Adaptive Human Behavior and Physiology (2017) 3:365–385
DOI 10.1007/s40750-017-0079-7

ORIGINAL ARTICLE

Does Exogenous Testosterone Modulate Men's Ratings of Facial Dominance or Trustworthiness?

Brian M. Bird[1] · **Shawn N. Geniole**[2,3] · **Anthony C. Little**[4] · **Benjamin J. P. Moreau**[5] · **Triana L. Ortiz**[2] · **Bernard Goldfarb**[6] · **Pierre L. Bonin**[6] · **Justin M. Carré**[2]

Received: 17 May 2017 / Revised: 25 August 2017 / Accepted: 29 August 2017 /
Published online: 13 September 2017
© Springer International Publishing AG 2017

Abstract Previous research indicates that men's testosterone levels, or personality and contextual variables known to influence testosterone levels, predict men's attributions of social and personality characteristics from faces. However, the correlational nature of many of these past findings precludes our ability to establish causal pathways. Here, across two pharmacological challenge experiments, we examined the extent to which testosterone reduced men's perceptions of trustworthiness from emotionally-neutral faces (Experiment 1, $N = 30$, within-subjects design) or sensitivity to dominance from men's faces that varied in characteristically dominant shape (Experiment 2, $N = 117$, between-subjects design). Results from Experiment 1 showed that administration of testosterone did not significantly lower men's perceptions of trustworthiness. An unexpected order effect (i.e., drug x order of administration interaction) showed that trustworthiness ratings were higher after testosterone, but only if men received testosterone on the first day and placebo on the second day; importantly, this effect was

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40750-017-0079-7) contains supplementary material, which is available to authorized users.

✉ Justin M. Carré
justinca@nipissingu.ca

[1] Department of Psychology, Simon Fraser University, Burnaby, BC, Canada

[2] Department of Psychology, Nipissing University, North Bay, ON, Canada

[3] Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Vienna, Austria

[4] Department of Psychology, University of Bath, Bath, UK

[5] Northern Ontario School of Medicine, Thunder Bay, ON, Canada

[6] Northern Ontario School of Medicine, Sudbury, ON, Canada

directionally opposite to that reported in the literature and to that predicted for the present study. Experiment 2 demonstrated that dominance perceptions did not vary as a function of whether men received testosterone or placebo. Supplementary analyses with linear mixed effects generally support the main findings across experiments, but also provide more nuanced details involving exploratory individual difference variables. Results from the present experiments provide important information to a growing body of research examining testosterone and complex social processes, and may help inform future research on the topic.

Keywords Testosterone · Hormones · Trust · Trustworthiness · Dominance · Masculinity · Facial perception · Instrasexual competition · Sexual dimorphism · Rivalry

Introduction

It is well established that humans make personality and social attributions from faces, often rapidly (Todorov et al. 2015). Such reflexive judgments may be adaptive, in part, because they allow for a quick appraisal of strength (Fink et al. 2007; Sell et al. 2009), dominance (Oosterhof and Todorov 2008; Watkins et al. 2010a, b; Watkins and Jones 2012, 2016), aggression (Carré et al. 2009a), or mate value (Bird et al. 2016a; Little et al. 2011), which may help to reduce the potential costs of intrasexual competition (Watkins et al. 2010a; Watkins and Jones 2012) or increase the likelihood of reproducing with a quality mate (Little 2014). An accumulating body of evidence suggests that testosterone, or contextual and personality factors known to correlate with testosterone (e.g., winning versus losing a competition, trait dominance), may play an important role in modulating such attributions to facial stimuli, and particularly ratings of trustworthiness or dominance (Bos et al. 2010; Carré et al. 2014; Watkins et al. 2010a, b; Watkins and Jones 2012, 2016). Previous findings raise the possibility that one proximal mechanism through which testosterone affects mating and dominance behavior is by altering perceptions of threat gleaned from facial stimuli (e.g., rise in testosterone following competition, or exogenous testosterone administration, predicts reduced ratings of trustworthiness: Carré et al. 2014; Bos et al. 2010; contest outcomes and dispositional dominance predict reduced perceptions of dominance: Watkins and Jones 2012; Watkins et al. 2010b). However, previous work in this area has been either largely correlational (e.g., Carré et al. 2014), or in the case of experimental designs, conducted exclusively in women (e.g., Bos et al. 2010). The present experiments were designed to address this gap by using experimental methods to examine the potential influence of exogenous testosterone on men's perceptions of trustworthiness and dominance from facial stimuli.

A large body of research suggests that testosterone plays a key role in modulating social behavior in both humans and non-human species, and most notably for dominance and social status. For example, higher testosterone predicts higher status in male primates (e.g., chimpanzees: Muehlenbein et al. 2004; Muller and Wrangham 2004; baboons: Beehner et al. 2006; Sapolsky 1991), and in humans, testosterone concentrations have been found to positively correlate with trait measures of dominance (e.g., Carré et al. 2009b; Sellers et al. 2007; Stanton and

Schultheiss 2009; Turan et al. 2014), as well as dominance behavior intended to gain or maintain social status (Mazur and Booth 1998). Moreover, rapid fluctuations in testosterone in response to competitive situations predict variables related to status, such as subsequent aggression (Carré et al. 2009a, 2013), competitive motivation (Carré and McCormick 2008; Mehta and Josephs 2006), and risk preferences (Apicella et al. 2014; see Carré and Olmstead 2015, for review). Similarly, and depending on individual difference factors like trait dominance or self-control, elevating testosterone via pharmacological challenge also maps onto corresponding status-relevant behaviors, such as increased aggression (Carré et al. 2017) and competitive decision-making (Mehta et al. 2015).

Many findings are consistent with the challenge hypothesis as applied to humans (originally developed in avian species: Wingfield et al. 1990), positing that higher testosterone levels may facilitate mating effort, whereas levels tend to drop when competition is attenuated, such as with greater levels of parental investment (Gettler et al. 2011) or when men are in committed relationships (Burnham et al. 2003; Gray et al. 2004). Encompassing such findings is also the broader life history framework, which suggests that testosterone may mediate trade-offs (e.g., mating versus parenting) that are inherent in species with limited resources (e.g., time, energy; see Zilioli and Bird 2017, for review). Moreover, previous findings are also largely consistent with the biosocial model of status framework (Mazur 1976, 1985; Mazur and Booth 1998), which posits that testosterone fluctuations in response to status contests (e.g., the winner-loser effect; see Geniole et al. 2017, for meta-analysis) serve to fine tune ongoing status-related behaviors such as aggression.

The previous findings and theoretical considerations raise the question as to what psychological mechanisms underlie testosterone's relationship with mating or parenting effort, and associated competitive behaviors (e.g., aggression). One potential mechanism is that testosterone influences the degree to which humans infer trustworthiness and dominance from faces, which in turn could influence decisions to engage in competition and/or pursue particular mates. Indeed, some recent evidence supports the link between testosterone and ratings of facial stimuli on these dimensions. For example, in a sample of 24 women, Bos et al. (2010) found that a single administration of testosterone significantly decreased facial ratings of trustworthiness, and especially so among those participants who had high levels of trust in the first place. In a conceptually similar manner, Boksem et al. (2013) found that an administration of testosterone to healthy females reduced trust behavior in a one-shot trust game. Importantly, however, these studies were conducted exclusively in women, posing difficulty for generalizing results to men.

