26)=-5.97$, $p<0.001$, and faster Incorrect RT, $r(24)=-3.68$, p=0.001. There was a trend towards winners having lower Total Incorrect, $t(73)=-1.77$, p=0.081.

3.4. Competition outcomes and T

There were no significant differences between participants with English or another language as their most comfortable language in Baseline T, PostComp T, or Pre–Post T. Thus, any differences between winners and losers were not due to possible ethnic or nationality differences in T. There was also no significant correlation between the time elapsed between the two saliva samples and the change in T, indicating that any differences in T between winners and losers were not due to any differences in time between samples. As expected, there were large and significant differences between men and women in Baseline T, $t(40)=15.99$, $p<0.001$, and PostComp T, $t(45)=19.43$, $p<0.001$, so analyses on competition outcomes and T were conducted separately by sex.

3.4.1. Men

Winners showed a smaller decrease in T levels than did losers (Fig. 1). At Baseline, men who would go on to win had significantly lower T than men who would go on to lose, $t(35)=-2.42$, p=0.021. There was a trend for men who won to have

Table 1
Pearson product–moment correlations between performance indicators (df=73)

	Total questions	Total correct	Total incorrect	Percent correct	RT	Correct RT	Incorrect RT
Total questions	1.00						
Total correct	0.76*	1.00					
Total incorrect	0.37*	−0.32*	1.00				
Percent correct	0.30*	−0.83*	−0.74*	1.00			
RT	−0.94*	−0.72*	−0.33*	−0.32*	1.00		
Correct RT	−0.90*	−0.71*	−0.29*	−0.35*	0.92*	1.00	
Incorrect RT	−0.87*	−0.54*	−0.50*	−0.07	0.94*	0.81*	1.00

* indicates $p<0.01$.

Table 2
Pearson product–moment correlations between performance indicators and testosterone in women and men

	Women			Men (*df*=35)		
	Baseline T (*df*=36)	Postcomp T (*df*=35)	Pre–Post T (*df*=35)	Baseline T (*df*=35)	Postcomp T (*df*=35)	Pre–Post T (*df*=35)
Total questions	−0.41*	−.23	−0.11	−0.08	−0.15	0.05
Total correct	−.25	−.21	0.07	−0.33*	−0.27	−0.26
Total incorrect	−.26	−.05	−0.24	0.39*	0.20	0.48**
Percent correct	−.01	−.10	0.20	−0.42**	−0.32<	−0.37*
RT	0.47**	0.31<	0.10	0.20	0.25	.01
Correct RT	0.46**	0.32<	0.06	0.14	0.17	.02
Incorrect RT	0.44**	0.24	0.16	0.08	0.16	−.08

< indicates $p<0.10$; * indicates $p<0.05$; ** indicates $p<0.01$.

significantly lower PostComp T than men who lost, $t(35)=-1.70$, $p=0.098$. However, men who won had a significantly smaller Pre–Post T than men who lost, $t(35)=-2.13$, $p=0.041$.

It is unlikely that this effect – losers showing a larger decrease in T than winners – is due simply to the higher baseline T in losers. Losers showed a larger decline in T as a percentage of their baseline T (11.83%) than winners (3.41%). In addition, losers showed a decrease in T over time, *paired* $t(9)=2.21$, $p=0.055$, that was nearly significant even with the small subsample, while there was no effect of time in winners that approached significance. Indeed, women showed a significant decrease over time, *paired t* $(37)=2.28$, $p=0.029$. So, all participants showed a significant (or nearly so) decrease in T over time except for male winners, further suggesting that winning attenuated a decrease in men's T.

We then conducted correlations between the T variables (Baseline, Postcomp, Pre–Post) and performance indicators to further explore possible associations (please see Table 2 for correlations).

3.4.2. Women

There were no significant differences (or notable trends toward differences) between winners and losers in Baseline T, PostComp T, or Pre–Post T. Please see Fig. 2.

Despite this lack of difference, there were significant correlations between the T variables and the performance indicators (please see Table 2).

4. Experiment 2

4.1. Goal

In experiment 2, our goal was to see if participants assigned to win (regardless of their ability) would show higher T parameters compared to participants assigned to lose (regardless of their ability). We also wanted to examine whether, in the present study, winners who would have won (Would-Be winners) in Experiment 1 (i.e. by their own ability) differed in their post-test T levels from winners who would have lost (Would-Be losers) in Experiment 1.

