Single dose testosterone administration impairs cognitive reflection in men

Gideon Nave*, Department of Marketing, The Wharton School of the University of Pennsylvania

Amos Nadler*, Department of Finance, Ivey School of Business, Western University

David Zava, ZRT Laboratories LTD

Colin Camerer, Department of Humanities and Social Sciences, Caltech

* These authors contributed equally to this work

Abstract

The sex steroid testosterone regulates reproductive behaviors such as intra-male fighting and mating in non-humans. Correlational studies have linked testosterone with aggression and disorders associated with poor impulse control, but the neuropsychological processes at work are poorly understood. Building on a dual-process framework, we identified an underlying mechanism for testosterone's behavioral effects in humans: reducing cognitive reflection. In the largest behavioral testosterone administration study to date, 243 men received either testosterone or placebo and took the Cognitive Reflection Test (CRT), that estimated their capacity to override incorrect intuitive judgments with deliberate correct responses. Testosterone administration reduced CRT scores. The effect was robust to controlling for age, mood, math skills, treatment expectancy and 14 other hormones, and held for each of the CRT questions in isolation. Our findings suggest a unified mechanism underlying testosterone's varied behavioral effects in humans, and provide novel, clear and testable predictions.

Keywords *testosterone, single administration, cognitive reflection, dual process, impulse control, Neuroeconomics*

The androgenic hormone testosterone (abbreviated "T") is produced in the adrenal glands, the male testes, and in smaller quantities in female ovaries. T affects physiology, brain development, and behavior throughout life. T is released into the bloodstream and in the brain in response to external stimuli, such as the presence of an attractive mate or winning a competition, modulating physiological and cognitive processes in a context-sensitive manner (Archer, 2006; Eisenegger, Haushofer, & Fehr, 2011; Mazur, 2005; Ronay & von Hippel, 2010).[i] In many non-human species, T levels rise amid breeding season to facilitate reproductive behaviors such as intra-male fighting and mating (Archer, 2006; Edwards, 1969; Mazur, 2005; Wingfield, Hegner, Dufty Jr, & Ball, 1990). Laboratory studies have further shown that T administration causally induces aggression, mating, and behavioral disinhibition in rodents and birds (Bing et al., 1998; Edwards, 1969; Svensson, Åkesson, Engel, & Söderpalm, 2003; Wingfield et al., 1990).

A largely open question is how T affects human cognition and decision-making. Studies in young men found correlations between endogenous T and physical aggression, sensation seeking, and impulse control disorders such as drug abuse, bulimia, and borderline personality disorder

(Campbell et al., 2010; Cotrufo et al., 2000; Dabbs, Carr, Frady, & Riad, 1995; Daitzman & Zuckerman, 1980; Garcıa, 2005; Martin et al., 2002; Reynolds et al., 2007). Moreover, prefrontal brain regions involved in impulse control contain androgen receptors (Finley & Kritzer, 1999), and an imaging study showed that decreased prefrontal activity mediated the correlation of endogenous T with rejections of unfair ultimatum bargaining offers (Mehta & Beer, 2010), a behavior that can be interpreted as impulsive based on other behavioral studies (Grimm & Mengel, 2011).

Due to the bi-directional influences between hormone levels and organisms' environment and behavior, cause and effect are conflated in correlational studies. Studies over the past decade addressed this important limitation by administering T under a placebo-controlled protocol, and observing its influence on behavioral outcomes. Although the hypotheses and behavioral measures testing them vary between studies, many findings are consistent with the presumption that T administration biases decision-making towards rapid, intuitive responses. For example, in males, T increased reactive aggression (Carré et al., 2016; Pope, Kouri, & Hudson, 2000)[ii], and reduced lying and deception (van Honk et al., 2016; Wibral, Dohmen, Klingmüller, Weber, & Falk, 2012), behaviors that are associated with high cognitive load and slow response times (Vrij, Granhag, Mann, & Leal, 2011). In females[iii], T increased cooperation in the public goods game (van Honk, Montoya, Bos, van Vugt, & Terburg, 2012)[iv], a response that is associated with short response times (Rand, Greene, & Nowak, 2012). However, when cooperation required over-riding one's intuitive response by incorporating the opinions of others, T reduced cooperation (Wright et al., 2012). T also increased ultimatum game generosity (Eisenegger, Naef, Snozzi, Heinrichs, & Fehr, 2010), a behavior that was shown to increase under cognitive load (Cappelletti, Güth, & Ploner, 2011).