In more recent work, Carré et al. (2014) used a competitive paradigm to examine the extent to which endogenous fluctuations in testosterone mapped onto ratings of trustworthiness from emotionally-neutral faces. Findings from this work revealed that a rise in testosterone following competition predicted decreased ratings of trustworthiness from emotionally-neutral faces—an effect that was found in men, but not in women. Other work tacitly suggests that testosterone may influence the degree to which men are sensitive to cues of dominance in same-sex faces. For example, Watkins and Jones (2012) found that men who were

primed with winning a contest—a contextual variable often predicting testosterone increases for winners more so than losers (Geniole et al. 2017)—were less sensitive to facial cues of dominance than those primed with losing. Such findings are consistent with past work showing that men who are higher in dominance—a trait found to positively correlate with testosterone (e.g., Carré et al. 2009b; Sellers et al. 2007)—or who are taller (where height is positively related to indices of dominance; see Buunk et al. 2008) are less sensitive to cues of dominance (i.e., were less likely to rate more masculine faces as dominant) in men's faces (Watkins et al. 2010a, b). Some recent experimental work has also found that a single administration of testosterone can increase healthy young men's perceptions of their own physical dominance as indicated by participants choosing a more dominant version of themselves in a "pick-your-own-face" task (Welling et al. 2016). Taken together, these findings suggest that acute changes in testosterone, or contextual and personality variables known map onto variability in testosterone, may play a role in men's perceptions of trustworthiness and dominance in faces.

One notable gap in the literature is that previous work has not used experimental methods to examine how testosterone influences men's perceptions of trustworthiness and dominance in other faces, which is a key limitation in being able to establish causality in these relationships. The present study was designed to fill this gap by conducting experiments to address the following research questions: 1) Does testosterone influence men's ratings of trustworthiness from emotionally-neutral faces? 2) Does testosterone influence men's perceptions of dominance from same-sex faces that vary in characteristically dominant shape? Based on the literature discussed above, it was hypothesized that a single administration of testosterone (versus placebo) to healthy young men would

Fig. 1 Example of an emotionally-neutral male (left) and female (right) face used in Experiment 1. Images are from the Karolinska database (image ID BM31NES and AF01NES, respectively)

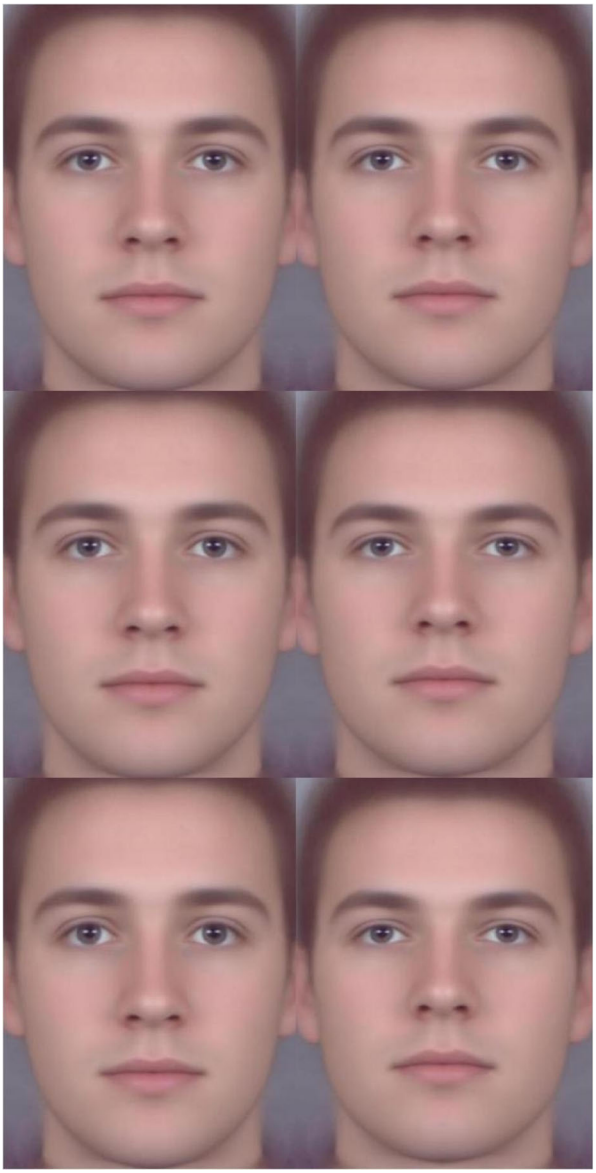

Fig. 2 Examples of images transformed for perceived dominance, with low perceived dominance (left) and high perceived dominance (right) for images transformed −/+ 10% (top), −/+ 25% (centre), and −/+ 50% (bottom). Participants saw transformed images of individual men. Composite images made by combining 50 male faces are presented here

predict 1) lower ratings of trustworthiness from emotionally-neutral faces, and 2) weaker sensitivity to differences in faces morphed to differing degrees of dominance.

Experiment 1: Ratings of Trustworthiness from Emotionally-Neutral Faces

Methods and Materials

Participants The participant sample consisted of 30 healthy young men (M_{age} = 21.21 years, SD = 2.19) who were part of a larger testosterone administration protocol. Self-reported participant ethnicities were Caucasian (n = 28), Latin American (n = 1), and Aboriginal / Indigenous (n = 1). Participants were recruited from a Northern Ontario Canadian university via the online research participation pool, posters displayed in common areas, or from communication with previous participants who agreed to be contacted for future studies. Prospective participants were first screened to determine eligibility based on the exclusionary criteria of 1) receiving prescription medication affecting hormone concentrations, 2) taking performance enhancing drugs, 3) diagnosis of a mental illness, 4) diagnosis of a heart condition, or 5) participation in an organized club or team where testosterone was a banned substance. All participants gave fully informed consent prior to the commencement of the protocol. Remuneration for participants consisted of $25 per hour, as well as partial course credit on the first day of testing (at the discretion of the individual course professors). Sample size was determined prior to data collection, and was in line with previous pharmacological challenge work (reviewed in Bos et al. 2012; see also Zilioli and Bird 2017). All procedures were approved by the Nipissing University Research Ethics Board.

Procedure The full protocol was conducted over three days. The first day of testing was a 1-h introductory session consisting of participant familiarization with the protocol, obtaining informed consent, and administration of demographic and self-report questionnaires. For separate research questions, participants also had their photographs taken (see Welling et al. 2016) as well as various anthropological measures (e.g., 2D:4D ratio, see Carré et al. 2015).

Hormone and Placebo Administration On the second day of testing, participants arrived at a local urology clinic for medical procedures, and first had 10 ml of blood taken from the antecubital area of the right arm to assess baseline levels of testosterone. Next, participants either received 150 mg of AndroGel®—a topical gel typically used for treating hypogonadism in men—or equivalent placebo, which were applied identically in both conditions to the upper arm and shoulder areas by a male research assistant who was blind to the experimental condition. Following administration, participants rested alone in a waiting room, and then had subsequent blood samples drawn (alternated between right and left arms) at 60 and 120 min post administration. Blood samples were allowed to clot, and were subsequently centrifuged at 3000 rpm, where serum was then extracted and stored at −60 °C until assayed. After their final blood draw, participants completed a series of computer-based tasks to assess attention (Hansen et al. 2017), social decision-making (Arnocky et al. 2016), cognition, and perception (Carré et al. 2015; Welling et al. 2016). Half of the participants provided trustworthiness ratings 2 h after drug administration, and the other half provided them 4 h after drug administration. Preliminary analyses indicated that time from drug

application (2 h vs. 4 h) did not interact with drug condition to predict ratings of trustworthiness (p = .79), and thus was not included in the statistical models. The time course was chosen based on previous pharmacokinetic work examining a 150 mg dose of AndroGel®, which found that testosterone levels begin to rise within about 2 h of administration, with peak concentrations reached at approximately 3 h post administration (Eisenegger et al. 2013). Additionally, other recent evidence suggests that a single administration of testosterone can rapidly modulate brain function within 45 to 90 min (Goetz et al. 2014; van Wingen et al. 2009). The third and final day of testing for Experiment 1 took place approximately two weeks after the second testing day, and was identical in nature to the second testing day, with the exception that participants received the opposite condition to which they received on the second day (i.e., participants receiving AndroGel® on day 2 then received placebo on day 3, and vice versa). Following testing on the third day, participants were asked to indicate which day they received testosterone, and a binomial test indicated they were no better than chance (p = .20).