4.2. Method

4.2.1. Participants

Seventy-four participants were recruited from the Department of Psychology undergraduate participant pool at Simon Fraser University, and participated for course credit. There were 43 women (mean age=20.23 years, SD=3.88) and 31 men (mean age=20.13 years, SD=2.79). No participants reported using exogenous hormones, including hormonal contraceptives. According to the Kinsey scales of sexual experience and fantasy [27], all participants were heterosexual (scoring 0 or 1 on both questions), except for three who were bisexual (scoring 2, 3, or 4 on one or both questions), one who was gay/lesbian/queer (scoring 5 or 6 on both questions), and two who were nonsexual (scoring 0 on both questions). There was one person who omitted responses to the scales.

There were 38 participants with English as their first language and 36 participants whose first language was another language (Cantonese/Chinese, Farsi, German, Hindi, Indonesian, Korean, Kuchi, Mandarin, Polish, Punjabi, Romanian, Sinhalese, Tagalog, and Vietnamese). There were 55 participants with English as the language they felt most comfortable speaking, and 19 with other languages (Cantonese/Chinese, Farsi, Hindi, Korean, Mandarin, and Romanian).

4.3. Materials and procedures

We used the competition task described in Experiment 1, except we programmed the test such that participants either 'won' or 'lost' as predetermined before they arrived (we alternated winners and losers). The procedure was otherwise the exact same as in Experiment 1. Two participants' saliva samples

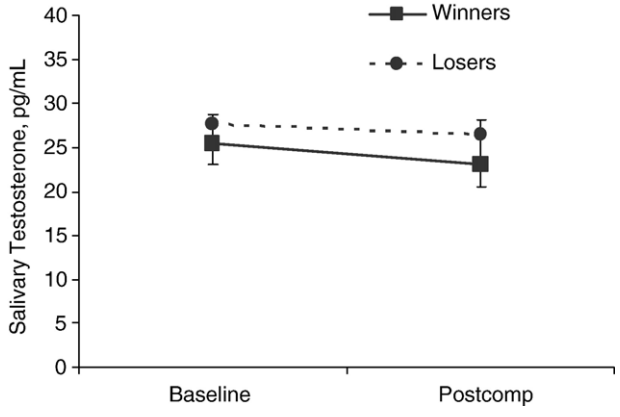

Fig. 2. Mean testosterone for women at baseline, post-competition, and the change from baseline to post-competition.

could not be analyzed, and four participants' competition parameters were not recorded due to computer difficulties.

5. Results

5.1. Sex similarities on competition parameters

There was no significant difference between women and men in the following, confirming that our task was not gender-specific: competition outcome (i.e. win/loss), $\chi^2(1)=0.11$, $p=0.816$, caring about winning, $t(72)=-.88$, $p=0.382$, feelings about the test (i.e. like/dislike), $t(72)=0.19$, $p=0.852$, level of challenge, $t(72)=-0.36$, $p=0.852$, factor involved in win (i.e. luck, ability), $t(72)=-0.27$, $p=0.782$, Total Questions, $t(68)=-0.85$, $p=0.400$, Total Correct, $t(68)=-0.01$, $p=0.989$, Total Incorrect, $t(68)=-1.40$, $p=0.166$, Percent Correct, $t(68)=0.35$, $p=0.728$, Average RT, $t(68)=0.70$, $p=0.486$, Average Correct RT, $t(68)=-0.46$, $p=0.648$, and Average Incorrect RT, $t(68)=0.52$, $p=0.608$. Women and men did not differ significantly in how competitive they reported being in competitions, $t(72)=-1.61$, $p=0.112$.

5.2. Competition outcomes, measures, language, and testosterone

There were no significant differences between winners and losers in any of the POMS subscales. In men, the correlation between Baseline T and competitiveness approached significance, $r(30)=0.30$, $p=0.091$, and there were significant correlations between believing that luck was the main factor in their performance and Baseline T $r(30)=0.45$, $p=0.009$, as well as PostComp T, $r(30)=0.45$, $p=0.009$. Also, PostComp T was significantly correlated with how much men cared about winning while doing the competition, $r(30)=0.36$, $p=0.043$. In women, there was a significant correlation between Pre–Post T and finding the task easy, $r(37)=0.33$, $p=0.042$. There were no other significant correlations between the T measures and the Post-Competition Questionnaire measures. There was no significant effect of English as a first language or as the most currently comfortable language on likelihood of winning or losing.