The current study aims to formally test whether and how T influences decision-making processes in humans. We build on the dual-process framework (Evans, 2003), according to which humans employ two types of information processing in the course of decision-making: "System 1" (intuitive) processes occur automatically, rapidly and effortlessly, but might provide sub-optimal responses. "System 2" (deliberate) processes are relatively slow and computationally demanding, but are more likely to produce accurate responses. An important function of system 2 is monitoring system 1 responses and overriding them when needed (akin to 'checking work' on an algebra problem).

Given the findings discussed above, we propose that T biases decision-making towards system 1 responses. We tested this hypothesis by randomly administering a single dose of either T or placebo to a sample of 243 males, and measuring its influence on performance in a task specifically designed to identify one's tendency towards either intuitive or deliberate information processing, the Cognitive Reflection Test (CRT, (Frederick, 2005)). The CRT is a 3-item questionnaire that assesses the capacity to monitor one's own intuitive judgments and override them when appropriate. CRT scores predict diverse behaviors, including preference of immediate gratification over greater delayed rewards and display of various decision-making biases such as the conjunction fallacy (Toplak, West, & Stanovich, 2011).

Here is an illustrative CRT question:

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

When faced with this question, an immediate incorrect answer (10 cents) automatically arises in most people's minds. Obtaining the correct answer ($0.05) requires simple addition and subtraction, yet detecting the incorrect answer requires cognitively reflecting on the validity of the intuitive answer by engaging in deliberate, yet easy to perform calculations (i.e., checking that the bat – ball difference is $1.00 and their sum is $1.10).

We hypothesized that T administration would increase participants' tendency to rely on their intuitive judgments, and therefore impair their CRT performance relative to participants who received a placebo. To rule out various confounding factors, namely T's potential influences on engagement, motivation, or arithmetic skills, participants also took part in an additional math task as a control. Participants also provided pre- and post-treatment saliva samples that were assayed by liquid chromatography tandem mass spectrometry (LC-MS/MS) as manipulation checks, and to control for levels of other hormones that might influence the task.

Finally, our study further allowed testing two previously reported associations between the CRT and hormonal variables: the 2D:4D digit ratio, a purported proxy of prenatal T exposure (Bosch-Domènech, Brañas-Garza, & Espín, 2014), and the stress steroid hormone cortisol (Margittai et al., 2015).

Methods

Participants

Two hundred and forty three males (mostly college students, see SOM, Table S1 for demographic details) were randomly administered either T (n=125) or placebo (n=118) topical gel under a double blind between-subjects protocol.[v] Sample size was chosen to be as large as possible given the study's budget constraints, making it the largest behavioral T administration experiment conducted to date. The institutional review boards of Caltech and Claremont Graduate University approved the study, all participants gave informed consent and no adverse events occurred.

Procedure

The timeline of our experimental procedure is illustrated in Fig. 1. Participants first arrived at the lab at 9:00 in the morning of their experimental session. They signed an informed consent form and proceeded to a designated room where their hands were scanned to obtain digit ratio measurement. Then, participants were randomly assigned to private cubicles where they completed demographic and mood questionnaires and provided the initial baseline saliva sample. Afterwards, participants proceeded to a designated room for T or placebo gel application.

