FINAL REPORT FOR

IFREE Foundation, Small Grants in Experimental Economics

Testing the causal influence of testosterone on rent- seeking and preferences for competition

Colin Camerer, Amos Nadler, Gideon Nave

California Institute of Technology 10.30.15

Preface

The current report to IFREE small grant program summarizes the research procedure and key findings of the project **"Testing the causal influence of testosterone on rent- seeking and preferences for competition"**.

The project is currently in advanced write-up stage, in preparation for submission to **Proceedings of the National Academy of Science**.

Introduction

Competition advances productivity, efficiency, and innovation, all of which create wealth and promote economic growth (Smith & Nicholson, 1887; Snow, 2002; Wennekers & Thurik, 1999). Over the past decade, lab and field studies have demonstrated substantial gender differences in risk taking, willingness to compete and performance amid competition (Buser, Niederle, & Oosterbeek, 2012; Charness & Gneezy, 2007, 2012; Croson & Gneezy, 2009; Flory, Leibbrandt, & List, 2010; Gneezy, Leonard, & List, 2006; Gneezy, Niederle, & Rustichini, 2003; Gneezy & Rustichini, 2004; Healy & Pate, 2011; Niederle & Vesterlund, 2010, 2011; Price, 2008; Sutter & Rützler, 2010). Despite the rapidly growing literature, little is known about the roots of gender differences in risk taking and competitiveness, and only recently have scholars started investigating its origins, and whether they arise from cultural (Gneezy, Leonard, & List, 2009), educational (Dreber, Von Essen, & Ranehill, 2011; Gneezy & Rustichini, 2004) or biological factors (Buser, 2012; Coates, Gurnell, & Rustichini, 2009; Coates & Herbert, 2008; Sapienza, Zingales, & Maestripieri, 2009).

Hormones and other biological factors have been shown to systematically influence decision-making and generate heterogeneity in economic behavior (Benjamin et al., 2012; Camerer, Loewenstein, & Prelec, 2004; Crockett & Fehr, 2013; Eisenegger, Naef, Snozzi, Heinrichs, & Fehr, 2010; Fehr & Camerer, 2007; Glimcher & Fehr, 2013; Kandasamy et al., 2014; Rangel, Camerer, & Montague, 2008). In particular, the androgen (sex steroid) hormone testosterone (abbreviated T) is the main driver of physiological dissimilarities between males and females, and a likely biological mechanism underlying many behavioral and cognitive gender differences (Miller & Halpern, 2014).

T is released to the blood stream and different brain regions in response to external stimuli, affecting physiological function and local information processing in cerebral structures context sensitively. Based on a well-established animal literature, competitiveness among the human behaviors that are most likely linked to T. A well-known theory, "The Challenge Hypothesis" (Archer, 2006; Wingfield, Hegner, Dufty Jr, & Ball, 1990)

postulates that before and during intra-male competitions over resources (e.g., territory) and mates amid breeding season, T levels rise and facilitate inter-male aggression. Accordingly, administering T to several male mammalian and bird species increased their mating and aggressive behavior in the lab.

Correlational studies in humans have previously linked human T levels (both in males and females) to risk taking, career choice and competitiveness (Apicella et al., 2008; Apicella, Dreber, & Mollerstrom, 2014; Booth, Shelley, Mazur, Tharp, & Kittok, 1989; Carré & McCormick, 2008; Carré, Putnam, & McCormick, 2009; Coates et al., 2009; Coates & Herbert, 2008; Mehta & Josephs, 2006, 2010; Mehta, Wuehrmann, & Josephs, 2009; Sapienza et al., 2009; Stanton et al., 2011). However, the evidence are mixed, and other studies failed to find substantial effects (Apicella et al., 2011; Zethraeus et al., 2009). T levels were also shown to be affected by competition outcomes: increased T levels were found in male wrestlers, tennis and chess players, trivia challengers and football fans after winning competitions (Bernhardt, Dabbs Jr, Fielden, & Lutter, 1998; Booth et al., 1989; Carré et al., 2009; Mazur, Booth, & Dabbs Jr, 1992; McCaul, Gladue, & Joppa, 1992), but the evidence are inconclusive: two recent lab studies found no differences in T levels between winners or losers, although the changes in T as a response to losing did predict subsequent behaviors, such as willingness to compete again (Apicella et al., 2014; Mehta & Josephs, 2006)

T levels have a daily (diurnal) cycle, decline secularly with age, and may vary by other demographic factors (e.g. profession). In addition, certain behaviors and environments might induce changes in T levels (rather than vice versa) and T interacts with other hormone and neurotransmitter systems in the brain (e.g., dopamine and serotonin). Thus, it is still an open question whether exists a causal link between T levels and competitive behavior, and whether T underlies gender differences in competitiveness. Several challenges accompany studies that rely on measures of endogenous T changes caused by indirect experimental treatments (e.g., winning or losing in a competition): first, hormone release is inherently noisy and subject to heterogeneity, thereby yielding unreliable control and identification. Second, one cannot establish claims

about causality using such manipulations as T reactivity in response to a short term behavioral manipulation rather than the rise in T levels per se, could be a correlate of the investigated behavior. Last, such treatments might influence various factors besides T levels.

The current study overcomes the above-mentioned methodological challenges by directly manipulating T pharmacologically, in a double blind, placebo-controlled protocol and random assignment of treatment groups. Previous lab studies have linked T administration with several economic behaviors, such as cooperation (van Honk, Montoya, Bos, van Vugt, & Terburg, 2012; Wright et al., 2012), honesty (Wibral, Dohmen, Klingmüller, Weber, & Falk, 2012), reduced trust (Boksem et al., 2013; Bos, Hermans, Ramsey, & Van Honk, 2012; Johnson & Breedlove, 2010), generous ultimatum game offers (Eisenegger et al., 2010; Zak et al., 2009) and rejection of low ultimatum offers (Burnham, 2007; Mehta & Beer, 2010). The generalized theory, which aims to explain the seemingly contradictory findings, is that T increases behaviors that are intended to obtain or maintain high social status (Eisenegger, Haushofer, & Fehr, 2011; Eisenegger, Naef, Snozzi, Heinrichs, & Fehr, 2012).

The current project aims to understand whether, and how, T has a causal influence on willingness to compete in males. Using an especially large sample, together with extended monitoring of a wide hormonal panel dataset - we test T as plausible biological mechanism generating gender differences in human competitive behavior.

Materials and methods

Subjects

There were n=243 male-only participants. Most (217, 89%) were students from a southern Californian college. Non-student participants were community members from surrounding cities. n=125 of subjects were randomly assigned to receive a standard dose of T and n=118 received placebos in a double blind exogenous administration paradigm.

Pre-screening criteria excluded everyone with relevant medical and psychological conditions (Klinefelter's syndrome, brain tumor, cancer,

psychiatric diagnosis/diagnoses, high blood pressure, liver disease, kidney disease, angina, cancer, hepatitis, renal/kidney impairment, history of epileptic seizures and hypersensitivity to soy/ alcohol), subjects using prescription drugs that may interfere with the study (oxyphenbutazone, insulin, corticosteroids, opioids), subjects who self reported consuming illegal drugs or alcohol in excess in the last 24 hours and non-native English speakers.

Personal, demographic, and treatment expectancy characteristics of the two treatment groups are summarized in Table 1 (note that 5 subjects did not report their age and were therefore excluded from all analyses in which age is used as a control variable). The right column of Table 1 also reports the p-value of two sample t-tests for differences between T and placebo group characteristics (a check on whether random assignment resulted in balance on all such variables). Two subjects (one from each treatment group) self reported taking T treatment on a regular basis; all analyses include these subjects and are robust to excluding them.
In order to reduce the potential effect of a female experimenter's presence on T-related behaviors, male researchers conducted all of the experimental sessions.

Table 1: Self-reported demographic data summary (standard errors in parentheses)

	All	T	Placebo	p-values for t-test of difference
N	243	118	125	
Age	23.63	24.42	22.78	0.08
	(0.46)	(0.77)	(0.49)	
Left-handed (proportion)	0.074	0.064	0.085	0.54
	(0.02)	(0.02)	(0.03)	
Heterosexual (proportion)	0.90	0.91	0.89	0.56
	(0.02)	(0.03)	(0.03)	
Treatment expectancy[1]	2.76	2.67	2.85	0.16
	(0.06)	(0.08)	(0.09)	
Married (proportion)	0.08	0.09	0.08	0.74
	(0.02)	(0.03)	(0.03)	
In a relationship (proportion)	0.38	0.34	0.42	0.20
	(0.03)	(0.05)	(0.04)	
Has children	0.06	0.08	0.04	0.23
	(0.02)	(0.02)	(0.02)	
Personal monthly income[2]	2.05	2.02	2.07	0.84
	(0.11)	(0.14)	(0.16)	

1 5 point scale, 1=definitely did not get testosterone, 2= 3=unsure 5= definitely got testosterone

2 5 point scale, 1=under $500/month, 2=$501-$1,000, 3= $1,001-$1,500/month, 4=$1,501-$2,000/month 5= over $2001/month

Experimental procedure

The timeline of the experimental procedure is summarized in Figure 1. Subjects first arrived at the lab at 9:00am in the morning of their experimental session. They signed an informed consent form and then proceeded to a designated room where their hands were scanned (to obtain digit ratio measurement, a possible proxy of prenatal T). Then, subjects were randomly assigned to private cubicles where they completed demographic and mood questionnaires. After all subjects completed those questionnaires, they provided an initial baseline saliva sample. After saliva sample collection, subjects proceeded to the gel administration room for T or placebo gel application. Participants were given a small plastic container and told that it contained either T gel or an inert substance and that the likelihood of receiving either was 50%. In the process of signing consent forms participants were explicitly told to have no skin contact with females, avoid vigorous exercise and showering, finish eating their lunch no later than 1:00pm, and return to the lab at 2:00pm well hydrated. These instructions and precautions were repeated again before they departed from the morning session. After gel application, subjects were instructed to thoroughly wash their hands in the adjacent bathroom with warm water and soap, avoid touching any part of their body until their hands were washed, and were given printed precautions prior to dismissal. All subjects returned to the lab at 2:00pm (with no incidents of lateness). Then they provided a second saliva sample and proceeded to the behavioral experiment. The time frame between gel application and behavioral experiment was chosen so that tasks took place when the T group subjects experienced elevated and stable blood T levels following drug administration (Eisenegger, von Eckardstein, Fehr, & von Eckardstein, 2013).