5.3. Task parameters

We checked performance of winners and losers, since they were randomly assigned and outcome was not ability-based. Winners and losers did not differ significantly on Total Incorrect, Correct RT, or Percent Correct. Losers had significantly higher Total Correct, $t(68)=-2.10$, $p=0.040$, and faster RT, $t(45)=2.41$, $p=0.020$, than winners. There was a trend for losers to have faster Incorrect RT, $t(41)=1.95$, $p=0.058$, and higher Total Questions, $t(68)=-1.86$, $p=0.067$, than winners. This did not disappear when we controlled for the effects of whether English was currently the most comfortable language.

5.4. Competition outcomes and T

As expected, there were large and significant differences between men and women in Baseline T, $t(33)=12.68$, $p<0.001$, and PostComp T, $t(32)=11.91$, $p<0.001$, so analyses on competition outcomes and T were conducted separately by sex.

5.4.1. Men

There were no significant differences by outcome (win/loss) in Baseline T, PostComp T, or Pre–Post T. Thus, there was no effect of randomly assigned competition outcome on T.

To see if T was correlated similarly with the performance indicators, we conducted the same set of correlations. There were trends for Baseline T to be correlated with Total Correct, $r(29)=-0.30$, $p=0.098$, Correct RT, $r(29)=0.31$, $p=0.092$, and Total Questions, $r(29)=-0.35$, $p=0.056$. There were trends for PostComp T to be correlated with Total Correct, $r(29)=-.34$, $p=0.065$, Incorrect RT, $r(29)=0.34$, $p=0.060$, and Percent Correct, $r(29)=-0.32$, $p=0.084$, and PostComp T was significantly correlated with Average RT, $r(29)=0.40$, $p=0.026$. Pre–Post T was significantly correlated with Incorrect RT, $r(29)=0.38$, $p=0.037$, and there was a trend for a correlation with Total Incorrect, $r(29)=0.33$, $p=0.070$, and Average RT, $r(29)=0.34$, $p=0.061$.

5.4.2. Women

There were no significant differences between winners and losers in Baseline T or PostComp T. There was a marginal trend for losers to have a greater Pre–Post T than winners, $t(37)=1.73$, $p=0.092$.

There were significant correlations between the T variables and the performance indicators. Baseline T was significantly correlated with Total Correct, $r(36)=-0.34$, $p=0.038$, Correct RT, $r(36)=0.42$, $p=0.010$, and Total Questions, $r(36)=-0.43$, $p=0.007$. There was a trend towards significant correlations between PostComp T and Total Questions, $r(36)=-0.28$, $p=0.095$. There were no significant correlations (or trends there towards) between Pre–Post T and the performance indicators.

5.5. Testosterone and experiment 1 outcome definitions

In Experiment 1, participants won or lost by responding correctly to 20 or more questions; in Experiment 2 participants were randomly assigned to win or lose. However, even in Experiment 2 we could differentiate between participants who would have won by their own ability (i.e. responded correctly to 20 or more questions) and those who would have lost by their own ability (i.e. did not respond correctly to at least 20 questions). Dividing participants this way provided Would-Be winners (seven men and 10 women) and Would-Be losers (24 men and 28 women). In men, the Would-Be losers had significantly higher Baseline T, $t(7)=2.63$, $p=0.036$, and PostComp T, $t(8)=2.45$, $p=0.043$, than the Would-Be winners, consistent with Experiment 1. There was no significant difference between men's Pre–Post T in the Would-Be winners and losers. In women, the Would-Be losers had significantly higher Baseline T than the Would-Be winners, $t(36)=3.89$, $p<0.001$, but there was no significant difference between the groups for PostComp T. However, women in the Would-Be losers had a significantly larger decrease (28.22%) in T from baseline to post-competition than women in the Would-Be winners (6.54%), $t(35)=-3.57$, $p=0.001$.