Testosterone Concentrations Serum samples were assayed for total testosterone concentrations using commercially available enzyme immunoassay kits (DRG International) with an analytical sensitivity of .085 ng/mL. All samples were assayed in duplicate, and the average of these duplicates was taken for statistical analyses. Intra- and inter-assay coefficients of variation were below 6%.

Facial Ratings of Trustworthiness Task For this task, participants were asked to provide ratings of trustworthiness for 70 emotionally-neutral faces (35 male, 35 female) from the Karolinska Directed Emotional Faces database (Lundqvist et al. 1998). Each face was presented on a black background via E-Prime software, and was visible for 1000 ms, after which participants were asked to answer the following question: "How trustworthy is this person?" Participant answers were indicated on a keypad using a 7-point likert scale ranging from 1 = not at all, to 7 = very much. See Fig. 1.

Analytic Approach

Analyses first compared participants' ratings of trustworthiness as averaged across facial stimuli, using mixed factor ANOVAs in the Statistical Package for the Social Sciences (SPSS, Version 20; IBM, 2011). Linear mixed modelling analyses were also employed, allowing us to account for variation in average trustworthiness ratings, and in the effects of testosterone across the various facial identities and participants. Specifically, participants may vary in their ratings of trustworthiness on average (necessitating what is referred to as a "random intercept" in mixed level modelling, at the level of the participant) and stimulus faces may vary in the extent to which they are rated as trustworthy on average (necessitating a "random intercept" at the level of the stimulus faces). Further, the effect of testosterone (vs. placebo), or the effects of any of the other factors, on ratings of trustworthiness may vary across participants (necessitating what is referred to as a "random slope" in mixed level modelling, at the level of the participant) or across stimulus faces (necessitating a "random slope" at the level of the stimulus faces). In mixed level modelling, rather than averaging ratings across

stimulus faces or participants, we can model the variability and control for it in our analyses.

For linear mixed modeling, consistent with the recommendations of Barr et al. (2013), we conduct all of our analyses with random intercepts and slopes, such that for each fixed factor included in the model (as well as their interactions), we also include the corresponding random slopes (if interactions are involved, we only include those of the highest-order; Barr 2013), unless otherwise specified. In Experiment 1, we also model the random intercept (but not the slopes of the fixed effects) for the interaction of stimulus and participant (see Judd et al. 2017). All predictors were centered around zero, with any two-level factors of interest contrast coded with a one-unit distance from each other (e.g., drug condition: placebo = −0.5; testosterone = 0.5), and any continuous factors standardized. Therefore, the coefficients estimated from the models represent the difference in ratings of trustworthiness either between the two groups (for two-level factors) or associated with a 1 *SD* change in the predictor variable (for continuous factors) (in Experiment 2, when modeling the selection of the more dominant face in a pair, these *b* weights represent the changes in log odds). Analyses were conducted in R (version 3.3.2, R Core Team 2016), using the *lme4* package (version 1.1–13, Bates et al. 2015). *P*-values for linear mixed models were obtained using the package *car* (version 2.1.3, Fox and Weisberg 2011). When significant interactions emerged, we conducted follow-up analyses within separate subgroups or used the package *reghelper* (version 0.3.3, Hughes 2017) to conduct simple slopes analyses. This package provided *t*- but not *p*- values or degrees of freedom for simple slopes. We test the significance of the simple slopes using the degrees of freedom for the interaction term, or, if not provided (as in the *glmer* command), using the formula in Preacher, Curran, and Bauer (2006).

Results and Discussion

Testosterone Concentrations

Data for testosterone concentrations are reported elsewhere (see Bird et al. 2016a; Carré et al. 2015). Briefly, on the day participants received AndroGel®, testosterone concentrations were significantly higher than the day they received placebo, which was evident at both 60 min and 120 min post administration (*p*s < .001). As expected, the conditions did not differ in testosterone concentrations at baseline. Within two-hours of drug administration, serum testosterone concentrations increased by 56.29% and 7.50% after AndroGel® and placebo, respectively ($t(29) = 9.07$, $p < .001$, Cohen's $d = 1.78$).

Trustworthiness Ratings

A preliminary 2 X 2 mixed factor ANOVA [within subject factor: drug condition (Testosterone vs. Placebo); between subject factor: order of drug administration (Testosterone then Placebo vs. Placebo then Testosterone)] on ratings of trustworthiness revealed a marginal effect of drug condition, [F (1, 28) = 3.69, p = .07, η_p^2 = .12]. Unexpectedly, an order effect was observed (i.e., drug condition x order interaction), [$F(1, 28) = 8.92$, $p = .006$, $\eta_p^2 = .242$]. Examination of this order effect revealed that

ratings of trustworthiness differed between testosterone and placebo day among those who received testosterone the first day and placebo the second day ([F (1, 14) = 9.57, p = .008, η_p^2 = .41], but not among those who received placebo the first day and testosterone the second day [F (1, 14) = .765, p = .40, η_p^2 = .05]. Importantly, the difference between ratings of trustworthiness among those who received testosterone the first day and placebo the second day was directionally opposite to that predicted, such that when on testosterone, men in this group gave higher ratings of trustworthiness (M = 3.96, SD = 0.81) than when on placebo (M = 3.71, SD = .82)—a difference driving the marginal main effect of drug. Including the sex of the stimuli face as a within-subject factor in the model did not reveal any 3-way interaction (drug x sex of face x order: [F (1, 28) = .00, p = .99, η_p^2 = .00]), or 2-way interactions (drug x sex of face: [F (1, 28) = .01, p = .93, η_p^2 = .00]; sex of face x order: [F (1, 28) = .1.79, p = .19, η_p^2 = .06]). A main effect of sex of stimuli face was observed [F (1, 28) = .18.896, p < .001, η_p^2 = .40]) such that female faces (M = 4.09, SD = 0.76) were rated as significantly more trustworthy than male faces (M = 3.55, SD = 0.8) regardless of what drug or order participants received.

Linear mixed modeling revealed similar results. When drug condition (Testosterone vs. Placebo), order of drug administration (Testosterone then Placebo vs. Placebo then Testosterone), and the sex of the stimuli faces (male vs. female) were entered as predictors of ratings of trustworthiness, along with their interaction terms,[1] there was a main effect of drug condition (testosterone > placebo, b = 0.10, se = 0.027, χ^2 = 13.73, p < .001), which was moderated by the order of drug administration (b = 0.32, se = 0.055, χ^2 = 33.15, p < .001): those who received testosterone first and placebo second showed larger drug effects (higher ratings of trustworthiness after T: b = 0.26, se = 0.037, χ^2 = 49.72, p < .001) than did those who received placebo first and testosterone second (b = −0.06, se = 0.041, χ^2 = 1.89, p = .17). There was also a main effect of the sex of the stimulus face (female > male faces, b = −0.54, se = 0.133, χ^2 = 16.36, p < .001), which was also moderated by the order of drug administration (b = 0.33, se = 0.076, χ^2 = 18.92, p < .001): Female faces were rated as more trustworthy than male faces, especially by participants who were randomly assigned to receive placebo first and testosterone second (b = 0.71, se = 0.15, χ^2 = 21.70, p < .001) (vs. testosterone first and placebo second: b = 0.37, se = 0.15, χ^2 = 9.11, p = .003). See supplementary material for the effects of exploratory individual difference analyses.