The experiment consisted of a battery of seven behavioral tasks. None of the tasks included feedback about the subjects' monetary payoffs (to avoid changes in T from changes in payoff). Only the final task included feedback regarding the subjects' performance relative to other participants (also to avoid outcome-related changes in T).

The rationale for conducting a battery of tasks (compared to a single experiment) is maximizing the knowledge gained from each human subject undergoing a pharmacological manipulation, a practice which is standard (Kocoska-Maras et al., 2011; Zethraeus et al., 2009) and looked

favorably upon by IRB review. Accordingly, we ensured that statistical tests for the CRT task alone survived correction for multiple comparisons (choosing only CRT out of the seven tasks for analysis) to avoid increased type-I error rate from multiple comparisons.

To maintain high-resolution monitoring of hormonal changes during the experiment and control for their influences throughout the study, a total of four saliva samples were collected throughout the experiment (details of collection frequency and time are in section 4 below). The accuracy and consistency of sampling times is crucial because the measured hormones have unique diurnal cycles which complicates comparing samples taken at different times of day. In order to standardize hormonal measurements among all subjects, we did not randomize the order of the behavioral tasks, in similar fashion to previous studies (Kocoska-Maras et al., 2011; Zethraeus et al., 2009). The behavioral battery lasted approximately two hours. The behavioral tasks reported here occurred between the second and third saliva samples. Following the experiment, subjects completed an exit survey. They indicated their expectancies about which of the two treatments they had received (see Table 1), and were privately paid in cash according to their performance.

Treatment administration

After providing the first saliva sample, participants were escorted in groups of 2-6 to a semi-private room where a research assistant provided a small plastic cup containing clear gel and stated it was equally likely to contain T or placebo (the cups were filled in advance by the lab manager, who did not interact with subjects and did not reveal the contents of the cup to the research assistant, so that the treatment was double-blind between assistant and subject). These cups contained either 10g of topical T 1% (2 x 50 mg packets Vogelxo® by Upsher-Smith) or volume equivalent of an inert placebo (80% alcogel, 20% Versagel®).

We chose to administer T using topical gel, as this is the only T administration method for which the pharmacokinetics of a single dose administration (i.e., time-course of post-treatment T levels change) has been investigated in healthy young men (S1). The single-dose study demonstrated that plasma T levels peaked 3 hours following exogenous topical administration, and that T measurements stabilized on high levels during the time window between 4 and 7 hours following administration.

Therefore we had all subjects return to the lab 4.5 hours after receiving gel, when the androgen levels are higher and stable.
Subjects were instructed to remove upper body clothing and apply the entire contents of the gel container to their shoulders, upper arms, and chest as demonstrated by the research assistant. During application they were told to wait until the gel fully dried before putting clothes back on, refrain from bathing, or any activity that might cause excessive perspiration before the afternoon session, finish eating no later than 1:00pm and return to the lab promptly at 1:55pm.
After self-administering the gel under the supervision of the experimenter, participants were instructed to thoroughly wash their hands with warm water and soap, avoid touching any part of their body before thorough washing and abstain from all skin-to-skin contact with females, as recommended by the gel manufacturers. All surfaces in the administration room were covered with medical grade isolation sheets and surfaces in the gel application area were cleaned with alcohol swabs after each experimental session. The adjacent bathroom where the sink was located was also thoroughly wiped, as were doorknobs and handles. These procedures are necessary because testosterone can be transmitted through physical contact with the gel.

Saliva samples

Each subject provided four saliva samples at predetermined sampling times throughout the study (see Figure 1): (1) Before treatment administration (all samples took place between 9:25 and 9:34 am) (2) upon return to the lab, just prior to starting the behavioral tasks (all samples took place between 1:55 and 2:15 pm); (3) in the middle of the behavioral tasks battery (between 3:02 and 3:38 pm) (4) a final sample following the one and only task involving performance feedback at the end of the experiment (between 4:10 and 4:44 pm). We chose to use saliva samples to avoid potential stress that might be induced by high-resolution blood drawing throughout the experimental session. Each saliva sample was time stamped. Fourteen (14) hormone measures were acquired using liquid chromatography tandem mass spectrometry (LC-MS/MS). No food or drinks were allowed into the laboratory, and the only water given to the participants was after their 3rd saliva draw (an hour before the 4th and final saliva draw).

Hormonal assay procedure

Salivary steroids (estrone, estradiol, estriol, testosterone, androstenedione, DHEA, 5-alpha DHT, progesterone, 17OH-progesterone, 11-deoxycortisol, cortisol, cortisone, and corticosterone) were measured by LC-MS/MS using an AB Sciex Triple Quad 5500. Internal standards were added to 1 ml of saliva and the steroids then extracted by C18 column chromatography with 0.1 M NH4OH wash followed by 10% acetone. Steroids were eluted from the SPE with 10% methanol in acetone and dried under nitrogen. The dried samples were subjected to derivatization—the process of transforming a compound into a derivative product of similar chemical structure—with pyridine-3-sulfonyl chloride for the estrogens (estrone (E1), estradiol (E2), and estriol (E3)) as outlined by Xi and Spink (2008). 40 μL sodium bicarbonate (50mM, pH 10) and 40 μL pyridine-3-sulfonyl chloride (3 mg/mL in acetonitrile) were added to the dried samples, and incubated at 60oC for 10 minutes. After derivatization, the samples were diluted with 80 μL of water and injected for LC-MS/MS analysis with analytical separation performed on an Agilent Poroshell 120 EC-C8 column and ionization by atmospheric pressure chemical ionization (APCI) in the positive ionization mode. Table S2 lists each analyte along with its validation results for the lower limit of quantitation (LLOQ is jargon for the lowest level of detection with coefficients of variation (CVs) < 20% over the linear range), linear range, and the inter-assay precision from the highest concentration to the LLOQ within the linear range. When salivary hormone levels of participants were below their LLOQ, we assigned values halfway between zero and their respective LLOQ (note that the true quantities of the hormone in the sample are never zero, even when they do not reach the detection threshold)

A series of one-sample Kolmogorov-Smirnov tests for conformity to Gaussian (Table S2) indicated that all hormonal measurement distributions were better approximated by a Gaussian following a log-transformation, as indicated by higher p-values (i.e., the Gaussian normality hypotheses were less likely to be rejected after log-transformations). Thus, all hormonal measurements were log-transformed prior to data analysis (e.g., fitting regression models) in order to make their distributions closer to Gaussian.

Table 2: Detection levels, precision and normality tests of hormonal assay

Analyte	LLOQ	Range	Precision	Proportion undetected, pre-treatment sample A	Proportion undetected, first post-treatment sample B	K-S test p-value	K-S test (log) p-value
Estrone pg/mL	0.5	0.5 - 510	8.7 - 13.7%	0.132	0.257	<0.01	0.56
Estradiol pg/mL	0.3	0.3 - 510	4.3 - 18.7%	0.128	0.329	0.06	0.88
Testosterone pg/mL	3.0	3.0 - 5100	3.0 - 18.1%	0	0.008	$<10^{-20}$	<0.01
Androstenedione pg/mL	5.0	5.0 - 2300	5.2 - 6.6%	0	0.008	$<10^{-20}$	0.008
DHEA pg/mL	20.0	20.0 - 1800	4.1 - 15.2%	0.004	0.012	0.002	0.98
DHT pg/mL	10.0	10.0 - 920	3.6 - 17.7%	0.786	0.473	$<10^{-11}$	0.02
Progesterone pg/mL	10.0	10.0 - 10000	4.8 - 10.8%	0.794	0.753	<0.01	0.03
17OH-Progesterone pg/mL	5.0	5.0 - 630	3.9 - 13.8%	0.004	0.061	0.003	0.98
11-Deoxycortisol pg/mL	5.0	5.0 - 410	6.8 - 16.6%	0.132	0.473	<0.01	0.04
Cortisol ng/mL	0.1	0.1 - 52	5.1 - 17.9%	0	0.008	<0.01	0.92
Cortisone ng/mL	0.1	0.1 - 81	4.1 - 14.9%	0	0.008	0.07	0.59
Corticosterone pg/mL	5.0	5.0 - 1800	4.6 - 17.5%	0.313	0.312	<0.01	0.08
Aldosterone pg/mL	10.0	10.0 - 560	8.9 - 18.8%	0.272	0.272	<0.06	0.39
Melatonin pg/mL	2.5	2.5-10000	5.2 - 15.9%	0.502	0.500	0.07	0.14

Note: P-values are calculated using a Kolmogorov-Smirnov test for the distributions of the second saliva sample compared to Gaussian, and for the log-transform (the null hypothesis is normal Gaussian distribution).