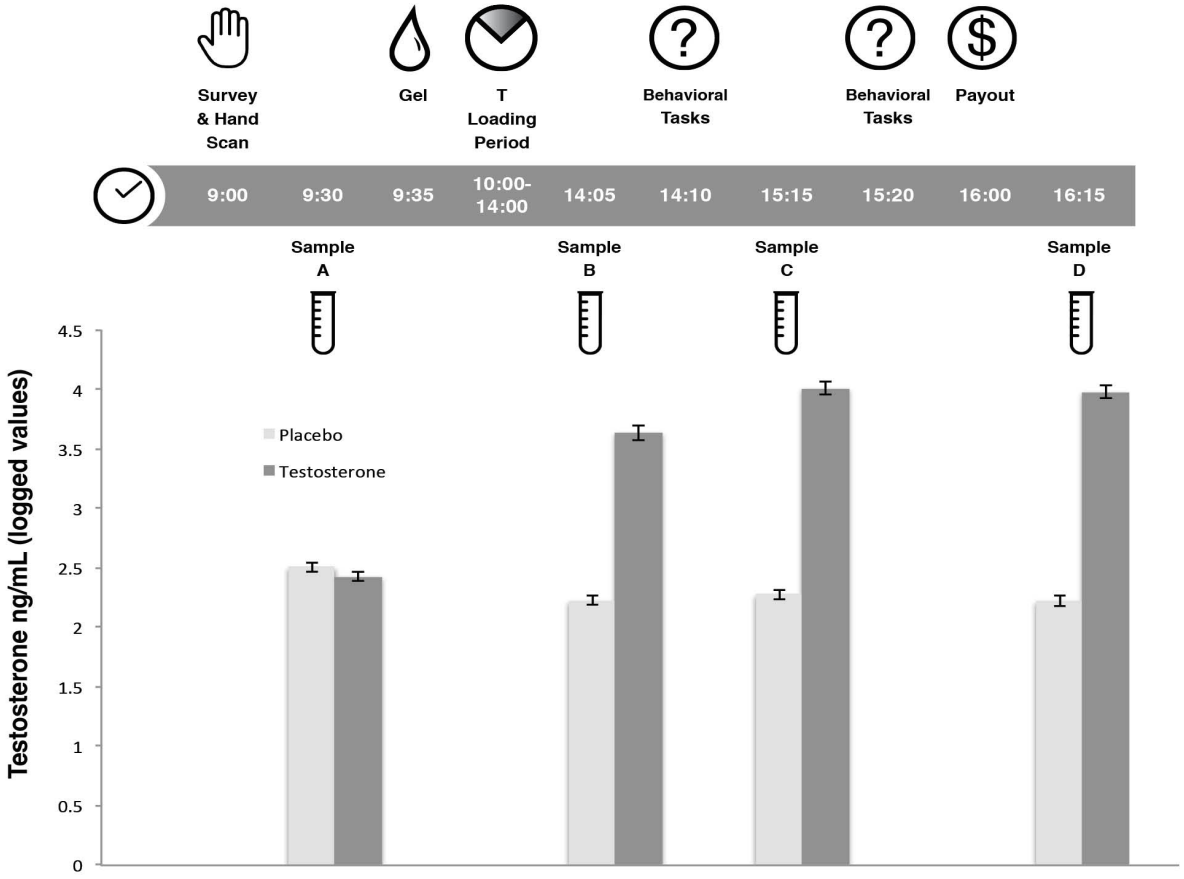
All participants returned to the lab at 2:00pm (with no incidents of lateness), provided a second saliva sample and began the behavioral experiment at the same cubicle they were assigned in the morning session. The time frame between gel application and behavioral experiment was chosen so that tasks took place when the T group participants 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 and none included feedback about the participants' monetary payoffs (to avoid endogenous changes in T from changes in payoff). Only the final task included feedback regarding the participants' 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 upon favorably by Institutional Review Boards. 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.

Four saliva samples were collected throughout the experiment to achieve high-resolution monitoring of hormonal changes during the experiment and control for their influences (further details of collection frequency and time below). The accuracy and consistency of sampling times was 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 participants, 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. Both of the behavioral tasks reported here were computerized and occurred in the first hour of the experiment, between the second and third saliva samples. Following the experiment, participants completed an exit survey, where they indicated their expectancies about which of the two treatments they had received, and then were privately paid in cash according to their performance. In order to reduce the potential effect of a female experimenter's presence on T levels and T-related behaviors, male researchers conducted all experimental sessions (Ronay & von Hippel, 2010).

Figure 1: Experiment timeline and salivary testosterone levels

Participants arrived at the lab at 9:00 am, had their hands scanned, completed an intake survey and provided baseline saliva sample "A" before application of either testosterone or placebo topical gel. After a four-hour loading period, participants returned 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 CRT and math tasks took place between saliva sample B and C.



Treatment administration

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 participants 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 of similar texture and viscosity (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 (Eisenegger et al., 2013). The single-dose study demonstrated that plasma T levels peaked 3 hours following exogenous topical

administration, and that T measurements stabilized at high levels during the time window between 4 and 7 hours following administration. Therefore we had all participants return to the lab 4.5 hours after receiving gel, when androgen levels were higher and stable.

Participants 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 research assistant, participants were instructed to thoroughly wash their hands with warm water and soap, avoid touching any part of their body before thoroughly washing hands, and abstain from all skin-to-skin contact with females, as recommended by T 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.

Measures

Saliva sampling. Each participant provided four saliva samples at predetermined sampling times throughout the session: (1) Before treatment administration (all samples collected between 9:25 and 9:34 am) (2) upon return to the lab, just prior to starting the behavioral tasks (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 multiple blood draws throughout the experimental session. Each saliva sample was time stamped. 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 assays. 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. Further details about the assay procedure are available in the SOM. A series of one-sample Kolmogorov-Smirnov tests for conformity to Gaussian (Table S2 in SOM) indicated that all hormonal measurement distributions were better approximated by a Gaussian distribution 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 in order to make their distributions closer to Gaussian. Three saliva samples (two from sample 2, and one from sample 3) could not be analyzed due to insufficient fluid and thus excluded from analyses involving these hormonal samples.