Findings from Experiment 1 suggest that a single administration of testosterone (versus placebo) did not significantly lower men's perceptions of trustworthiness from emotionally-neutral faces. While the main effect of drug condition approached significance using a mixed ANOVA approach, and was significant using linear mixed modeling, the effect was directionally opposite to that reported elsewhere using correlational paradigms (Carré et al. 2014) such that overall, men here gave somewhat *higher* ratings of trustworthiness when on testosterone than when on placebo. However,

[1] Our mixed-level model predicting ratings of trustworthiness from drug condition, order of drug administration, sex of stimuli, and their interactions was specified as: lmer(Trust ~ DrugCondition * OrderAdministration * StimSex + (1 + DrugCondition: StimSex | ParticipantID) + (1 + DrugCondition: OrderAdministration | StimulusID) + (1 | ParticipantID: StimulusID), data = [DATAFILENAME], REML = FALSE). Note that fixed effects are those appearing after "~" but before the first bracket, whereas random effects are those terms that appear within the brackets after the "~".

an unexpected drug x order of administration interaction found with both analytic approaches suggested that a main effect (in the opposite direction of that predicted) was driven by increased ratings of trustworthiness among men who received testosterone on the first day and placebo on the second. The interpretation of such an effect is not clear, and thus requires a cautious approach. It is possible that carry-over effects might obscure any true difference in ratings of trustworthiness between testosterone and placebo day, such that effects of testosterone might be evident if it is the first time that participants make the judgments, but would not be evident if participants have already rated the faces (as the case for those receiving placebo first and testosterone second). Such an explanation may not be correct, however, as Carré et al. (2014) found changes in testosterone in response to competition predicted changes in trustworthiness ratings using the same set of stimuli on the same day. An important statistical consideration here is that when splitting analyses by drug order, the sample size is halved, making it possible that increased ratings of trustworthiness after testosterone (among those who received testosterone first and placebo second) is spurious. Also of consideration is that findings here conceptually contrast with other reports, where women receiving a single administration of testosterone showed decreased ratings of trustworthiness from emotionally neutral faces (Bos et al. 2010), in addition to decreases in trust behavior during a financial investment task (Boksem et al. 2013). Experiment 1 did not show that men had decreased ratings of trustworthiness after receiving testosterone, and if anything, there was a trend toward increased ratings of trustworthiness; however, given the presence of an order effect, this is interpreted not as convincing evidence for testosterone increasing trustworthiness, but rather a lack of evidence for testosterone decreasing trustworthiness, in contrast to our hypothesis.

It is important to note that previous testosterone administration studies in women using a sublingual preparation typically increase testosterone levels well above the normal physiological range (e.g., Bos et al. 2010). The dose used in Experiment 1 (150 mg) increased men's testosterone levels into the high normal range. While it is possible that higher doses of testosterone may have produced the predicted behavioral effects, such an increase in testosterone was not needed in previous work to elicit reduced ratings of trustworthiness from emotionally-neutral faces (Carré et al. 2014), and as noted, the direction of the effect was not in line with our hypothesis. Carré et al. (2014) found that testosterone responses to competition predicted reduced ratings of trustworthiness for men, but not women, suggesting that perhaps acute changes in hormones are better predictors of future behavior for men than women, which also contrasts with previous findings in which a single administration of testosterone reduced ratings of trustworthiness in women (Bos et al. 2010). One difference in Experiment 1 here versus Carré et al. (2014) is that there was no competitive condition in Experiment 1, which is perhaps notable, as competition related factors may play an important role in personality judgments from faces (e.g., Watkins and Jones 2012, 2016). In other words, an acute rise in testosterone may not be sufficient on its own to potentiate differences in trustworthiness perceptions, but rather may also rely on important contextual variables such as winning competitive encounters or experiencing threats (e.g., provocation). This idea is developed further in the general discussion.

It seems possible that acute changes in testosterone from pharmacological challenge may only predict ratings of more salient cues of threat, such as those indicating dominance. Experiment 2 was designed to address this question by examining the

effect of testosterone on men's perceptions of dominance in same-sex faces that vary in characteristically dominant shape.

Experiment 2: Ratings of Dominance from Men's Faces Varying in Dominant Shape

Methods and Materials

Participants

The participant sample consisted of 120 healthy young men between the ages of 18 and 35 (M_{age} = 25.27, SD = 4.98) who were part of a larger research protocol. Subjects were recruited to participate in a study at a medical research facility in Northern Ontario via advertising on local media sites, poster recruitment at local colleges and universities, as well as through the medical facility's research database. Similar to Experiment 1, each prospective participant underwent initial interviews to determine eligibility, with identical exclusionary criteria to that reported in Experiment 1. Self-reported participant ethnicities were primarily Caucasian (77.5%), followed by Aboriginal/Indigenous (13.1%), Asian (4.1%), and other (3.3%). Prior to testing, each participant provided written informed consent. Three participants did not complete the dominance task, rendering our final sample size 117 for Experiment 2. All procedures were approved by the corresponding university's research ethics board, and testing was done in accordance with the Declaration of Helsinki.

Stimuli

High and Low Perceived Dominance Composites Composite images were created starting from a set of 50 Caucasian male photographs with front-on pose, neutral expression, and under standard lighting conditions, which were randomly selected from a larger image set. Images were rated for dominance by 17 participants (11 men, 6 women, M_{age} = 26.4, SD = 4.3), who rated the faces on a 7-point Likert scale (1 being low dominance, and 7 being high dominance). Scores were averaged to provide a mean perceptual dominance score for each image, and the top and bottom scoring images were used to make high and low perceived dominance composites (high M = 4.73, SD = 0.37, low M = 3.06, SD = 0.42).

Each image was delineated with 179 landmark-points using Psychomorph (Tiddeman, Burt, & Perrett, 2001). Composite images were made by creating two average images made up of either the 15 top or 15 bottom scoring facial photographs (Benson and Perrett 1993; Tiddeman, Burt, & Perrett, 2001). Faces were aligned on interpupillary distance and made symmetrical before being used for transforms. See Fig. 2.

Transformed Perceived Dominance Images For transforms, 20 neutral, front-on Caucasian male face images were taken from the Radboud face set (Langner et al., 2010). Faces were delineated as described above and transformed in shape only using the linear differences between the landmark locations of the high perceived dominance

and low perceived dominance composite images (Rowland and Perrett 1995; Tiddeman et al. 2001). Starting from each base face, six transformed images were generated representing high (+50%, +25%, +10%) and low (−50%, −25%, −10%) perceived dominance. This process created 120 images in total which were paired as high and low perceived dominance images of corresponding levels (i.e., −50% vs. +50%, −20% vs. +20%, −10% vs. +10%) giving 60 image pairs in total and with 20 pairs at each of three levels of perceived dominance difference.

Procedure

Testing for the full protocol was conducted in a single session, where participants reported to the lab at one of two times (10:00 am or at 1:00 pm). Upon arrival, participants completed informed consent, had various anthropomorphic measurements taken, and completed a battery of online demographic and self-report individual difference variables as part of the larger protocol.