Hormonal changes following treatment and manipulation check

As expected, we found significant post-treatment differences between groups with respect to all hormones influenced by T treatment (see Figure 1), either as an upstream (androstenedione) or downstream (5-alpha DHT) metabolite of T (Horton & Tait, 1966). Additionally, we found a decrease in progesterone 170H resulting from an increase in T (which is common, according to personal communication from ZRT chief scientist Dr. David Zava). The changes in saliva T measures were similar in magnitude to those reported in previous studies following topical gel administration of T and progesterone, e.g. (Du et al., 2013; Mayo, Macintyre, Wallace, & Ahmed, 2004).

We observed no significant differences between treatment groups in hormones that were not expected to change following short-term T treatment (e.g., aldosterone, cortisol, cortisone, melatonin) in all four

saliva measurements throughout the experiment (i.e., the pre-treatment and the three post-treatment measurements). The pre-treatment and first post-treatment mean hormonal saliva levels are summarized in Table 3; note that differences between morning and afternoon hormonal levels were affected by diurnal cycles in both treatment groups.

From assays conducted during the first 13 sessions of the study, we identified that 72 out of 184 pre-treatment baseline saliva samples (in both treatment groups) presented measurements with higher T level that are expected in normal young men (greater than 400 pg/mL), while all other measurements (including T metabolites) were hormonally typical. We traced the cause of these abnormal levels to T gel spread to common surfaces (e.g., door knobs, mouse pads).

Crucially, we concluded that the high values were caused by local contamination of saliva tubes, but physiological levels were unaffected by superficial contact with the dry nuisance T gel. We reached this conclusion because: (a) we observed normal pre-treatment levels of T metabolites, namely DHT and androstenedione in all subjects; (b) none of the placebo group participants showed abnormally high values of T metabolites in the post-treatment measurements; (c) Only five out of 118 subjects from the placebo group showed consistent elevated T levels in all of the 3 post-treatment saliva measurements; (d) a previous investigations found that even after skin-to-skin contact, interpersonal T transfer is highly unlikely, where local contamination of a tube indeed might occur (Rolf, Knie, Lemmnitz, & Nieschlag, 2002). Thus, we conclude that biofluid levels, which could influence subjects' behavior, were unaffected by superficial contact. The results are robust (in terms of both the effect size and significance level) to inclusion/ exclusion of the five placebo subjects with post-treatment contaminated samples.

In response to this finding, we modified sterile isolation protocol to reduce the spread of the dried T gel by including thorough cleaning of keyboards, computer mice, chair backs, displays, and all doorknobs with a bleach-alcohol solution after each session as well as asking subjects to carefully wipe hands with a wet tissue before collecting each saliva sample. New pens were used for each session while old and possibly contaminated pens were removed from testing area. Clipboards and other miscellaneous objects that participants did or could interact with were cleaned, and an aerosol "air sanitizer" that bonds to VOCs (volatile organic compounds) was sprayed into the air. Following the adoption of this strict sterilization

protocol, we found a drastic reduction in incidence of high T samples in the pre-treatment measurements, to a total of 5 participants out of 58 in the following four sessions (sessions 14-17).

Figure 1: Experiment timeline and salivary testosterone levels

Subjects arrived at the lab at 9 am, had their hands scanned, filled an intake survey and gave a baseline saliva sample "A" before application of either testosterone or placebo topical gel. After a four-hour loading period, subjects came back to the lab and took part in a battery of behavioral tasks. Three additional saliva samples ("B", "C" and "D") were collected during the experiment, all of which indicated elevated T levels in the treatment group compared to placebo. The behavioral tasks (CRT and Math) took place between saliva sample B and C.

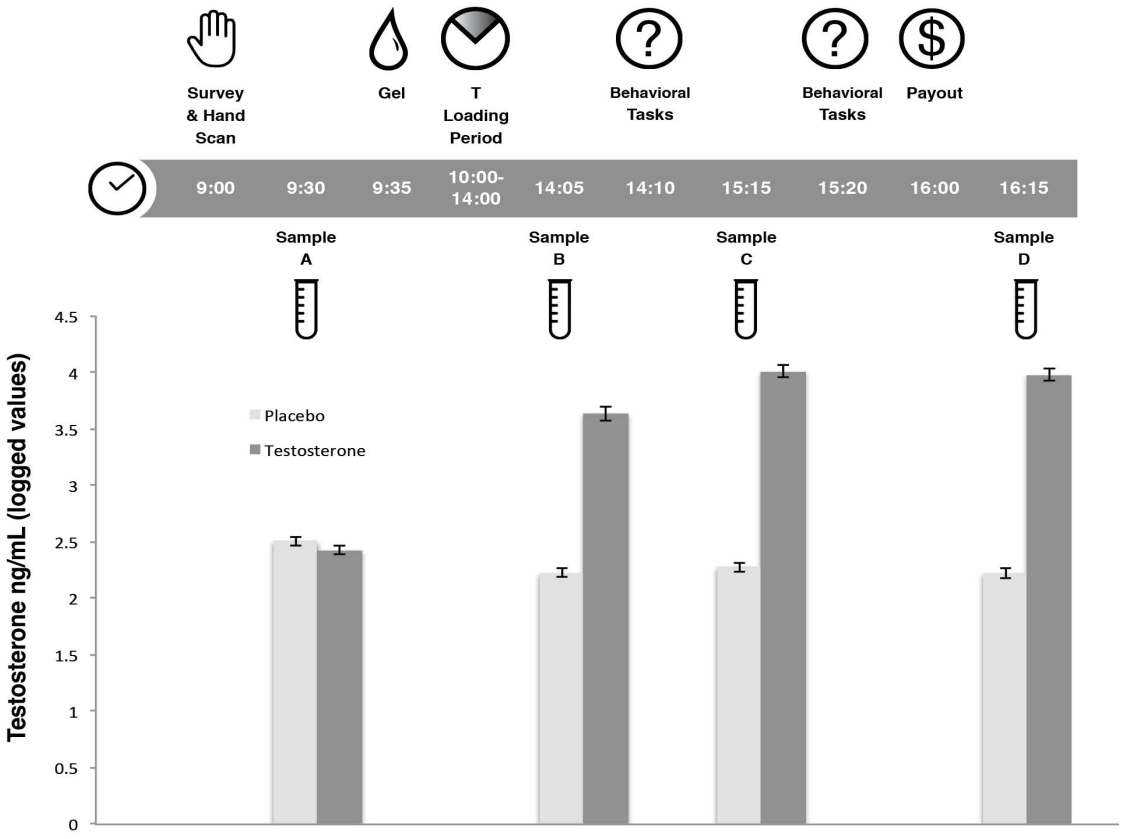

Table 3: hormone panel data measurements (pg/mL) summary statistics (standard errors in parentheses)

	T		Placebo		Two-tailed p-value from t-test of T-Placebo equality	
Sampling time[1]	**9:30**	**2:00**	**9:30**	**2:00**	**9:30**	**2:00**
Cortisol	3.44	1.32	3.61	1.35	0.68	0.81
	(0.20)	(0.08)	(0.36)	(0.09)		
Cortisone	13.54	8.17	13.32	8.05	0.70	0.76
	(0.44)	(0.25)	(0.37)	(0.30)		
Corticosterone	34.75	5.04	28.87	5.33	0.37	0.76
	(5.02)	(0.59)	(4.04)	(0.78)		
DHEA	194.43	113.23	207.31	116.91	0.35	0.67
	(9.30)	(5.78)	(10.31)	(6.39)		
Estrone	1.27	0.81	1.14	0.75	0.29	0.42
	(0.09)	(0.04)	(0.08)	(0.06)		
Estradiol	0.55	0.41	0.55	0.38	0.86	0.44
	(0.03)	(0.02)	(0.03)	(0.02)		
Progesterone	8.16	9.12	9.65	10.91	0.36	0.41
	(0.88)	(1.32)	(1.40)	(1.73)		
Deoxycortisol11	19.28	6.48	18.41	7.24	0.66	0.35
	(1.55)	(0.52)	(1.25)	(0.62)		
Aldosterone	18.29	14.29	17.92	16.41	0.82	0.14
	(1.20)	(0.93)	(1.13)	(1.10)		
Melatonin	14.80	2.46	4.96	1.50	0.27	0.23
	(8.58)	(0.76)	(0.56)	(0.18)		
DHEA7	85.58	75.18	82.32	74.02	0.69	0.84
	(5.94)	(4.13)	(5.55)	(4.25)		
Testosterone	480.13	11433.73	614.46	250.78	0.27	0.00
	(73.96)	(1368.32)	(96.53)	(25.19)		
Androstenedione	98.82	381.31	101.57	73.73	0.63	0.00
	(3.29)	(48.19)	(4.84)	(2.67)		
DHT	10.58	79.14	11.59	8.16	0.75	0.00
	(2.27)	(14.24)	(2.11)	(0.84)		
Progesterone 170H	29.24	14.11	29.93	17.15	0.79	0.01
	(1.93)	(0.75)	(1.77)	(0.87)		

1 Main effects of time (afternoon vs. treatment) are due to the diurnal cycles of the hormones (Hurwitz, Cohen, & Williams, 2004)

Digit ratio measurements

The ratio of second (index) finger length to fourth (ring) finger (abbreviated 2D:4D) is considered a proxy for pre-natal T exposure (Manning, 2002), although this hypothesis is still under debate (e.g., (Voracek, 2014)). Subjects' 2D:4D ratios were measured by two independent raters using hand scans and digital calipers (correlation between the two raters was ~.95). The right hand digit ratio was not calculated for one subject due to a broken finger, and therefore he was excluded from all analyses that use the right hand digit ratio as control. Correlation between the digit ratios of the left and right hands was 0.64, p=0.0001.