Mood questionnaire. Participants completed the PANAS-X scale (Watson & Clark, 1999), both pre-treatment (in the morning) and post-treatment (in the afternoon). Three participants did not answer all of the negative affect items in their questionnaires, and five participants did not complete all of the positive affect items; these participants were excluded from analyses that include these scales as control variables.

Digit ratio measurement. The ratio of second (index) finger length to fourth (ring) finger (abbreviated 2D:4D) is considered a proxy for pre-natal T exposure. Participants' 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. Regression models (tables S5-S7 in SOM) are reported using the right hand measurements. All of the results hold when replacing the right hand 2D:4D by either the left hand digit ratio or the averaged digit ratio of both hands.

Cognitive reflection test (CRT). The CRT is designed to assess a specific cognitive function: the ability to detect and suppress an intuitive, rapid ("system 1") incorrect answer in favor of a reflective and deliberative ("system 2") correct answer.

The test consists of the following three questions:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Participants solved the CRT without time pressure, and were told they would be paid $1 for each correct answer and an additional bonus of $2 if they correctly solved all three questions. Thus, they could have earned as much as $5 in just a few minutes (to put this amount in perspective, the minimum wage in California, which is typical for student jobs, is $9 / hour).[vi]

The CRT has never been conducted in the laboratory's subject pool, mitigating the concern that participants were previously exposed to the questions or their solutions (Toplak, West, & Stanovich, 2014). In line with this notion, a recent meta-analysis of 118 studies, comprising of 44,558 participants across 21 countries, shows that the exposure concern is fully driven by studies conducted online using Amazon Mechanical Turk (Brañas-Garza et al., 2015).

Math task. Participants completed a math task to control for their arithmetic skills, engagement levels, attention, and motivation. They had five minutes to correctly add as many sets of five two-digit numbers as possible. Participants could use pen and paper but were not allowed to use a calculator. The two-digit numbers in each problem were randomly drawn and presented in the following way on the computer screen (participants entered their summation of the five numbers in the blank box on the right):

21	35	48	29	83	

Once a participant submitted an answer, a new problem appeared. Participants received $1 for each correct answer and $0 for an incorrect answer.

Treatment expectancy. One previous study indicated an effect of participants' beliefs about the

treatment they had received on behavior (Eisenegger et al., 2010). We therefore asked participants 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 SOM, table S1). Two participants did not report their treatment expectancy and therefore were excluded from all analyses in which this measure was used as a control.

Results

Manipulation check

We observed elevated levels of T and its metabolites (e.g., dihydrotestosterone) in the saliva measurements of the T group but not in the placebo group (Fig. 1). There were no treatment effects on either mood, treatment expectancy, or levels of all other measured hormones, ruling out these potential indirect treatment influences on the task (see SOM Tables S1, S3 and S4 for further details).

The influence of testosterone on the cognitive reflection test

In line with our main hypothesis, the T group achieved significantly lower CRT scores compared to placebo, with 20% fewer correct answers, (β=-0.43, 95% confidence interval (CI)= [-0.72 -0.16], t(241)=-3.07, p=0.002, Cohen's *d*: -0.42, CI = [-0.70 -0.15], see Fig. 2a, full analyses details and all models are summarized in Section 4 of the SOM).[vii] Moreover, relative to the placebo group, incorrect intuitive answers were more common, and correct answers were less common in the T group, for each of the three CRT questions analyzed separately (see Fig. 2c-e and SOM). Participants who received T also gave incorrect answers more quickly, and correct answers more slowly, than participants who received placebo (SOM, Table S9). These differences are consistent with our functional proposition of T-induced bias toward system 1 intuitions.

Several factors other than cognitive reflection, such as reduced engagement, motivation, or arithmetic skills, might have lowered CRT scores following T treatment. To control for these potential influences, participants performed a separate arithmetic task of adding sequences of five two-digit numbers under time pressure (5 minutes) with the incentive of $1 for each correct answer.[viii] While arithmetic scores explained a substantial part of the between-participants variance in CRT scores (β=0.08, CI=[0.04 0.11], t(240)=4.69, p<0.001), they were unaffected by T administration (β=0.04, CI= [-1.01 1.08], t(241)=0.07, p=0.94, Cohen's *d*: 0.01, CI = [-0.25 0.26]). Crucially, the effect of T on CRT scores remained highly significant after controlling for arithmetic performance, age, treatment expectancy, affective state, 2D:4D digit ratio, and the levels of all other measurable hormones that were not affected by the pharmacological manipulation (SOM, Table S5). Further analysis corroborated that CRT scores were influenced by levels of T, rather than by other metabolites that were affected by T treatment (SOM, Table S6). Finally, all of the effects hold either when using pre-task (sample 2) or post-task (sample 3) hormonal measurements (Table S7), ensuring that the behavioral effects were not caused by endogenous hormonal fluctuations that might have occurred during the experimental battery.