Hormone and Placebo Administration Following completion of the online questionnaires, participants received their initial blood draw, where a phlebotomist drew 10 mL of blood from the antecubital area of the arm. All blood samples were allowed to clot, were centrifuged at 3000 rpm, and subsequently had serum extracted and stored in −60 °C refrigeration until assayed. Following the initial blood draw, participants were randomized in a double blind fashion to receive either 150 mg of AndroGel® or equivalent placebo. In both conditions, a male research assistant who was blind to the condition applied topical gel to the upper arm and shoulder area. Following gel application, participants rested for one hour, and then received their second blood draw, after which they spent approximately one hour completing a series of computer-based tasks assessing social perception (Bird et al. 2016a), cognition, and decision-making abilities (Carré et al. 2017). The third and fourth blood draws were spread out across the remainder of the protocol, with the entirety of the protocol taking just over three hours. Testing for the facial dominance task occurred at approximately 2 h 45 min after gel administration. As in Experiment 1, at the end of the protocol, participants were asked if they were aware of which condition they were in; a binomial test determined they were at chance level for accuracy ($p = 1.0$).

Testosterone Concentrations Serum samples were assayed for total testosterone concentrations using commercially available enzyme immunoassay kits (DRG International) with an analytical sensitivity of .085 ng/mL. All samples were assayed in duplicate, and the average of these duplicates was taken for statistical analyses. Intra- and inter-assay coefficients of variation were 7.38% and 16.03%, respectively.

Facial Ratings of Dominance Task For their task, participants were asked to provide ratings of dominance to 60 pairs of men's faces. As described in the stimuli section, 20 men's faces were morphed to have either high or low dominance, at morph levels of 10%, 25%, or 50%. Thus, stimuli included 20 pairs consisting of 10% more dominant/10% less dominant, 20 pairs of 25% more dominant/25% less dominant, and 20 pairs of 50% more dominant/50% less dominant. Participants were tasked with selecting which

face in each pair they thought was more dominant, followed by indicating the degree to which the selected face was more dominant than the other face in the pair (i.e., "how much more dominant?"), with the following options: slightly more dominant, somewhat more dominant, more dominant, or much more dominant. The stimuli faces, the morph level pairs, and the side of the screen on which the more or less dominant version of the face appeared, were all randomized across participants.

The number of more dominant morph faces that the participant selected as the more dominant face in each pair was used as a total dominance score out of 20 for each morph level. The degree of dominance was coded in line with previous work investigating dominance perception (e.g., Watkins et al. 2010a; Watkins and Jones 2012): if the face selected as more dominant was the *less dominant morph*, responses were coded as much more dominant = 1, more dominant = 2, somewhat more dominant = 3, slightly more dominant = 4; if the face selected as more dominant was the *more dominant morph*, responses were coded as slightly more dominant = 5, somewhat more dominant = 6, more dominant = 7, and much more dominant = 8.

Analytic Approach

As with Experiment 1, analyses were conducted first with mixed factor ANOVAs in SPSS, comparing participants' average dominant face selections and degree of dominance ratings at each morph level. Further analyses were conducted using linear mixed modeling. For the morph level here, we ran two sets of regression analyses—one in which the 25% morph level was the reference group, and one in which the 10% morph level was the reference group. Linear mixed effects were once again conducted in R using the packages identified for Experiment 1. As noted earlier, for the selection of the more dominant face in a pair, the *b* weights represent the change in log odds.

Results

Testosterone Concentrations

As reported previously (Bird et al. 2016a; Carré et al. 2017), testosterone concentrations were significantly higher in the testosterone group relative to the placebo group at 60 mins, 75 mins, and 180 mins post-administration (all ps < .001). As expected, there were no differences in testosterone levels at baseline (p = .49).

Dominance Ratings

A 3 × 2 mixed factor ANOVA [within subject factor: Dominance Morph Level (10% vs. 25% vs. 50%); between subject factor: Drug condition (Testosterone vs. Placebo)] on the number of dominant faces selected as more dominant revealed no effect of drug, [F (1, 115) = 0.028, p = .868, η_p^2 = .00], a significant main effect of morph [F(2, 230) = 84.80, p < .001, η_p^2 = .42], but no drug condition x morph interaction [F(2, 230) = .532, p = .59, η_p^2 = .01]. Pairwise comparisons to probe the main effect of morph suggested that regardless of drug condition, and as would be expected,

participants selected as dominant more of the 50% dominant faces (M = 14.22, SE = 0.29) than they did the 25% dominant faces (M = 12.60, SE = 0.24; $p < .001$) and 10% dominant faces (M = 10.62, SE = 0.23; $p < .001$), and more of the 25% dominant faces than the 10% dominant faces ($p < .001$). A separate 3 × 2 mixed factor ANOVA (same independent variables as above) on the degree of dominance ratings similarly revealed no effect of drug condition, [F (1, 115) = 0.245, p = .622, η_p^2 = .00], a significant main effect of morph [F (2, 230) = 80.682, $p < .001$, η_p^2 = .41], but no drug condition x morph interaction [F(2, 230) = .408, p = .67, η_p^2 = .00]. Pairwise comparisons to probe the main effect of morph suggested that regardless of drug, participants attributed greater degree of dominance ratings to the 50% dominant faces (M = 5.27, SE = 0.057) than they did the 25% dominant faces (M = 4.91, SE = 0.039; $p < .001$) and 10% dominant faces (M = 4.58, SE = 0.035; $p < .001$), and more of the 25% dominant faces than the 10% dominant faces ($p < .001$).

Using a generalized linear mixed-effects model on the choice of face (dominant vs non-dominant version),[2] there was a main effect of the degree of manipulation, consistent with results above, such that participants more often chose the dominant face when the pairs differed by 25% versus 10% (b = 0.42, se = 0.069, χ^2 = 37.15, $p < .001$), 50% versus 25% (b = 0.42, se = 0.069, χ^2 = 36.69, $p < .001$), and 50% versus 10% (b = 0.84, se = 0.087, χ^2 = 91.89, $p < .001$), but there was no main effect of drug condition (b = −0.08, se = 0.083, χ^2 = 0.88, p = .35) or a drug condition by manipulation interaction (null effects for all contrasts: 25% vs 10%, p = .41; 50% vs 25%, p = .91; 50% vs 10%, p = .60).

A linear mixed-effects model on the degree to which the faces were rated as dominant also indicated a main effect of manipulation: there was a greater difference in perceived dominance between faces that were manipulated by 25% versus 10% (b = 0.34, se = 0.054, χ^2 = 38.50, $p < .001$), 50% vs 25% (b = 0.36, se = 0.057, χ^2 = 39.85, $p < .001$), and 50% vs 10% (b = 0.69, se = 0.068, χ^2 = 105.25, $p < .001$). There was no main effect of drug condition (b = −0.05, se = 0.072, χ^2 = 0.54, p = .46), or interactions between drug condition and the manipulations (null effects for all contrasts: 25% vs 10%, p = .48; 50% vs 25%, p = .43; 50% vs 10%, p = .92). See supplementary materials for the effects of individual differences.