6. General discussion

The present study investigated the effects of a computer-based vocabulary competition on testosterone (T) levels in women and men, and we manipulated whether competition outcome was ability- or chance-determined. We found that men who lost the competition exhibited a larger decrease in T than men who won the competition, but only when they won or lost based on their own ability, suggesting that earned wins attenuate the T decrement associated with the task or with time. In addition, the effect on men's T could be directly attributed to a match between ability and outcome, because men who would have lost based on their ability but were assigned to win in the chance-determined paradigm (i.e. Experiment 2) did not show the same loss-mediated decrease in T as men who lost based on their own ability in Experiment 1. This suggests that good performance (i.e. ability) alone does not mediate the competition effect in men, nor does external recognition (i.e. win outcome) alone; instead, the competition effect relies on a good performance being externally recognized with an appropriate outcome. One limiting factor in this interpretation is the smaller sample size of would-be losers in Experiment 2 compared to losers in Experiment 1. The data from the women suggest a much less straightforward picture. In women, neither ability- nor chance-determined outcome led to changes in T; a mismatch between ability and outcome did. That is, women who would have lost based on their ability – but in the chance-based paradigm where outcome was assigned – had a larger decrease in T than women who would have won. However, women who lost based on their ability or were assigned to lose did not show this decrement (discussed more below).

Our results confirm the competition effect and parallel previous findings from sports paradigms [9,8,10], game paradigms [20,22], and a reaction time test [21]. While the mechanism for such an effect is unclear (even in non-human animals), the decrease from baseline to post-competition T among other groups except for male winners and female would-be winners suggests that the task induced smaller amounts of T to be released, but that winning produced larger pulses of T (or blocked the induction of smaller pulses), counteracting the tasks' reductive effects on T. The finding of an attenuated T decrease rather than an explicit increase begs future study. For example, our selection of 15–20 min post competition was based on previous findings of experimental effects on T at approximately 15 min [33]; it is possible that T measurements would be higher at earlier sampling points. Our findings extend the literature on the competition effect to show that a cognitively-demanding test also shows the same effect, and that, interestingly, participants do not need to be competing (or believe they are competing) against other individuals to show increased T parameters following winning relative to losing. Our study is the first to illustrate that the competition effect does not depend on direct competition, and begs the question of whether winning direct competitions versus individual competitions might show a differential effect on T in humans, since they both show higher T relative to losing. One possibility is that our finding of attenuated decreases in winner's T may be idiosyncratic to the solitary nature of our competition. Though an increase in T following a win makes sense in an evolutionary context, its functionality remains unknown. We expect that the increased T may provide small and brief increases in confidence, willingness to compete, attention, or mood. It may also stimulate transient increases in sperm production, presumably coordinated with increased sexual opportunities as gonadotropins stimulate T release as well as sperm production [34]. Alternatively (or complementarily), decreased T may increase behaviors aimed at promoting social bonds (i.e. bond-maintenance behaviors; 4). Our paradigm should prove useful in future studies of androgen–behavior associations, perhaps identifying which behaviors can be affected by transient fluctuations in androgens, by having participants engage in the behavior of interest after the test.

Ours is the first report to include both men and women on the same task and discover results that differ for women and men. As noted above, men's T was affected only by ability-determined outcomes, while women's T was affected only by a mismatch between ability and outcome. The results for women are difficult to interpret, as neither ability nor outcome alone was sufficient to induce differential effects on T. Instead, women who would have lost showed the same decrease in T as men who had ability-based losses. Previous studies have failed to find a competition effect in women (e.g. [15]), but since previous studies that have included both women and men have failed to demonstrate an effect in men, and many studies with men find no effect, it has been difficult to interpret the null finding in women. The results from our study suggest that there may be an effect in women, but one that does not mirror the effect in men and one that is complex and unclear as of yet. One possibility is that the effect in females is a non-functional byproduct of a functional male androgenic response (see [35] for related discussion). Another speculative possibility is an androgenic increase of adrenal origin in response to the stress of good performance gone unrecognized and mislabeled, since adrenal contributions are a large portion of overall androgens in women. Measuring frustration or stress, which we did not, might address this. This is contrast to evidence of the Challenge Hypothesis in women (i.e. that androgens are elevated with social challenge), from studies of female rugby players [15] and soccer players [16] where androgens increase over competition as they have been shown to do in men. One important point for interpreting the present difference in findings by sex is the presence of an opposite-sex experimenter for the men. All participants were tested by a woman and almost all were heterosexual; perhaps the presence of an opposite-sex tester affects the physiological response to winning as it does to brief conversations [36]. It is, however, unlikely that the complex finding in women is because our task was less relevant or important to women, or that women were less experienced with our task, since we found no sex differences in relevant task parameters or task performance. Our finding of no difference in overall competitiveness between women and men parallels previous findings [25].