Additional analyses

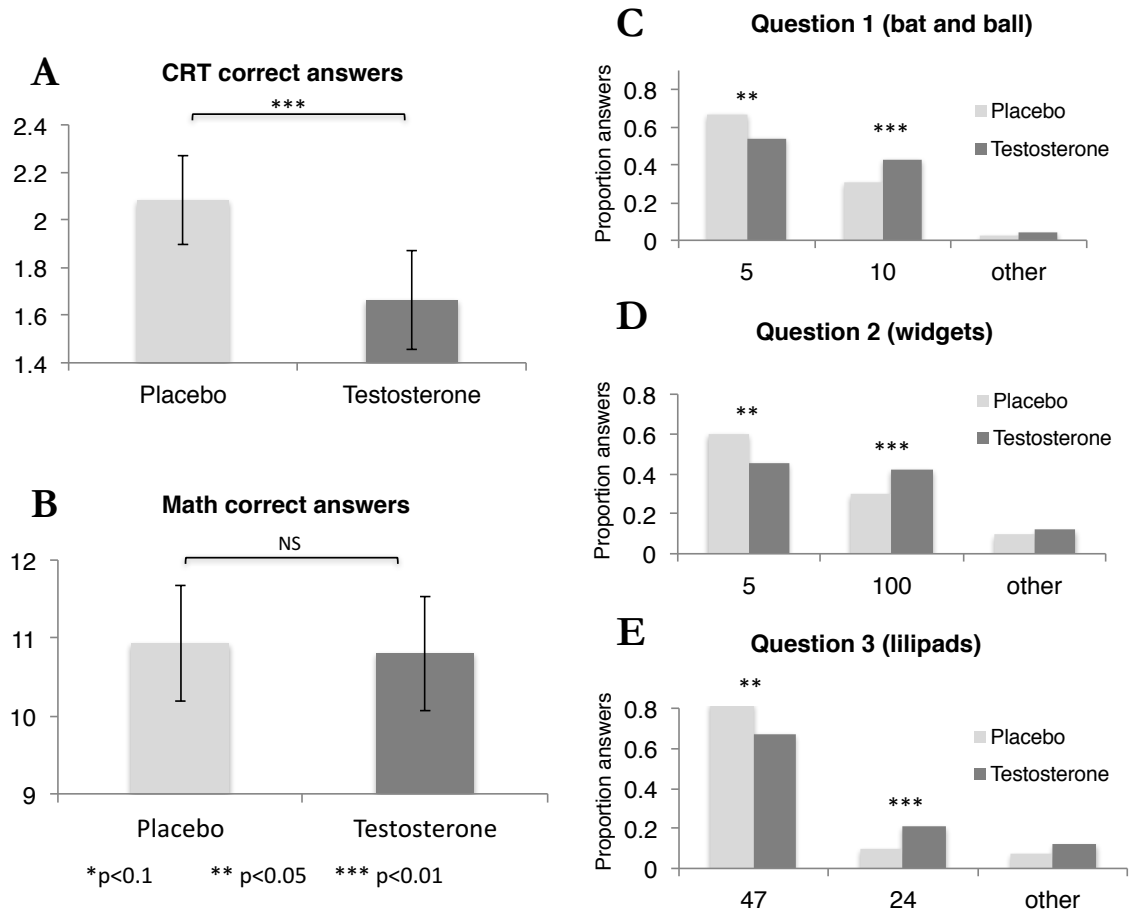
We conducted two additional analyses of other hormones and proxies, by testing their influences on CRT performance based on previous findings. These analyses were conducted using linear regression models that included control for math performance.

First, we did not find an association between CRT scores and the 2D:4D ratio, a purported proxy of prenatal T exposure (right hand: β=-1.852, CI= [-5.84 2.142], t(238)=-0.913, p=0.362; left hand: β=-0.988, CI= [-5.082 3.105], t(238)=-0.476, p=0.634), although the coefficients were negative, in line with a previous report (Bosch-Domènech et al., 2014). However, perhaps because our sample was roughly half the size of the aforementioned study and used a pharmacological manipulation, the absence of effect might have been a false negative outcome.

Second, and in line with a previous administration study (Margittai et al., 2015), we found a negative correlation between endogenous cortisol levels and CRT scores (r(241)=-0.131, p=0.043, CI=[-0.253 -0.004]).