General Discussion

Previous studies have suggested that one of the mechanisms through which testosterone might promote aggressive, competitive, and mate seeking behavior is through altering perceptions of others' trustworthiness (Bos et al. 2010; Carré et al. 2014) or dominance of self (Welling et al. 2016) and others (Watkins and Jones 2012). The two experiments reported in this paper are the first to test the causal effects of testosterone on men's ratings of trustworthiness in same- and opposite-sex faces, and men's sensitivity to cues of dominance in same-sex faces. Experiment 1 found that an acute rise in testosterone

[2] We use the following model specifications: glmer(Dominance_SelectAface ~ DrugCondition * ManipulationDegree + (1 + ManipulationDegree | ParticipantID) + (1 + DrugCondition: ManipulationDegree | StimulusID), family = binomial(link = "logit"), nAGQ = 0, control = glmerControl (optimizer = "nloptwrap"), data = s2final). The "nAGQ = 0" and "optimizer = "nloptwrap" commands were used to increase the speed of model fitting.

from exogenous administration did not significantly lower men's ratings of trustworthiness from emotionally-neutral faces when compared to the same men rating faces after receiving placebo. An unpredicted effect showed that drug interacted with order of drug administration to predict ratings of trustworthiness: specifically, testosterone predicted *higher* ratings of trustworthiness, but only for those who received testosterone on the first day. As noted in the discussion for Experiment 1, such a finding should be interpreted with caution for at least two reasons: 1) the effect was not predicted, and thus we have no immediate theoretical explanation, and 2) splitting analyses by order significantly reduces the sample size, increasing the chance that such an effect is spurious.

Experiment 2 found that, using a between-subjects design, an acute rise in testosterone did not significantly alter men's perceptions of dominance from same-sex faces that varied in characteristically dominant shape. While the findings from both experiments did not support our hypotheses, they nevertheless provide a critical test of the potential causal relationship between testosterone and men's perceptions of important social and personality variables as gleaned from faces, contributing to an important and growing body of literature examining relationships between hormones and attributions to facial stimuli. Moreover, and more generally, the contributions of null findings on previously found (or theorized) relationships are important for advancement in the field (e.g., Bird et al. 2016b; Kandrik et al. 2017) and will help to inform future studies on the subject.

The findings in the present study raise the question as to what might account for the differences between these and previously reported findings. When compared to studies showing that testosterone administration can reduce ratings of trustworthiness from faces (Bos et al. 2010), two key differences should be noted. Firstly, in their study, Bos et al. (2010) exclusively examined women, and secondly, they found that reduced perceptions of trustworthiness were exclusive to women who had high levels of trust in the first place. As previously mentioned, studies examining exogenous testosterone in women have typically used doses that increase testosterone levels above those attainable in any natural setting. While undoubtedly useful for examining causal effects of hormones, the generalizability of the results from such studies can become somewhat nebulous when considering the levels that testosterone typically reaches in humans.

Given the methodological considerations noted above (i.e., sex, dose), it is possible that the differences in the present study are a result of using men instead of women, and/or using a dose that raises men' s testosterone levels into the high normal range, versus the supraphysiological range. Such hard-and-fast distinctions seem somewhat unlikely, though, as this same dose (150 mg) and population of men has shown effects in other facial perception tasks such as men's ratings of women's attractiveness as short- or long-term mates (Bird et al. 2016a), as well as inferring cognitive states from photographs of the eye region (Carré et al. 2015). It should be noted, however, that the previous work in men found that the relationship between testosterone and ratings from faces depended on certain individual difference or contextual variables (e.g., short-or long-term mating context, Bird et al. 2016a; 2D:4D ratio and psychopathic traits, Carré et al. 2015). Including some of the most relevant individual difference variables in exploratory analyses, however, left the findings in the present study effectively unchanged (with some exceptions; see supplementary materials).

Although the findings presented here contrast with studies examining exogenous testosterone in women, they also contrast with previous correlational findings in men, showing that testosterone responses to competition predicted reduced ratings of trustworthiness from emotionally-neutral faces (Carré et al. 2014). Perhaps one of the most notable distinctions between the two studies is that Carré and colleagues used a competitive paradigm where some individuals won the competition and others lost, whereas this manipulation was not used in Experiment 1 here. Recent evidence also suggests that acute stress may be associated with reduced perceptions of trustworthiness from racial "outgroup" faces, perhaps because acute stress increases vigilance to threat (Salam et al., 2017). Given that competitive situations may be considered at least mildly stressful (e.g., Buckert et al., 2017), and perhaps more so under competitive conditions involving provocation, this may be another reason why reduced ratings of trustworthiness were previously found within a competitive context, but not within the pharmacological challenge here (but see, for example, Prasad et al., 2017, for evidence that acute stress blocks behavioral effects of testosterone; see also Mehta & Prasad, 2015). It is possible, then, that any relationship between men's testosterone and ratings of trustworthiness may be contextually-dependent (or psychologically-dependent, as in the case of subjective acute stress; Salam et al., 2017), but such a dependent relationship will require further testing.

The null findings in Experiment 2 are also worth discussion. Previous research suggests that priming men with different contest outcomes can influence perceptions of dominance from same sex faces, such that those primed with losing show greater sensitivity to dominance than those primed with winning (Watkins and Jones 2012). While not explicitly tested in that research, another body of work shows that on the whole, winners of competitions will show a rise in testosterone relative to losers (Geniole et al., 2017). The absence of a contest manipulation in the present study precludes the direct comparison to previous work, but the results of Experiment 2, using a relatively large sample of healthy young men, suggest that on its own, testosterone in the high normal range does not appear sufficient to alter men's perceptions of dominance from faces that vary in characteristically dominant shape. Moreover, when controlling for previously noted moderating variables potentially influencing men's perceptions of other men's dominance, results in Experiment 2 were largely the same. One possibility is that testosterone has more of an effect on men's perceptions of their own dominance (see Welling et al. 2016) than it does on perceptions of other men's dominance, but this possibility will require a direct test in future studies.

Limitations and Future Directions

Although the present experiments found no evidence for a significant influence of testosterone on men's reduced perceptions of trustworthiness or reduced sensitivity to dominance from others' faces, the possibility still exists that testosterone is an important variable for facial perception in both men and/or women. Should such a relationship exist, it may depend on other important contextual variables, such as being primed with exposure to opposite sex faces (e.g., Watkins et al. 2013), or with winning or

losing a dominance competition (Carré et al. 2014; Watkins and Jones 2012, 2016), which was not tested in the present study. Indeed, in addition to research showing contextual effects on ratings of trustworthiness or dominance, other work shows that contest-related outcomes can also influence men's perceptions of facial attractiveness (Welling et al. 2013). Given the focus on the association between competitive outcomes and sensitivity to cues in same-sex or opposite-sex faces, future research may benefit from combining pharmacological challenge with contextual contest manipulations. In line with this, some recent evidence in women suggests that testosterone administration may affect competitive decision-making in more complex interactions between victory-defeat experience and individual differences in trait dominance (Mehta et al. 2015).

Another consideration is the timeline of testosterone administration and behavioral testing. Previous investigations in women have typically assessed behavior at approximately 3 to 4 h after peak concentrations (Zilioli and Bird 2017). However, other investigations in men have found that the assessment of behavior at much earlier time points (e.g., 1–3 h post-administration) produces behavioral effects, including impacts on face perception (see Bird et al. 2016a; Carré et al. 2015) and aggressive behavior (Carré et al. 2017), suggesting that testosterone may act through a relatively-rapid, non-genomic mechanism (Foradori et al. 2008). However, this does not preclude the possibility that testosterone may affect perceived trustworthiness or dominance at much later time points (e.g., several hours, or on a subsequent day) via binding to the intracellular androgen receptor, and modulating transcription of target genes (i.e., the classical genomic model; briefly discussed in Rahman and Christian 2007; see also Foradori et al. 2008). In line with this idea is work in animal models showing that injections producing an acute rise in testosterone following a victory can influence aggressive behavior more than 24 h after the original rise in testosterone (Trainor et al. 2004). In the present study, behavior was assessed at one time point within a few hours of administration, and thus future investigations on trust and dominance (or other forms of human perceptual and behavior phenomena) may seek to assess behavior at later time points to test this potential.