Mood questionnaire

We measured mood using the PANAS-X scale (Watson & Clark, 1999), both pre-treatment (in the morning) and post-treatment (in the afternoon). Table 4 shows a modest decrease in both affect measures over time (morning vs. afternoon), and no treatment or time x treatment interaction, indicated by the output of 2-way analysis of variance (ANOVA) with an interaction term, ruling out this indirect way in which T might affect cognition and behavior. Three subjects did not answer all of the negative affect items in their questionnaires, and five subjects did not complete all of the positive affect items; these subjects were excluded from analyses that include these scales as control variables.

Table 4: Positive / negative affect (PANAS-X) summary statistics

	All		Testosterone		Placebo		ANOVA: p-values		
Time	**Morning**	**Afternoon**	**Morning**	**Afternoon**	**Morning**	**Afternoon**	**T**	**Time**	**T x time**
Positive affect	2.72	2.61	2.72	2.63	2.72	2.60	0.85	0.16	0.85
	(0.05)	(0.06)	(0.06)	(0.08)	(0.07)	(0.09)			
Negative affect	1.53	1.45	1.53	1.46	1.53	1.43	0.77	0.13	0.84
	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)	(0.05)			

Treatment expectancy

One previous study indicated an effect of subjects' beliefs about the treatment they had received on behavior (Eisenegger et al., 2010). We therefore asked subjects to indicate their expectancy about whether they had received placebo or T using a 5-point scale. There were no significant

differences between the groups on this expectancy measure (see table S1). Two subjects did not report their treatment expectancy and therefore were excluded from all analyses in which this measure was used as a control.

Behavioral tasks

Competition

To test for causal influence of T on human competitiveness, we measured T's effects on competitiveness and performance amid competition using an established experimental paradigm. The task was shown to generate robust gender differences in competition entry (Niederle & Vesterlund, 2005, 2010, 2011) and is associated with real world gender differences in career choice (Buser, 2012).

The experimental task was programmed in z-Tree (Fischbacher, 2007) and consisted of four rounds. The basic task in each round required subjects to correctly add up as many sets of five two-digit numbers as possible in 5 minutes. Participants were not allowed to use a calculator but could use pen and paper. The numbers in each problem were randomly drawn and presented in the following way (such that participants filled in the sum in the blank box):

21	35	48	29	83	

Once a participant had submitted an answer, a new problem appeared. Crucially, participants were not be given feedback on their performance relative to others and payoffs until the end of the experimental session.

Each session consisted of five rounds, each of which was a variant of the basic task that differed with respect to its payoff scheme. Subjects were informed about the payoff schemes only immediately before each round, and at the end of the experiment were paid according to their performance in one randomly chosen round.

Round 1: Piece rate. Subjects were rewarded $1 for each correct answer. This round always took place first in order to estimate subjects' performance *before* they were introduced to the competitive context.

Round 2: Tournament. Subjects were randomly matched into groups of four anonymous partners. The subject who had the most correct answers in each group was rewarded with $2 for each correct answer, while the others received $0. In the case of a tie, the winner was chosen at random. We conducted an additional treatment where competition winners were rewarded $3 for each correct answer.

Round 3: Choice of competition scheme for future performance. Before performing the task, subjects chose how much of their payout would be calculated according to the tournament scheme; the rest would be paid based on the piece rate scheme.[1] For the tournament, subjects had to compete against their group members' performance in round 2, and were notified that their decision would not impact the payoffs of their partners. This task allowed us to assess subjects' willingness to compete regardless of their beliefs about the decisions of the other participants to compete, and regardless of possible consequences of their decision on other players. Again, we conducted an additional treatment with higher competition incentives, where winners were rewarded $3 for each correct answer.

Round 4. Choice of competition scheme for past piece-rate performance. In the final round, subjects did not perform the task. Instead, they indicated what percentage of their performance in *round 1* would be paid according to the tournament scheme; the reminder would be paid according to the piece-rate scheme. For the tournament option, individuals' performance was compared with round 1 performance of four randomly selected partners. Round 4 tested whether T had an effect on compensation scheme choice even when no future or past tournament performance were involved.

Belief elicitation and ranking. Following the final round, we asked participants to guess their rank of performance relative to that of their (anonymous) partners in rounds 1 and 2. We also elicited their beliefs about winning these rounds using an incentive-compatible method (Karni, 2009).[2]

---

[1] We changed two feature of the original experimental design (Niederle & Vesterlund, 2005): first, we chose lowered the incentives to compete from $4 to either $2 or $3. Second, we used a linear allocation rather than a binary decision to enter the competition (Gneezy & Pietrasz, 2013). The changes aimed to reduce the possibility of a ceiling effect, as 75% of men were shown to choose the competition scheme in the original study.

[2] This specific belief elicitation method was used as it is incentive compatible for every risk preferences. This aspect is especially important as risk might be influenced by T levels (Apicella et al., 2008; Apicella et al., 2014; Stanton et al., 2011)

All pay auctions

Competitive enterprises are often quite costly and losers are often not compensated for their investment; therefore, over-spending efforts on competition, also known as "rent-seeking" (Tollison, 1982; Tullock, 2001) might reduce efficiency and welfare (Krueger, 1974). We measured rent-seeking behavior using an all-pay auction, where the winner is determined according to the highest bid but all players pay their bids. Over-bidding relative to the rational model (Nash equilibrium prediction) commonly emerges as an empirical regularity in all-pay auction experiments (Dechenaux, Kovenock, & Sheremeta, 2014; Gneezy & Smorodinsky, 2006) and bids have been shown to increase with level of competition among bidders (Toelch, Jubera-Garcia, Kurth-Nelson, & Dolan, 2014).

In each of the three rounds of the task, subjects were randomly matched with a different number (2, 4 and 16) of other participants using a stranger protocol[3] and were endowed with 100 units of experimental points. Each subject had to decide which part of his endowments to bid for a reward of 100 additional experimental points, such that the winner was determined according to the highest bid, and all players paid their bids. Participants were not given feedback on their payoffs until the end of the entire experiment, when they were paid according to one randomly selected round (using an exchange ratio of $1=10 experimental points). The average share of endowment used for bidding is our measure of rent-seeking behavior.

Results

Competition

Comparing T and placebo groups' behavior allows us to test whether T levels had a causal influence on competitiveness.

Performance

Fig. 2 shows the cumulative distribution of problem solved across the two treatment groups (collapsed across incentive treatments; there was no

[3] In the first two rounds the number of bidders was 2 and 4; in the last round the number of bidders was the entire group of participants in the session, either 12 or 16.

significant effect of incentive level on performance). Evidently, the distributions are almost identical across the two groups, implying that T had no effect on the basic performance in the task. The absence of effect holds in both the piece-rate and tournament schemes. The average performance in Task 2 (competition) was greater than in Task 1, reflecting either a learning effect or greater motivation to preform under competition. However, we found no interaction between incentive scheme and treatment. Thus, testosterone had no influence on *performance* under competition.

Tournament entry (tasks 3 and 4)

The amount of tournament investment in task 3 allows us to measure T's influence on the willingness to enter a competition. While task performance and beliefs were (as expected) a strong predictor of tournament investment, there was no significant difference between the two treatment groups, across all performance quartiles (see Fig. 2, Table 2).

As the roles of risk and feedback aversion in the tournament investment decisions made in rounds 3 and 4 are equivalent, round 4 decisions allowed us to measure whether these factors might alone generate observed differences in willingness to compete between the two groups. We found no significant differences between the two groups also with respect to task 4's decisions (Fig.3, Table 3). If anything, testosterone decreased investment in tournament (relative to placebo) in the top quartile, but the effect did not survive correction for multiple hypotheses.

Beliefs and guessed rank

After the completion of all tasks, we (a) asked subjects to guess their rankings relative to other subjects, both under piece rate and competition schemes (in a similar manner to (Niederle & Vesterlund, 2005), we awarded a correct guess with $1), and (b) elicited subjects' belief about the likelihood that they were the best performing participants in their group, in an incentive compatible manner (Karni, 2009). In line with previous studies, we found that subjects generally over-estimated their ranking, with over 50% of subjects guessing that they are the best performing subjects in their groups, compared to less than 10% guessing that they are the worst performing (both should have been 25% if subjects were objective). However, T effects were again insignificant - both for guessed

ranking and beliefs about winning (Fig. 4, Tables 4,5). The effect of T on beliefs was absent across all performance quartiles; again, if anything, we found an effect in the opposite direction for the placebo group's worst quartile in task 1, but it did not survive correction for multiple hypotheses.