Figure 2: behavioral results

(a) Mean CRT scores under placebo and testosterone treatments. **(b)** Mean arithmetic scores under placebo and testosterone treatment **(c-e)** proportions of answers given to each of the CRT questions separately. The left bar represents the correct, deliberate answer; the middle bar represents the incorrect intuitive answer; the right bar represent incorrect answers that are different from the intuitive one. Error bars denote 95% confidence intervals.



Discussion

Our results demonstrate a clear causal effect of T on human cognition and decision-making. We now relate this effect to previous findings in the literature.

There is extensive evidence that T increases instinctive[ix] responses (Tinbergen, 1951) with sensitivity to context in animals. In non-human species, T levels typically rise during breeding season, to facilitate reproductive behaviors such as mating and intra-male aggression (Edwards, 1969; Eisenegger et al., 2011; Mazur, 2005; Wingfield et al., 1990). In humans, the analogous effects are release of T and its precursors during competition, challenge, presence of an attractive mate, and in anticipation of sexual activity (Archer, 2006; Eisenegger et al., 2011; Mazur, 2005; Ronay & von Hippel, 2010). Our results corroborate the proposed T-initiates-instinctive-behavior neurocognitive pathway if one thinks of intuitive CRT responses as cognitive instincts. In this account, T's effect on cognition is an evolutionary vestige (or repurposing) that blunts careful deliberation in favor of rough and rapid processing.

While reduced cognitive reflection is maladaptive in contexts where deliberation is helpful (such as the CRT), the effect of T embodies several advantages. Deliberate information processing is slow and effortful. Therefore, T-induced bias towards rapid automatic responses is beneficial when acting slowly is costly (e.g., during intra-male physical challenges), and when reproductive success depends on innate instincts (e.g., during copulation). Furthermore, T-induced decrease in cognitive reflection is a biologically plausible mechanism, encompassing several advantages that we list below.

First, T is a steroid hormone that, once released, is capable of rapidly reaching many target locations in the central and peripheral nervous systems simultaneously. This makes T an ideal mechanism for context-sensitive coordination between anatomically distant systems, when reproductive success is on the line (Roney, 2016).[x] Along this line of argumentation, the steroid hormone cortisol, which is also released in contexts where it is critical to respond fast (as a part of the HPA axis stress response), has similar influence on the CRT (Margittai et al., 2015).

Second, androgen receptors are found in high density in both prefrontal regions involved in down-regulation of automatic responses (Finley & Kritzer, 1999), and in regions of the limbic system that control reproductive behavior and rapid information processing, such as the amygdala and the hypothalamus (Clancy, Bonsall, & Michael, 1992). Further, several imaging studies have linked T with reduced activation of these prefrontal brain regions (e.g., (Mehta & Beer, 2010)), and functional decoupling between them and the limbic systems (e.g., (Bos, Hermans, Ramsey, & Van Honk, 2012; Spielberg et al., 2014; Volman, Toni, Verhagen, & Roelofs, 2011).

Last, reduced cognitive reflection facilitates a context-sensitive decrease in the *probability* of engaging in slow and effortful cognition, while keeping the *capacity* to preform them intact, as demonstrated by our math control task. This feature mitigates the potential downside of reduced deliberate information processing when a decision-maker with elevated T faces a challenge that cannot be met by a rapid cognitive shortcut. Once a rapid and effortless action is available, however, the probability of translating it into action increases, and the time it takes to do so is shorter.

The hypothesis that T reduces cognitive reflection has many testable implications. Conditions known to elevate T, such as winning contests and presence of attractive mates, should reduce cognitive reflection. In tasks where rapid intuitions are useful (e.g., physical fighting) increased T will boost performance, and in situations where deliberation is needed T might impair performance. Furthermore, our findings provide a parsimonious explanation for the diverse, sometimes seemingly contradictory, behavioral influences of T reported in the literature. A bias towards intuitive responses could explain why T administration promotes aggressive responses to provocation (Carré et al., 2016; Pope, Kouri, & Hudson, 2000), decrease deception (van Honk et al., 2016; Wibral, Dohmen, Klingmüller, Weber, & Falk, 2012), and increase cooperation under some conditions (van Honk, Montoya, Bos, van Vugt, & Terburg, 2012) while decreasing it in others (Wright et al., 2012). Future investigation of T's influence on decision-making tasks should take into account the possibility that T might influence the decision-making *process*, rather than induce goal-directed status seeking per se (Eisenegger, Haushofer, & Fehr, 2011; Mazur, 2005).