The present study focused exclusively on men. Although using men may be considered a strength in that it expands a growing body of administration work done exclusively with women, it is simultaneously limited in that it remains unclear if similar methods would affect women in the same manner. Where possible (i.e., in countries where testosterone is not restricted for use in women), it may be useful to study the effects of testosterone administration simultaneously on men and women.

Conclusion

In summary, the present study found that in two experiments, utilizing randomized, placebo-controlled designs, administration of exogenous testosterone to healthy young men did not reduce their perceptions of trustworthiness from emotionally-neutral faces, or alter perceptions of dominance from faces that varied in dominant shape. Future research may seek to disentangle the potential independent or interactive effects of context (e.g., winning or losing a competition, exposure to opposite-sex individuals) with testosterone dynamics, and perceptions of social and personality variables from same- or opposite-sex faces.

Funding This work was supported by a Joseph-Armand Bombardier CGS Doctoral Scholarship to BMB, as well as a Natural Sciences and Engineering Research Council of Canada Discovery Grant to JMC, and a Northern Ontario Heritage Fund Corporation Grant to JMC.

Compliance with Ethical Standards

Conflict of Interest The authors declare no conflicts of interest.

References

Apicella, C. L., Dreber, A., & Mollerstrom, J. (2014). Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. *Psychoneuroendocrinology, 39*, 58–64. https://doi.org/10.1016/j.psyneuen.2013.09.025.

Arnocky, S., Taylor, S. M., Olmstead, N. A., & Carré, J. M. (2016). The effects of exogenous testosterone on men's moral decision-making. *Adaptive Human Behavior and Physiology,* 1–13. doi:https://doi.org/10.1007/s40750-016-0046-8.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*, 328.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, B. B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Beehner, J. C., Bergman, T. J., Cheney, D. L., Seyfarth, R. M., & Whitten, P. L. (2006). Testosterone predicts future dominance rank and mating activity among male chacma baboons. *Behavioral Ecology and Sociobiology, 59*, 469–479. https://doi.org/10.1007/s00265-005-0071-2.

Benson, P. J., & Perrett, D. I. (1993). Extracting prototypical facial images from exemplars. *Perception, 22*(3), 257–262. https://doi.org/10.1068/p220257.

Bird, B. M., Welling, L. L., Ortiz, T. L., Moreau, B. J., Hansen, S., Emond, M., et al. (2016a). Effects of exogenous testosterone and mating context on men's preferences for female facial femininity. *Hormones and Behavior, 85*, 76–85. https://doi.org/10.1016/j.yhbeh.2016.08.003.

Bird, B. M., Jofré, V. S. C., Geniole, S. N., Welker, K. M., Drug, S., Maestripieri, D., et al. (2016b). Does the facial width-to-height ratio map onto variability in men's testosterone concentrations? *Evolution and Human Behavior, 37*(5), 392–398. https://doi.org/10.1016/j.evolhumbehav.2016.03.004.

Boksem, M. A., Mehta, P. H., Van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., et al. (2013). Testosterone inhibits trust but promotes reciprocity. *Psychological Science, 24*(11), 2306–2314. https://doi.org/10.1177/0956797613495063.

Bos, P. A., Terburg, D., & van Honk, J. (2010). Testosterone decreases trust in socially naive humans. *Proceedings of the National Academy of Sciences, 107*(22), 9991–9995. https://doi.org/10.1073/pnas.0911700107.

Bos, P. A., Panksepp, J., Bluthé, R. M., & van Honk, J. (2012). Acute effects of steroid hormones and neuropeptides on human social–emotional behavior: a review of single administration studies. *Frontiers in Neuroendocrinology, 33*(1), 17–35. https://doi.org/10.1016/j.yfrne.2011.01.002.

Burnham, T. C., Chapman, J. F., Gray, P. B., McIntyre, M. H., Lipson, S. F., & Ellison, P. T. (2003). Men in committed, romantic relationships have lower testosterone. *Hormones and Behavior, 44*(2), 119–122. https://doi.org/10.1016/S0018-506X(03)00125-9.

Buunk, A. P., Park, J. H., Zurriaga, R., Klavina, L., & Massar, K. (2008). Height predicts jealousy differently for men and women. *Evolution and Human Behavior, 29*(2), 133–139. https://doi.org/10.1016/j.evolhumbehav.2007.11.006.

Carré, J. M., & McCormick, C. 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. https://doi.org/10.1016/j.yhbeh.2008.04.008.

Carré, J. M., & Olmstead, N. A. (2015). Social neuroendocrinology of human aggression: examining the role of competition-induced testosterone dynamics. *Neuroscience, 286*, 171–186.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009a). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20*, 1194–1198. https://doi.org/10.1111/j.1467-9280.2009.02423.x.

Carré, J. M., Putnam, S. K., & McCormick, C. M. (2009b). Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. *Psychoneuroendocrinology, 34*(4), 561–570. https://doi.org/10.1016/j.psyneuen.2008.10.018.

Carré, J. M., Campbell, J. A., Lozoya, E., Goetz, S. M., & Welker, K. M. (2013). Changes in testosterone mediate the effect of winning on subsequent aggressive behaviour. *Psychoneuroendocrinology, 38*(10), 2034–2041. https://doi.org/10.1016/j.psyneuen.2013.03.008.

Carré, J. M., Baird-Rowe, C. D., & Hariri, A. R. (2014). Testosterone responses to competition predict decreased trust ratings of emotionally neutral faces. *Psychoneuroendocrinology, 49*, 79–83. https://doi.org/10.1016/j.psyneuen.2014.06.011.

Carré, J. M., Ortiz, T. L., Labine, B., Moreau, B. J., Viding, E., Neumann, C. S., & Goldfarb, B. (2015). Digit ratio (2D: 4D) and psychopathic traits moderate the effect of exogenous testosterone on socio-cognitive processes in men. *Psychoneuroendocrinology, 62*, 319–326. https://doi.org/10.1016/j.psyneuen.2015.08.023.

Carré, J. M., Geniole, S. N., Ortiz, T. L., Bird, B. M., Videto, A., & Bonin, P. L. (2017). Exogenous testosterone rapidly increases aggressive behavior in dominant and impulsive men. *Biological Psychiatry, 82*, 249–256.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Eisenegger, C., von Eckardstein, A., Fehr, E., & von Eckardstein, S. (2013). Pharmacokinetics of testosterone and estradiol gel preparations in healthy young men. *Psychoneuroendocrinology, 38*(2), 171–178. https://doi.org/10.1016/j.psyneuen.2012.05.018.

Fink, B., Neave, N., & Seydel, H. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology, 19*(1), 82–87. https://doi.org/10.1002/ajhb.20583.

Foradori, C. D., Weiser, M. J., & Handa, R. J. (2008). Non-genomic actions of androgens. *Frontiers in Neuroendocrinology, 29*(2), 169–181. https://doi.org/10.1016/j.yfrne.2007.10.005.

Fox, J., & Weisberg, S. (2011). *An an {R} companion to applied regression* (5th ed.). Thousand Oaks: Sage.

Geniole, S. N., Bird, B. M., Ruddick, E. L., & Carré, J. M. (2017). Effects of competition outcome on testosterone concentrations in humans: an updated meta-analysis. *Hormones and Behavior, 92*, 37–50. https://doi.org/10.1016/j.yhbeh.2016.10.002.

Gettler, L. T., McDade, T. W., Feranil, A. B., & Kuzawa, C. W. (2011). Longitudinal evidence that fatherhood decreases testosterone in human males. *Proceedings of the National Academy of Sciences, 108*(39), 16194–16199. https://doi.org/10.1073/pnas.1105403108.