Figure 2a: Piece rate performance (correct answers in task 1) CDF

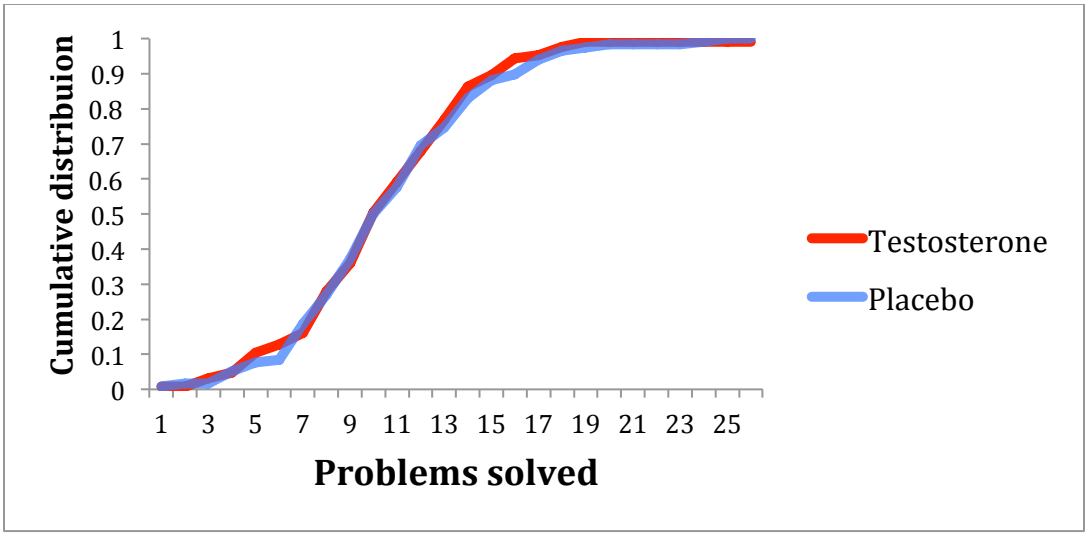

Figure 2b: Tournament performance (correct answers in task 2) CDF

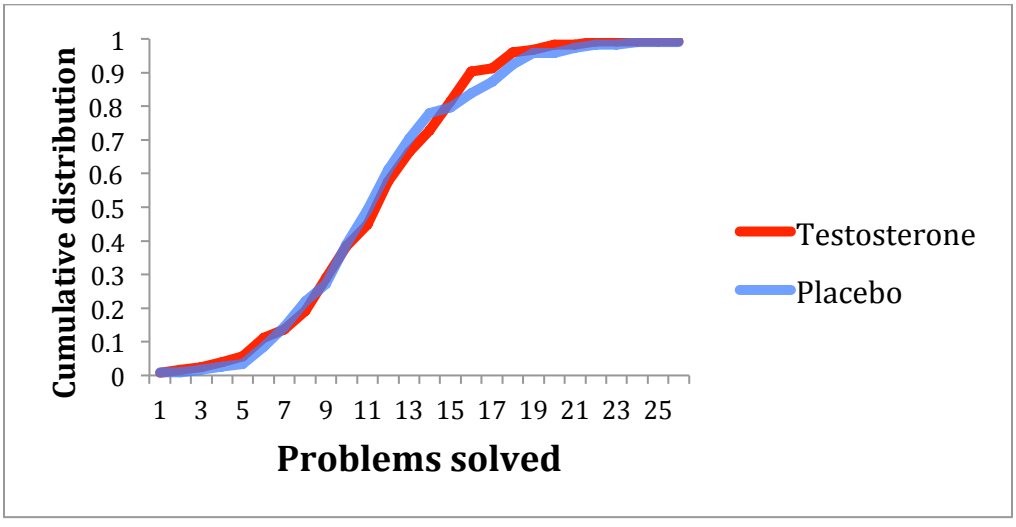

Figure 3a: Tournament investment (task 3) by performance quartile on task 2 (low incentives condition)

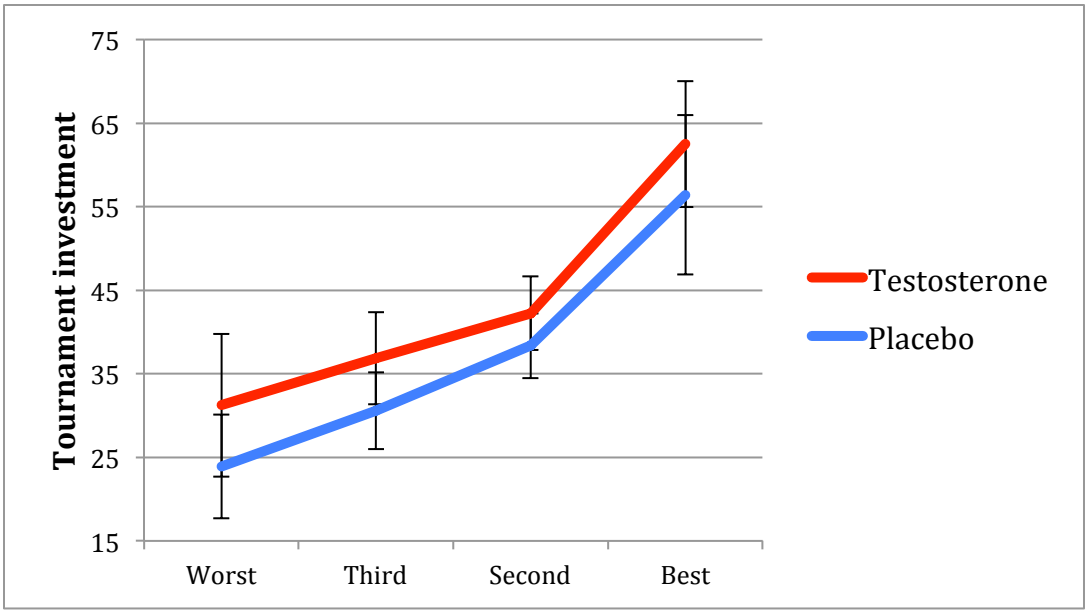

Figure 3b: Tournament investment (task 3) by rank, task 2 (low incentives)

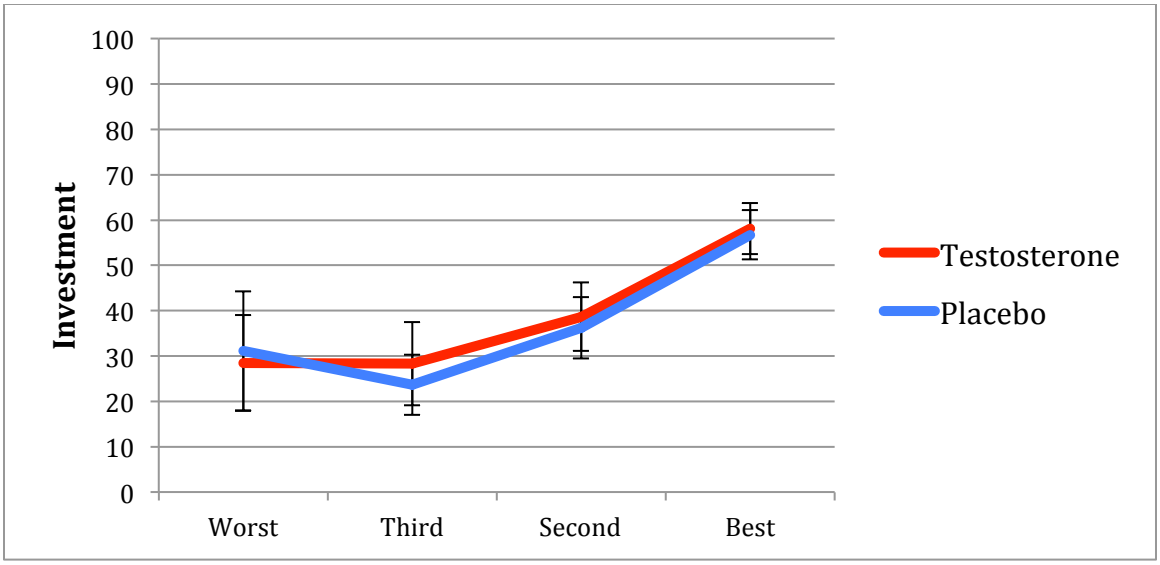

Table 5
Linear regression. Dependent variable: tournament investment (Task 3)

	(A)	(B)	(C)	(D)
Intercept	43.97***	14.79*	10.86	9.71
	(3.38)	(6.48)	(6.64)	(6.25)
Incentives	20.19***	20.85***	18.42***	4.64**
	(4.96)	(4.72)	(4.79)	(2.66)
Treatment (T=1)	1.38	1.46	2.64	4.04
	(4.40)	(4.18)	(4.17)	(3.93)
Performance (Task 2)		2.43***	1.97***	1.51**
		(5.19)	(0.50)	(0.48)
Beliefs			0.17*	0.07
			(0.07)	(0.07)
Task 4 investment				0.33
				(0.06)
N	243	243	243	243
Adjusted R^2	0.05	0.15	0.16	0.26