Out study is also the first to test the effects of a chance-determined and ability-determined competition task on T. Whether a win must be earned or whether merely winning without effort can affect T has been a key unresolved issue in

research on the Competition Effect. Though the majority of past studies have used (usually implicitly) ability-determined paradigms, one study using a chance-determined paradigm reported increases in T for winners of a series of coin tosses [22] while another found no effect on T for winners of a laboratory lottery [10]. Our findings accord with Mazur and Lamb and suggest that, in men, the win must be earned. Our participants, however, attributed their outcome similarly in between luck and ability for both the ability-determined Experiment 1 and the chance-determined Experiment 2. Therefore, the question of how participants actually match the feel of earning their outcome with the actual outcome is open to question.

We also found that male winners and female would-be winners had lower T at baseline and post-competition, suggesting that higher T is associated with poorer performance on our competition task. For men, this deficit associated with higher T appears to be related to the verbal meaning component of the task, since T was negatively correlated with the number of questions answered correctly, but not associated with response time. Female winners did not show a difference in T at baseline or post-competition, but did show a negative correlation between T and the number of questions answered. Though not reported, additional analyses from previous studies with women [30] show a similar significant negative correlation between T and vocabulary test scores. In addition, higher T was associated with longer response times, suggesting that the poorer performance on the task associated with higher T was related to the response time component for women. Past studies have not noted this negative association between T and vocabulary scores within women or men, but pertinent studies have not reported within-sex correlational analyses [37,38]. One study [39] did find that gifted adolescents had lower T than their non-gifted counterparts, paralleling the finding from the present study since vocabulary tests are often used in cognitive studies as proxies for general intelligence. The (negative) association between T and vocabulary performance raises questions about the utility of vocabulary tests as hormone-neutral control tasks in studies of androgens and cognition.

Acknowledgements

We would like to thank Lisa Dawn Hamilton for help with data collection, and Jared van Snellenberg and Jennifer Whitman for help with programming. This research was supported by operating grant 0194522 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to NVW. SMvA was supported by an NSERC Post-Graduate Scholarship, an NSERC-UNESCO-L'Oreal For Women in Science Supplement, an IODE War Memorial Scholarship, and a C.D. Nelson scholarship and other awards from Simon Fraser University. We would like to acknowledge the Endocrine Core Lab at Yerkes National Primate Research Center, Emory University, for conducting the hormone assays.