Goetz, S. M., Tang, L., Thomason, M. E., Diamond, M. P., Hariri, A. R., & Carré, J. M. (2014). Testosterone rapidly increases neural reactivity to threat in healthy men: a novel two-step pharmacological challenge paradigm. *Biological Psychiatry, 76*(4), 324–331. https://doi.org/10.1016/j.biopsych.2014.01.016.

Gray, P. B., Chapman, J. F., Burnham, T. C., McIntyre, M. H., Lipson, S. F., & Ellison, P. T. (2004). Human male pair bonding and testosterone. *Human Nature, 15*(2), 119–131. https://doi.org/10.1007/s12110-004-1016-6.

Hansen, S., McAuliffe, J., Goldfarb, B., & Carré, J. M. (2017). Testosterone influences volitional, but not reflexive orienting of attention in human males. *Physiology & Behavior, 175*, 82–87.

Hughes, J. (2017). Reghelper: helper functions for regression analysis. R package, version 0.3.3. https://CRAN.R-project.org/package=reghelper.

IBM Corp. (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: designs, analytic models, and statistical power. *Annual Review of Psychology, 68*, 601–625.

Kandrik, M., Hahn, A. C., Han, C., Wincenciak, J., Fisher, C. I., DeBruine, L. M., & Jones, B. C. (2017). Does the interaction between cortisol and testosterone predict men's facial attractiveness? *Adaptive Human Behavior and Physiology*, 1–7. https://doi.org/10.1007/s40750-017-0064-1.

Little, A. C. (2014). Facial attractiveness. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(6), 621–634. https://doi.org/10.1002/wcs.1316.

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 366*(1571), 1638–1659. https://doi.org/10.1098/rstb.2010.0404.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

Mazur, A. (1976). Effects of testosterone on status in primate groups. *Folia Primatologica, 26*(3), 214–226.

Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces, 64*, 377–402.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*(03), 353–363.

Mehta, P. H., & Josephs, R. A. (2006). Testosterone change after losing predicts the decision to compete again. *Hormones and Behavior, 50*(5), 684–692. https://doi.org/10.1016/j.yhbeh.2006.07.001.

Mehta, P. H., van Son, V., Welker, K. M., Prasad, S., Sanfey, A. G., Smidts, A., & Roelofs, K. (2015). Exogenous testosterone in women enhances and inhibits competitive decision-making depending on victory–defeat experience and trait dominance. *Psychoneuroendocrinology, 60*, 224–236. https://doi.org/10.1016/j.psyneuen.2015.07.004.

Muehlenbein, M. P., Watts, D. P., & Whitten, P. L. (2004). Dominance rank and fecal testosterone levels in adult male chimpanzees (Pan troglodytes schweinfurthii) at Ngogo, Kibale National Park, Uganda. *American Journal of Primatology, 64*(1), 71–82. https://doi.org/10.1002/ajp.20062.

Muller, M. N., & Wrangham, R. W. (2004). Dominance, aggression and testosterone in wild chimpanzees: a test of the 'challenge hypothesis. *Animal Behaviour, 67*(1), 113–123. https://doi.org/10.1016/j.anbehav.2003.03.013.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*(4), 437–448. https://doi.org/10.3102/1076998603100443.

Rahman, F., & Christian, H. C. (2007). Non-classical actions of testosterone: an update. *Trends in Endocrinology and Metabolism, 18*(10), 371–378. https://doi.org/10.1016/j.tem.2007.09.004.

Rowland, D. A., & Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications, 15*(5), 70–76.

Sapolsky, R. M. (1991). Testicular function, social rank and personality among wild baboons. *Psychoneuroendocrinology, 16*(4), 281–293.

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society of London B: Biological Sciences, 276*(1656), 575–584. https://doi.org/10.1098/rspb.2008.1177.

Sellers, J. G., Mehl, M. R., & Josephs, R. A. (2007). Hormones and personality: testosterone as a marker of individual differences. *Journal of Research in Personality, 41*(1), 126–138. https://doi.org/10.1016/j.jrp.2006.02.004.

Stanton, S. J., & Schultheiss, O. C. (2009). The hormonal correlates of implicit power motivation. *Journal of Research in Personality, 43*(5), 942–949. https://doi.org/10.1016/j.jrp.2009.04.001.

Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Prototyping and transforming facial texture for perception research. *IEEE Computer Graphics and Applications, 21*, 42–50. https://doi.org/10.1109/38.946630.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831.

Trainor, B. C., Bird, I. M., & Marler, C. A. (2004). Opposing hormonal mechanisms of aggression revealed through short-lived testosterone manipulations and multiple winning experiences. *Hormones and Behavior, 45*(2), 115–121. https://doi.org/10.1016/j.yhbeh.2003.09.006.

Turan, B., Guo, J., Boggiano, M. M., & Bedgood, D. (2014). Dominant, cold, avoidant, and lonely: basal testosterone as a biological marker for an interpersonal style. *Journal of Research in Personality, 50*, 84–89. https://doi.org/10.1016/j.jrp.2014.03.008.

Watkins, C. D., & Jones, B. C. (2012). Priming men with different contest outcomes modulates their dominance perceptions. *Behavioral Ecology*. https://doi.org/10.1093/beheco/arr221.

Watkins, C. D., Fraccaro, P. J., Smith, F. G., Vukovic, J., Feinberg, D. R., DeBruine, L. M., & Jones, B. C. (2010a). Taller men are less sensitive to cues of dominance in other men. *Behavioral Ecology, 21*(5), 943–947. https://doi.org/10.1093/beheco/arq091.

Watkins, C. D., Jones, B. C., & DeBruine, L. M. (2010b). Individual differences in dominance perception: dominant men are less sensitive to facial cues of male dominance. *Personality and Individual Differences, 49*(8), 967–971.

Watkins, C. D., Debruine, L. M., Feinberg, D. R., & Jones, B. C. (2013). A sex difference in the context-sensitivity of dominance perceptions. *Evolution and Human Behavior, 34*(5), 366–372. https://doi.org/10.1016/j.evolhumbehav.2013.06.004.

Welling, L. L., Persola, L., Wheatley, J. R., Cárdenas, R. A., & Puts, D. A. (2013). Competition and men's face preferences. *Personality and Individual Differences, 54*(3), 414–419. https://doi.org/10.1016/j.paid.2012.10.014.

Welling, L. L., Moreau, B. J., Bird, B. M., Hansen, S., & Carré, J. M. (2016). Exogenous testosterone increases men's perceptions of their own physical dominance. *Psychoneuroendocrinology, 64*, 136–142. https://doi.org/10.1016/j.psyneuen.2015.11.016.

van Wingen, G. A., Zylicz, S. A., Pieters, S., Mattern, C., Verkes, R. J., Buitelaar, J. K., & Fernández, G. (2009). Testosterone increases amygdala reactivity in middle-aged women to a young adulthood level. *Neuropsychopharmacology, 34*(3), 539–547. https://doi.org/10.1038/npp.2008.2.

Watkins, C. D., & Jones, B. C. (2016). Competition-related factors directly influence preferences for facial cues of dominance in allies. *Behavioral Ecology and Sociobiology, 70*(12), 2071–2079. https://doi.org/10.1007/s00265-016-2211-2

Wingfield, J. C., Hegner, R. E., Dufty Jr., A. M., & Ball, G. F. (1990). The "challenge hypothesis": theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. *The American Naturalist, 136*(6), 829–846.

Zilioli, S., & Bird, B. M. (2017). Functional significance of men's testosterone reactivity to social stimuli. *Frontiers in Neuroendocrinology*. https://doi.org/10.1016/j.yfrne.2017.06.002.