Signif. codes: 0 '*'; 0.001 '**'; 0.01 '*'; 0.05**

Figure 4a: Tournament investment (task 4) by performance quartile on task 1 (low incentive condition)

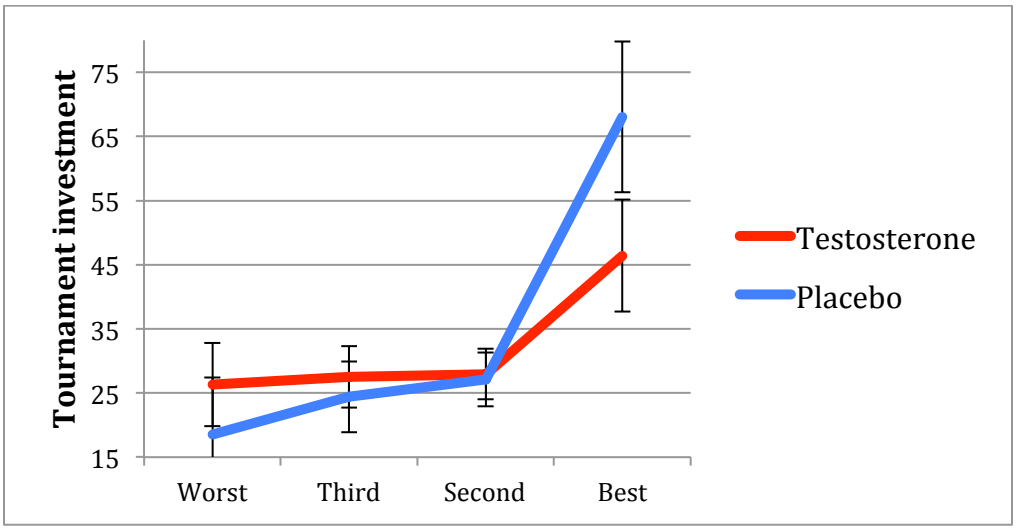

Figure 4b: Tournament investment (task 4) by rank, task 2 (low incentives)

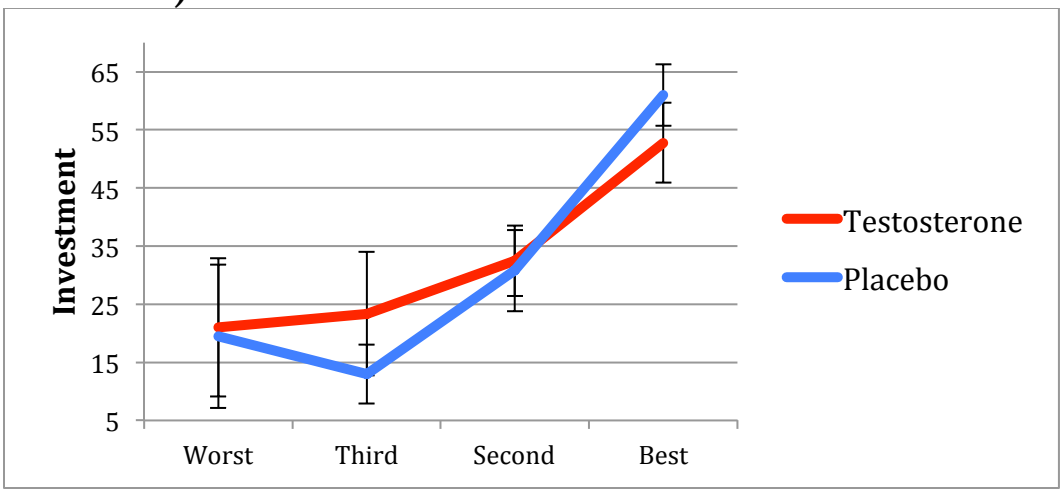

Table 6
Linear regression. Dependent variable: tournament investment (Task 4)

	(A)	(B)	(C)
Intercept	43.97***	3.05	-9.55
	(3.38)	6.77	(6.49)
Incentives	20.19***	24.67***	16.46***
	(4.96)	(4.93)	(4.69)
Treatment (T=1)	1.383	-6.03	-2.63
	(4.40)	(4.35)	(4.02)
Performance (Task 1)		3.09***	2.02***
		(0.53)	(0.51)
Beliefs			0.49***
			(0.07)
N	243	243	243
Adjusted R^2	0.07	0.18	0.31

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05**

Figure 5a: Guessed rank by treatment group

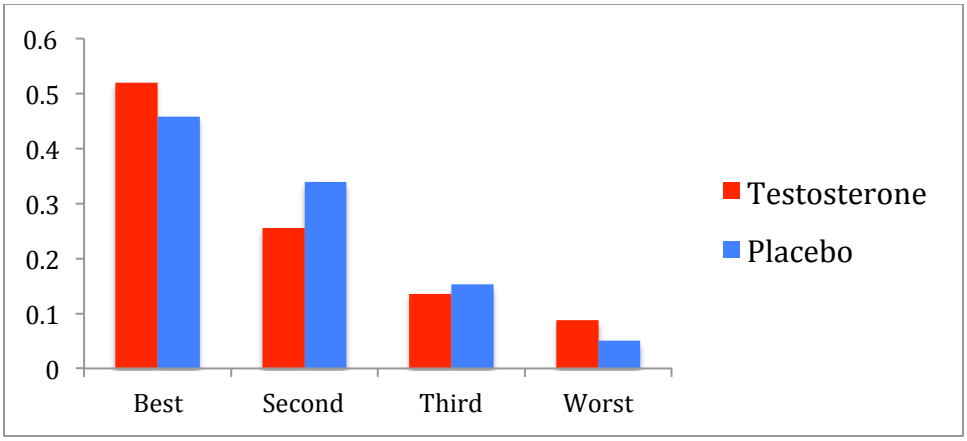

Figure 5b: Task 2 beliefs (winning) by performance quartile

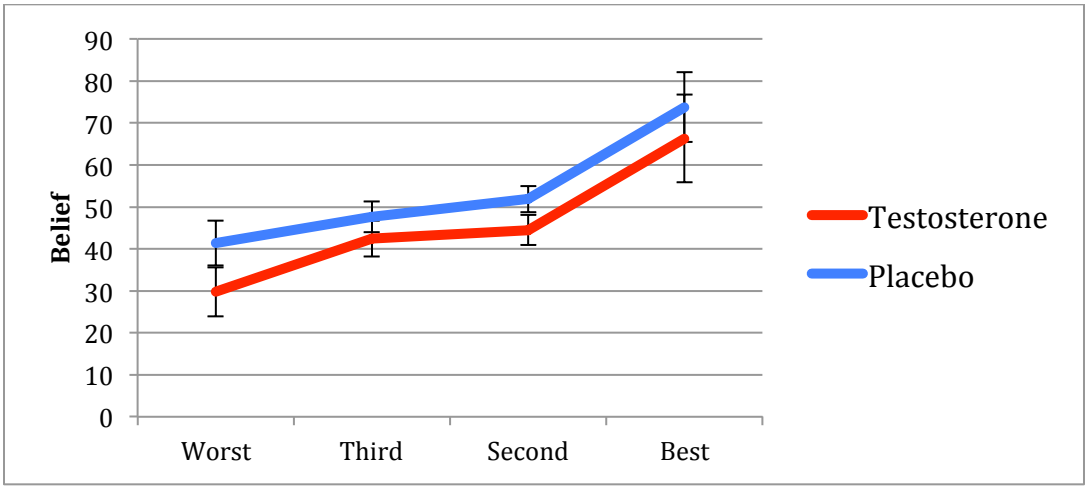

Figure 5c: Task 1 beliefs (winning) by performance quartile

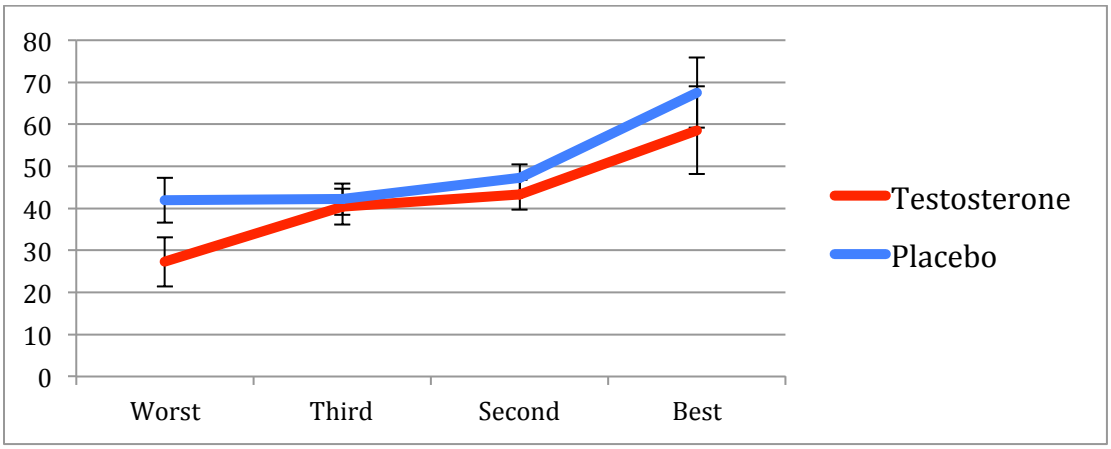

Table 7

Linear regression. Dependent variable: Beliefs (Task 2)