Finally, our study has important public health implications. Western society has experienced an exogenous T 'shock' over the past decade from a rapidly growing T replacement therapy industry, with annual sales estimated at over $2B USD in 2013 (Von Drehle, 2014). The possibility that T might have deleterious influences on certain aspects of judgment and decision-making should be investigated further and taken into account by users, therapists, and policy makers.

References

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

Bing, Ola, Heilig, Markus, Kakoulidis, Pirros, Sundblad, Charlotta, Wiklund, Lisa, & Eriksson, Elias. (1998). High doses of testosterone increase anticonflict behaviour in rat. *European Neuropsychopharmacology, 8*(4), 321-323.

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.

Bosch-Domènech, Antoni, Brañas-Garza, Pablo, & Espín, Antonio M. (2014). Can exposure to prenatal sex hormones (2D: 4D) predict cognitive reflection? *Psychoneuroendocrinology, 43*, 1-10.

Brañas-Garza, Pablo, Kujal, Praveen, & Lenkei, Balint. (2015). Cognitive Reflection Test: Whom, how, when.

Campbell, Benjamin C, Dreber, Anna, Apicella, Coren L, Eisenberg, Dan TA, Gray, Peter B, Little, Anthony C, . . . Lum, J Koji. (2010). Testosterone exposure, dopaminergic reward, and sensation-seeking in young men. *Physiology & behavior, 99*(4), 451-456.

Cappelletti, Dominique, Güth, Werner, & Ploner, Matteo. (2011). Being of two minds: Ultimatum offers under cognitive constraints. *Journal of Economic Psychology, 32*(6), 940-950.

Carré, Justin M, Geniole, Shawn N, Ortiz, Triana L, Bird, Brian M, Videto, Amber, & Bonin, Pierre L. (2016). Exogenous testosterone rapidly increases aggressive behavior in dominant and impulsive men. *Biological psychiatry*.

Clancy, Andrew N, Bonsall, Robert W, & Michael, Richard P. (1992). Immunohistochemical labeling of androgen receptors in the brain of rat and monkey. *Life sciences, 50*(6), 409-417.

Cotrufo, Paolo, Monteleone, P, d'Istria, M, Fuschino, A, Serino, I, & Maj, M. (2000). Aggressive behavioral characteristics and endogenous hormones in women with bulimia nervosa. *Neuropsychobiology, 42*(2), 58-61.

Dabbs, James M, Carr, Timothy S, Frady, Robert L, & Riad, Jasmin K. (1995). Testosterone, crime, and misbehavior among 692 male prison inmates. *Personality and individual Differences, 18*(5), 627-633.

Daitzman, Reid, & Zuckerman, Marvin. (1980). Disinhibitory sensation seeking, personality and gonadal hormones. *Personality and Individual Differences, 1*(2), 103-110.

Edwards, David A. (1969). Early androgen stimulation and aggressive behavior in male and female mice. *Physiology & Behavior, 4*(3), 333-338.

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, 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.

Evans, Jonathan St BT. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences, 7*(10), 454-459.

Finley, SK, & Kritzer, MF. (1999). Immunoreactivity for intracellular androgen receptors in identified subpopulations of neurons, astrocytes and oligodendrocytes in primate prefrontal cortex. *Journal of neurobiology, 40*(4), 446-457.

Frederick, Shane. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 25-42.

Garcıa, Anton Aluja Luis F. (2005). Sensation seeking, sexual curiosity and testosterone in inmates. *Neuropsychobiology, 51*, 28-33.

Grimm, Veronika, & Mengel, Friederike. (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters, 111*(2), 113-115.

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.

Margittai, Zsofia, Nave, Gideon, Strombach, Tina, van Wingerden, Marijn, Schwabe, Lars, & Kalenscher, Tobias. (2015). Exogenous cortisol causes a shift from deliberative to intuitive thinking. *Psychoneuroendocrinology*.

Martin, Catherine A, Kelly, Thomas H, Rayens, Mary Kay, Brogli, Bethanie R, Brenzel, Allen, Smith, W Jackson, & Omar, Hatim A. (2002). Sensation seeking, puberty, and nicotine, alcohol, and marijuana use in adolescence. *Journal of the American academy of child & adolescent psychiatry, 41*(12), 1495-1502.

Mazur, Allan. (2005). *Biosociology of dominance and deference*: Rowman & Littlefield Publishers.