References

[1] Bernstein IS, Rose RM, Gordon TP. Behavioral and environmental events influencing primate testosterone levels. J Hum Evol 1974;3:517–25.
[2] Setchell JM, Dixson AF. Changes in the secondary sexual adornments of male mandrills (*Mandrillus sphinx*) are associated with gain and loss of alpha status. Horm Behav 2001;39:177–84.
[3] Wingfield JC, Hegner RE, Dufty Jr AM, Ball GF. The "challenge hypothesis": theoretical implications for patterns of testosterone secretions, mating systems, and breeding strategies. Am Nat 1990;136:829–46.
[4] van Anders, SM, Watson, NV. Social neuroendocrinology: Effects of states and behaviours on human endocrine function. Hum Nat in press.
[5] Archer J. Testosterone and human aggression: an evaluation of the Challenge Hypothesis. Neurosci Biobehav Rev 2006;30:319–45.
[6] Salvador A. Coping with competitive situations in humans. Neurosci Biobehav Rev 2005;19:195–205.
[7] Mazur A, Booth A. Testosterone and dominance in men. Behav Brain Sci 1997;21:353–97.
[8] Elias M. Serum cortisol, testosterone, and testosterone-binding globulin responses to competitive fighting in human males. Aggress Behav 1981;7:215–24.
[9] Booth A, Shelley G, Mazur A, Tharp G, Kittock R. Testosterone and winning and losing in human competition. Horm Behav 1989;23:556–71.
[10] Mazur A, Lamb TA. Testosterone, status, and mood in human males. Horm Behav 1980;14:236–46.
[11] Gonzalez-Bono E, Salvador A, Serrano MA, Ricarte J. Testosterone, cortisol, and mood in a sports team competition. Horm Behav 1999;35:55–62.
[12] Filaire E, Maso F, Sagnol M, Ferrand C, Lac G. Anxiety, hormonal responses, and coping during a judo competition. Aggress Behav 2001;27:55–63.
[13] Salvador A, Simon V, Suay F, Llorens L. Testosterone and cortisol responses to competitive fighting in human males: a pilot study. Aggress Behav 1987;13:9–13.
[14] Suay F, Salvador A, Gonzalez-Bono E, Sanchis C, Martinez M, Martinez-Sanchis S, et al. Effects of competition and its outcome on serum testosterone, cortisol, and prolactin. Psychoneuroendocrinology 1999;24:551–66.
[15] Bateup HS, Booth A, Shirtcliff EA, Granger DA. Testosterone, cortisol, and women's competition. Evol Hum Behav 2002;23:181–92.
[16] Edwards DA, Wetzel K, Wyner DR. Intercollegiate soccer: saliva cortisol and testosterone are elevated during competition, and testosterone is related to status and social connectedness with teammates. Physiol Behav 2006;87:135–43.
[17] Linnamo V, Pakarinen A, Komi PV, Kraemer WJ, Hakkinen K. Acute hormonal responses to submaximal and maximal heavy resistance and explosive exercise in men and women. J Strength Cond Res 2005;19:566–71.
[18] Neave N, Wolfson S. Testosterone, territoriality, and the 'home advantage'. Physiol Behav 2003;78:269–75.
[19] Mazur A, Susman EJ, Edelbrock S. Sex difference in testosterone response to a video game contest. Evol Hum Behav 1997;18:317–27.
[20] Mazur A, Booth A, Dabbs Jr JM. Testosterone and chess competition. Soc Psychol Q 1992;55:70–7.
[21] Gladue BA, Coechler M, McCaul KD. Hormonal response to competition in human males. Aggress Behav 1989;15:409–22.
[22] McCaul KD, Gladue BA, Joppa M. Winning, losing, mood, and testosterone. Horm Behav 1992;26:486–504.
[23] Bernhardt PC, Dabbs Jr JM, Fielden JA, Lutter CD. Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. Physiol Behav 1998;65:59–62.
[24] Oliveira RF, Lopes M, Carneiro LA, Canario AV. Watching fights raises fish hormone levels. Nature 2001;409:475.
[25] Cashdan E. Are men more competitive than women? Br J Soc Psychol 1998;37:213–29.
[26] Cashdan E. Hormones and competitive aggression in women. Aggress Behav 2003;29:107–15.
[27] Kinsey AC, Pomeroy WB, Martin CE. Sexual Behavior in the Human Male. Philadelphia: W.B. Saunders Co; 1948.
[28] Thurstone TG. Primary Mental Abilities for Grades 9–12. Chicago: Science Research; 1962.
[29] NeuroBehavioral Systems (www.neurobs.com). Presentation, Version 0.71 (Computer software). Albany, CA; 2004.
[30] van Anders SM, Hampson E. Testing the prenatal androgen hypothesis: measuring digit ratios, sexual orientation, and spatial abilities in adults. Horm Behav 2005;47:92–8.

[31] McNair DM, Lorr M, Droppleman LF. Profile of Mood States. San Diego: EdITS; 1971.
[32] Kinsey AC, Pomeroy WB, Martin CE. Sexual Behavior in the Human Male. Philadelphia: WB Saunders; 1948.
[33] Schultheiss OC, Wirth MM, Torges CM, Pang JS, Villacorta MA, Welsh KM. Effects of implicit power motivation on men's and women's implicit learning and testosterone changes after social victory or defeat. J Pers Soc Psychol 2005;88:174–88.
[34] Nelson RJ. An Introduction to Behavioral Endocrinology. 2nd Ed. Sunderland Mass: Sinauer Associates, Inc; 2000.
[35] Ketterson ED, Nolan Jr V, Sandell M. Testosterone in females: mediator of adaptive traits, constraint on sexual dimorphism, or both? Am Nat 2005;166: S85–98.
[36] Roney JR, Mahler SV, Maestripieri D. Behavioral and hormonal responses of men to brief interactions with women. Evol Hum Behav 2003;24: 365–75.
[37] Davison KK, Susman EJ. Are hormone levels and cognitive ability related during early adolescence? Int J Behav Dev 2001;25:416–28.
[38] Gouchie C, Kimura D. The relationship between testosterone levels and cognitive ability patterns. Psychoneuroendocrinology 1991;16:323–34.
[39] Ostatkinova D, Laznibatova J, Putz Z, Mataseje A, Dohnanyiova M, Pastor K. Salivary testosterone levels in intellectually gifted and non-intellectually gifted preadolescents: an exploratory study. High Abil Stud 2000;11:41–54.