	(A)	(B)
Intercept	54.81***	22.91***
	(2.99)	(5.56)
Incentives	13.48**	14.21***
	(4.40)	(4.05)
Treatment (T=1)	-6.94	-6.85
	(3.89)	(3.59)
Performance (Task 2)		2.65***
		(0.40)
N	243	243
Adjusted R^2	0.04	0.18

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05**

Table 8

Linear regression. Dependent variable: Beliefs (Task 1)

	(A)	(B)
Intercept	50.14***	16.76***
	(2.87)	(4.03)
Incentives	15.05***	16.76***
	(4.22)	(4.03)
Treatment (T=1)	-7.26	-6.95
	(3.74)	(3.56)
Performance (Task 1)		2.19***
		(0.43)
N	243	243
Adjusted R^2	0.05	0.14

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

All pay auction

We found no T effects on auction bidding, neither in groups of two ($t(242=0.71, p=0.47, CI = [-6.39\ 13.71]$), nor in groups of four ($t(242=0.73, p=0.46, CI = [-5.94\ 13.00]$) or groups of sixteen subjects ($t(242=0.042, p=0.97, CI = [-8.71\ 9.10]$).

Conclusion

We pharmacologically manipulated T levels in men under a double blind protocol to test for its influence on competitiveness and rent seeking. We found no main treatment effects of T on several behavioral measures, that based on the animal literature and folk wisdom should have been affected by the treatment.

Our sample size was especially large - at the order of 4 times more than a typical T pharmacological behavioral study (e.g., (Eisenegger et al., 2010; Wibral et al., 2012)) and more than ten times greater than several other T administration studies of human decision making published in top academic journals (van Honk et al., 2012). We have used an FDA approved pharmacological treatment, the pharmacokinetics of which are clearly mapped in our target population (Eisenegger et al., 2013). Our manipulation check, using mass-spectrometry - the platinum standard of hormonal assays (table 3) revealed significant changes in the levels of T and its metabolites, and no changes in other hormones (e.g., cortisol).

In addition, we used a task that reliably generates gender differences (Buser et al., 2012; Niederle & Vesterlund, 2005). All of the variables that were expected to influence willingness to compete (i.e., skill, beliefs and incentives) had significant behavioral effects - implying high level of engagement among the subjects. Our null result is quite striking: the confidence interval around the estimated treatment effect (Table 3) rules out treatment effects greater than 8 out of 100 points with a 97.25% probability.

Yet, our conclusions should be taken with a grain of salt. There are few simple cause and effect relationships in nature, and there are several

variables that we have not yet examined in our rich data set. Therefore, our future analysis plan is to rigorously search for patterns in the data while carefully maintaining high methodological standards, in order to avoid over-fitting. In particular, we intend to investigate the possible link between behavior and the *measured* androgens levels (rather than a binary treatment effect) - in particular, DHT, a potent T metabolite with a T receptor affinity 7 times greater). We also intend to investigate the link between competitiveness and other hormones such as cortisol, and the interaction between cortisol and T (Mehta & Josephs, 2010). Last, we intend to seek whether a link exists between competitiveness and the digit ratio measure or facial masculinity – both of which were previously correlated with several economic behaviors (Coates et al., 2009; Coates & Herbert, 2008; Jia, van Lent, & Zeng, 2014; Manning & Taylor, 2001; Neave, Laing, Fink, & Manning, 2003)

References

Apicella, Coren L, Dreber, Anna, Campbell, Benjamin, Gray, Peter B, Hoffman, Moshe, & Little, Anthony C. (2008). Testosterone and financial risk preferences. *Evolution and Human Behavior, 29*(6), 384-390.

Apicella, Coren L, Dreber, Anna, Gray, Peter B, Hoffman, Moshe, Little, Anthony C, & Campbell, Benjamin C. (2011). Androgens and competitiveness in men. *Journal of Neuroscience, Psychology, and Economics, 4*(1), 54.

Apicella, Coren L, Dreber, Anna, & Mollerstrom, Johanna. (2014). Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. *Psychoneuroendocrinology, 39*, 58-64.

Archer, John. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience & Biobehavioral Reviews, 30*(3), 319-345.

Benjamin, Daniel J, Cesarini, David, van der Loos, Matthijs JHM, Dawes, Christopher T, Koellinger, Philipp D, Magnusson, Patrik KE, . . . Johannesson, Magnus. (2012). The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences, 109*(21), 8026-8031.

Bernhardt, Paul C, Dabbs Jr, James M, Fielden, Julie A, & Lutter, Candice D. (1998). Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. *Physiology & Behavior, 65*(1), 59-62.

Boksem, Maarten AS, Mehta, Pranjal H, Van den Bergh, Bram, van Son, Veerle, Trautmann, Stefan T, Roelofs, Karin, . . . Sanfey, Alan G. (2013). Testosterone inhibits trust but promotes reciprocity. *Psychological science*, 0956797613495063.

Booth, Alan, Shelley, Greg, Mazur, Allan, Tharp, Gerry, & Kittok, Roger. (1989). Testosterone, and winning and losing in human competition. *Hormones and Behavior, 23*(4), 556-571.

Bos, Peter A, Hermans, Erno J, Ramsey, Nick F, & Van Honk, Jack. (2012). The neural mechanisms by which testosterone acts on interpersonal trust. *NeuroImage, 61*(3), 730-737.

Burnham, Terence C. (2007). High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society B: Biological Sciences, 274*(1623), 2327-2330.

Buser, Thomas. (2012). The impact of the menstrual cycle and hormonal contraceptives on competitiveness. *Journal of Economic Behavior & Organization, 83*(1), 1-10.

Buser, Thomas, Niederle, Muriel, & Oosterbeek, Hessel. (2012). Gender, competitiveness and career choices: National Bureau of Economic Research.

Camerer, Colin F, Loewenstein, George, & Prelec, Drazen. (2004). Neuroeconomics: Why economics needs brains. *The Scandinavian Journal of Economics, 106*(3), 555-579.

Carré, Justin M, & McCormick, Cheryl M. (2008). Aggressive behavior and change in salivary testosterone concentrations predict willingness to engage in a competitive task. *Hormones and Behavior, 54*(3), 403-409.

Carré, Justin M, Putnam, Susan K, & McCormick, Cheryl M. (2009). Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. *Psychoneuroendocrinology, 34*(4), 561-570.

Charness, Gary, & Gneezy, Uri. (2007). Strong evidence for gender differences in investment. *Available at SSRN 648735*.

Charness, Gary, & Gneezy, Uri. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization, 83*(1), 50-58.

Coates, John M, Gurnell, Mark, & Rustichini, Aldo. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences, 106*(2), 623-628.

Coates, John M, & Herbert, Joe. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proceedings of the National Academy of Sciences, 105*(16), 6167-6172.

Crockett, Molly J, & Fehr, Ernst. (2013). Social brains on drugs: tools for neuromodulation in social neuroscience. *Social cognitive and affective neuroscience*, nst113.

Croson, Rachel, & Gneezy, Uri. (2009). Gender differences in preferences. *Journal of Economic literature*, 448-474.

Dechenaux, Emmanuel, Kovenock, Dan, & Sheremeta, Roman M. (2014). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 1-61.

Dreber, Anna, Von Essen, Emma, & Ranehill, Eva. (2011). Outrunning the gender gap—boys and girls compete equally. *Experimental Economics, 14*(4), 567-582.

Du, Joanna Y, Sanchez, Puy, Kim, Lila, Azen, Colleen G, Zava, David T, & Stanczyk, Frank Z. (2013). Percutaneous progesterone delivery via cream or gel application in postmenopausal women: a randomized cross-over study of progesterone levels in serum, whole blood, saliva, and capillary blood. *Menopause, 20*(11), 1169-1175.

Eisenegger, Christoph, Haushofer, Johannes, & Fehr, Ernst. (2011). The role of testosterone in social interaction. *Trends in cognitive sciences, 15*(6), 263-271.

Eisenegger, Christoph, Naef, M, Snozzi, R, Heinrichs, M, & Fehr, E. (2012). Eisenegger et al. reply. *Nature, 485*(7399), E5-E6.

Eisenegger, Christoph, Naef, Michael, Snozzi, Romana, Heinrichs, Markus, & Fehr, Ernst. (2010). Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature, 463*(7279), 356-359.

Eisenegger, Christoph, von Eckardstein, Arnold, Fehr, Ernst, & von Eckardstein, Sigrid. (2013). Pharmacokinetics of testosterone and estradiol gel preparations in healthy young men. *Psychoneuroendocrinology, 38*(2), 171-178.

Fehr, Ernst, & Camerer, Colin F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in cognitive sciences, 11*(10), 419-427.

Fischbacher, Urs. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171-178.

Flory, Jeffrey A, Leibbrandt, Andreas, & List, John A. (2010). Do competitive work places deter female workers? a large-scale natural field experiment on gender differences in job-entry decisions: National Bureau of Economic Research.

Glimcher, Paul W, & Fehr, Ernst. (2013). *Neuroeconomics: Decision making and the brain*: Academic Press.