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.
Pope, Harrison G, Kouri, Elena M, & Hudson, James I. (2000). Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: a randomized controlled trial. *Archives of general psychiatry, 57*(2), 133-140.
Rand, David G, Greene, Joshua D, & Nowak, Martin A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427-430.
Reynolds, Maureen D, Tarter, Ralph, Kirisci, Levent, Kirillova, Galina, Brown, Stacy, Clark, Duncan B, & Gavaler, Judith. (2007). Testosterone levels and sexual maturation predict substance use disorders in adolescent boys: A prospective study. *Biological psychiatry, 61*(11), 1223-1227.
Ronay, Richard, & von Hippel, William. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Social Psychological and Personality Science, 1*(1), 57-64.
Roney, James R. (2016). Theoretical frameworks for human behavioral endocrinology. *Hormones and Behavior, 84*, 97-110.
Spielberg, Jeffrey M, Forbes, Erika E, Ladouceur, Cecile D, Worthman, Carol M, Olino, Thomas M, Ryan, Neal D, & Dahl, Ronald E. (2014). Pubertal testosterone influences threat-related amygdala–orbitofrontal cortex coupling. *Social cognitive and affective neuroscience*, nsu062.
Svensson, Anders I, Åkesson, Pernilla, Engel, Jörgen A, & Söderpalm, Bo. (2003). Testosterone treatment induces behavioral disinhibition in adult male rats. *Pharmacology Biochemistry and Behavior, 75*(2), 481-490.
Tinbergen, Niko. (1951). The study of instinct.
Toplak, Maggie E, West, Richard F, & Stanovich, Keith E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275-1289.
Toplak, Maggie E, West, Richard F, & Stanovich, Keith E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147-168.
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.
van Honk, Jack, Will, Geert-Jan, Terburg, David, Raub, Werner, Eisenegger, Christoph, & Buskens, Vincent. (2016). Effects of testosterone administration on strategic gambling in poker play. *Scientific reports, 6*.
Volman, Inge, Toni, Ivan, Verhagen, Lennart, & Roelofs, Karin. (2011). Endogenous testosterone modulates prefrontal–amygdala connectivity during social emotional behavior. *Cerebral Cortex*, bhr001.
Von Drehle, D. (2014). Manopause!? Aging, insecurity and the $2 billion testosterone industry. *Time, US Edition, 184*, 36-43.
Vrij, Aldert, Granhag, Pär Anders, Mann, Samantha, & Leal, Sharon. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science, 20*(1), 28-32.
Watson, David, & Clark, Lee Anna. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.

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.
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.

Acknowledgement

Funding for this work generously provided by Caltech, Ivey Business School, IFREE, Russell Sage Foundation, University of Southern California, INSEAD, and Stockholm School of Economics. Special thank you to Jorge Barraza, Austin Henderson, and Garrett Thoelen who assisted this study, David Kimball for LC-MS/MS assay testing, and Justine Carre' and James Roney for their insightful comments on earlier versions of the manuscript.

---

i Some of the cited studies have used saliva measurements of endogenous T. As correlations between saliva and blood T were shown to be high (at the order of .95), and as T is a small molecule that crosses the blood brain barrier, saliva T is considered reliable proxy of blood and brain levels.

ii The Pope et al. study administered supraphysiologic T dose over an extended period.

iii The T system is sexually dimorphic; therefore, it remains to be seen whether these findings extrapolate to males.

iv The effect depended on the right hand digit ratio (2D:4D), i.e., there was no main effect.

v A within-subject crossover design, in which both placebo and T are administered to each subject in two different sessions, could not be used because there is only a small bank of CRT items and their answers are likely to be easily remembered from one session to the next.

vi A recent meta-analysis shows that incentivizing the CRT has no impact on performance (Brañas-Garza, Kujal, & Lenkei, 2015). We incentivized the task to have real consequences to participant's decisions, make the study more realistic, improve attention, and eliminate noisy outlying responses.

vii The task was a part of a behavioral battery containing several unrelated tasks. The hypothesis test statistic remained significant after controlling for multiple hypothesis testing using a conservative Bonferroni correction ($p<0.015$, two tailed, corrected).

viii Simple arithmetic calculations are the type of "system 2" processes required for checking the correctness of the intuitive CRT responses, once their validity is challenged by cognitive reflection (e.g., adding the \$0.10 "intuitive" ball price and the \$1.10 bat price and seeing that they add up to \$1.20, rather than \$1.10).

ix Ethnologists have long defined five main types of animal behaviors as "instinctive": flight, freeze, feed, fight and mate (Tinbergen, 1951). The latter two are strongly linked with testosterone. Fittingly, the first two behaviors are associated with stress response that has also been linked to reduced system 2 information processing.

x The release of hormones into the bloodstream makes them ideal for coordinating between different information processing systems in the body and the brain, as opposed to neurotransmitters, that are released locally and affect information processing only a small number of target cells at the same time.