Gneezy, Uri, Leonard, Kenneth L, & List, John A. (2006). Gender differences in competition: the role of socialization. *manuscript, University of Chicago*.

Gneezy, Uri, Leonard, Kenneth L, & List, John A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica, 77*(5), 1637-1664.

Gneezy, Uri, Niederle, Muriel, & Rustichini, Aldo. (2003). Performance in competitive environments: Gender differences. *QUARTERLY JOURNAL OF ECONOMICS-CAMBRIDGE MASSACHUSETTS-, 118*(3), 1049-1074.

Gneezy, Uri, & Pietrasz, Aniela. (2013). When half of the men are more competitive than all women: Working paper.

Gneezy, Uri, & Rustichini, Aldo. (2004). Gender and competition at a young age. *American Economic Review*, 377-381.

Gneezy, Uri, & Smorodinsky, Rann. (2006). All-pay auctions—an experimental study. *Journal of Economic Behavior & Organization, 61*(2), 255-275.

Healy, Andrew, & Pate, Jennifer. (2011). Can Teams Help to Close the Gender Competition Gap?*. *The Economic Journal, 121*(555), 1192-1204.

Horton, R, & Tait, JF. (1966). Androstenedione production and interconversion rates measured in peripheral blood and studies on the possible site of its conversion to testosterone. *Journal of Clinical Investigation, 45*(3), 301.

Hurwitz, Shelley, Cohen, Richard J, & Williams, Gordon H. (2004). Diurnal variation of aldosterone and plasma renin activity: timing relation to melatonin and cortisol and consistency after prolonged bed rest. *Journal of Applied Physiology, 96*(4), 1406-1414.

Jia, Yuping, van Lent, Laurence, & Zeng, Yachang. (2014). Testosterone and financial misreporting. *Available at SSRN 2265510*.

Johnson, Ryan T, & Breedlove, S Marc. (2010). Human trust: Testosterone raises suspicion. *Proceedings of the National Academy of Sciences, 107*(25), 11149-11150.

Kandasamy, Narayanan, Hardy, Ben, Page, Lionel, Schaffner, Markus, Graggaber, Johann, Powlson, Andrew S, . . . Coates, John. (2014). Cortisol

shifts financial risk preferences. *Proceedings of the National Academy of Sciences, 111*(9), 3608-3613.

Karni, Edi. (2009). A mechanism for eliciting probabilities. *Econometrica, 77*(2), 603-606.

Kocoska-Maras, Ljiljana, Zethraeus, Niklas, Rådestad, Angelique Flöter, Ellingsen, Tore, von Schoultz, Bo, Johannesson, Magnus, & Hirschberg, Angelica Lindén. (2011). A randomized trial of the effect of testosterone and estrogen on verbal fluency, verbal memory, and spatial ability in healthy postmenopausal women. *Fertility and sterility, 95*(1), 152-157.

Krueger, Anne O. (1974). The political economy of the rent-seeking society. *The American Economic Review*, 291-303.

Manning, John T. (2002). *Digit ratio: A pointer to fertility, behavior, and health*: Rutgers University Press.

Manning, John T, & Taylor, Rogan P. (2001). Second to fourth digit ratio and male ability in sport: implications for sexual selection in humans. *Evolution and Human Behavior, 22*(1), 61-69.

Mayo, A, Macintyre, H, Wallace, AM, & Ahmed, SF. (2004). Transdermal testosterone application: pharmacokinetics and effects on pubertal status, short-term growth, and bone turnover. *The Journal of Clinical Endocrinology & Metabolism, 89*(2), 681-687.

Mazur, Allan, Booth, Alan, & Dabbs Jr, James M. (1992). Testosterone and chess competition. *Social Psychology Quarterly*, 70-77.

McCaul, Kevin D, Gladue, Brian A, & Joppa, Margaret. (1992). Winning, losing, mood, and testosterone. *Hormones and Behavior, 26*(4), 486-504.

Mehta, Pranjal H, & Beer, Jennifer. (2010). Neural mechanisms of the testosterone–aggression relation: The role of orbitofrontal cortex. *Journal of Cognitive Neuroscience, 22*(10), 2357-2368.

Mehta, Pranjal H, & Josephs, Robert A. (2006). Testosterone change after losing predicts the decision to compete again. *Hormones and Behavior, 50*(5), 684-692.

Mehta, Pranjal H, & Josephs, Robert A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior, 58*(5), 898-906.

Mehta, Pranjal H, Wuehrmann, Elizabeth V, & Josephs, Robert A. (2009). When are low testosterone levels advantageous? The moderating role of individual versus intergroup competition. *Hormones and Behavior, 56*(1), 158-162.

Miller, David I, & Halpern, Diane F. (2014). The new science of cognitive sex differences. *Trends in cognitive sciences, 18*(1), 37-45.

Neave, Nick, Laing, Sarah, Fink, Bernhard, & Manning, John T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1529), 2167-2172.

Niederle, Muriel, & Vesterlund, Lise. (2005). Do women shy away from competition? Do men compete too much? : National Bureau of Economic Research.

Niederle, Muriel, & Vesterlund, Lise. (2010). Explaining the gender gap in math test scores: The role of competition. ***The Journal of Economic Perspectives*****, 129-144.**

Niederle, Muriel, & Vesterlund, Lise. (2011). Gender and competition. ***Annu. Rev. Econ., 3*****(1), 601-630.**

Price, Joseph. (2008). Gender differences in the response to competition. ***Industrial & Labor Relations Review, 61*****(3), 320-333.**

Rangel, Antonio, Camerer, Colin, & Montague, P Read. (2008). A framework for studying the neurobiology of value-based decision making. ***Nature Reviews Neuroscience, 9*****(7), 545-556.**

Rolf, C, Knie, U, Lemmnitz, G, & Nieschlag, E. (2002). Interpersonal testosterone transfer after topical application of a newly developed testosterone gel preparation. ***Clinical endocrinology, 56*****(5), 637-641.**

Sapienza, Paola, Zingales, Luigi, & Maestripieri, Dario. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. ***Proceedings of the National Academy of Sciences, 106*****(36), 15268-15273.**

Smith, Adam, & Nicholson, Joseph Shield. (1887). ***An Inquiry Into the Nature and Causes of the Wealth of Nations*****: T. Nelson and Sons.**

Snow, Marcellus S. (2002). Competition as a discovery procedure. ***Quarterly Journal of Austrian Economics, 5*****(3), 9-23.**

Stanton, Steven J, O'Dhaniel, A, McLaurin, R Edward, Kuhn, Cynthia M, LaBar, Kevin S, Platt, Michael L, & Huettel, Scott A. (2011). Low-and high-testosterone individuals exhibit decreased aversion to economic risk. ***Psychological science*****.**

Sutter, Matthias, & Rützler, Daniela. (2010). Gender differences in competition emerge early in life.

Toelch, Ulf, Jubera-Garcia, Esperanza, Kurth-Nelson, Zeb, & Dolan, Raymond J. (2014). Competition strength influences individual preferences in an auction game. ***Cognition, 133*****(2), 480-487.**

Tollison, Robert D. (1982). Rent seeking: A survey. ***Kyklos, 35*****(4), 575-602.**

Tullock, Gordon. (2001). Efficient rent seeking ***Efficient Rent-Seeking*** **(pp. 3-16): Springer.**

van Honk, Jack, Montoya, Estrella R, Bos, Peter A, van Vugt, Mark, & Terburg, David. (2012). New evidence on testosterone and cooperation. ***Nature, 485*****(7399), E4-E5.**

Voracek, Martin. (2014). No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D: 4D): a comprehensive meta-analysis and critical evaluation of research. ***Evolution and Human Behavior, 35*****(5), 430-437.**

Watson, David, & Clark, Lee Anna. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.

Wennekers, Sander, & Thurik, Roy. (1999). Linking entrepreneurship and economic growth. ***Small business economics, 13*****(1), 27-56.**

Wibral, Matthias, Dohmen, Thomas, Klingmüller, Dietrich, Weber, Bernd, & Falk, Armin. (2012). Testosterone administration reduces lying in men. *PloS one, 7*(10), e46774.

Wingfield, John C, Hegner, Robert E, Dufty Jr, Alfred M, & Ball, Gregory F. (1990). The "challenge hypothesis": theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. *American Naturalist*, 829-846.

Wright, Nicholas D, Bahrami, Bahador, Johnson, Emily, Di Malta, Gina, Rees, Geraint, Frith, Christopher D, & Dolan, Raymond J. (2012). Testosterone disrupts human collaboration by increasing egocentric choices. *Proceedings of the Royal Society B: Biological Sciences, 279*(1736), 2275-2280.

Zak, Paul J, Kurzban, Robert, Ahmadi, Sheila, Swerdloff, Ronald S, Park, Jang, Efremidze, Levan, . . . Matzner, William. (2009). Testosterone administration decreases generosity in the ultimatum game. *PloS one, 4*(12), e8330.

Zethraeus, Niklas, Kocoska-Maras, Ljiljana, Ellingsen, Tore, Von Schoultz, BO, Hirschberg, Angelica Lindén, & Johannesson, Magnus. (2009). A randomized trial of the effect of estrogen and testosterone on economic behavior. *Proceedings of the National Academy of Sciences, 106*(16), 6535-6538.