

Running head: TESTOSTERONE AND PERSONALITY STATES IN COMPETITION

**Effects of male testosterone and its interaction with cortisol on self- and observer-rated
personality states in a competitive mating context**

Tobias L. Kordsmeyer* & Lars Penke

Department of Psychology & Leibniz ScienceCampus Primate Cognition

University of Goettingen

Gosslerstr. 14

37073 Goettingen, Germany

*Corresponding author.

E-mail address: tob.kor@gmail.com, phone: +49 551 3920706

Author notes:

Author contributions: Both TLK and LP designed the study and wrote the manuscript. TLK collected and analysed the data.

Funding: We acknowledge support by the Leibniz Association through funding for the Leibniz ScienceCampus Primate Cognition.

Declarations of interest: The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Acknowledgement: We would like to thank Nicole Storp and Yasmin Thies for helpful comments on an earlier draft of this article and Mira Hallmann for help in data collection.

**Effects of male testosterone and its interaction with cortisol on self- and observer-rated
personality states in a competitive mating context**

Abstract

Increases in men's testosterone (T) levels after intrasexual competitions and exposure to females facilitate competitive and courtship behaviours, suggesting T reactivity should affect relevant personality state changes. How exactly T reactivity, also under potential buffering effects of Cortisol (C), relates to personality state changes is unclear. In a preregistered study, we aimed at inducing T increases in young men ($N=165$) through dyadic intrasexual competitions while exposed to a female experimenter. We investigated self-reported and video-based observer-rated personality state changes, captured by the Interpersonal Circumplex and social impressions, in relation to hormonal levels. Results revealed increases in self-reported competitiveness and observer-rated self-assurance, relative to a control group, moderated by T reactivity and partly by TxC interactions, providing insights into hormone-personality response-links.

Keywords testosterone, cortisol, male competition, female exposure, interpersonal circumplex

1. Introduction

Humans compete for access to mates and social status in order to reproduce and eventually reach higher biological fitness (Puts, 2016). A wide range of fine-grained mechanisms have evolved over human evolutionary history to support pursuing these goals. One important mechanism implicated here is the endocrinological system. Hormones act as physiological coordinators, influencing several organs and processes simultaneously, including human perception and behavioural dispositions (Roney, 2016). The steroid hormone testosterone (T) plays a key role in human competitive behaviour. It has been shown to partly mediate a trade-off between mating and parenting effort (Muehlenbein & Bribiescas, 2005), in particular by increasing aggression and risk-taking at the expense of survival and nurturing behaviour (Muller, 2017). T has further been suggested to modulate social signalling in intrasexual (dominance contests) and intersexual (mate attraction) contexts (Puts, 2010). T levels are also related to relationship status and parenthood, in that lower T has been found in partnered, compared to single, individuals (van Anders & Watson, 2006), and in fathers relative to non-fathers (Gettler, McDade, Feranil, & Kuzawa, 2011). T is particularly relevant in men compared to women, presumably due to different trade-offs concerning reproductive strategies and hence intrasexual competition being more prevalent (Archer, 2009; Puts et al., 2015). If T responses to competitive and mating-related situations adaptively trigger relevant behavioural tendencies and social signals, they may mediate changes in corresponding personality states (i.e., temporary characteristics of how an individual behaves, feels and thinks in a given situation; e.g., Geukes, Nestler, Hutteman, Küfner, & Back, 2017), which may also be accurately detectable by observers. In the present preregistered study, we seek to clarify the role of T and its effects on personality states in an intrasexually competitive situation among men.

1.1. Challenge hypothesis and female exposure studies

According to the challenge hypothesis, which was originally proposed for seasonally breeding birds (Wingfield, Hegner, Dufty, & Ball, 1990) and later applied to humans (Archer 2006), there is an acute increase in T levels in the face of various kinds of challenges, such as mating opportunities or

intrasexual competition, inducing corresponding behavioural tendencies, such as aggressiveness, competitiveness and courtship behaviour (for recent discussions, see Carré & Archer, 2018; Wingfield, 2017). The challenge hypothesis stems from the field of behavioural ecology, where acute changes (such as T reactivity) are termed plasticity (Dingemanse, Kazem, Reale, & Wright, 2010) and correspond to state changes in personality psychology. One contemporary domain where such T reactivity becomes apparent in humans is sports competitions. A T increase in male competitors has been shown in anticipation of, during and/or after engaging in sports competitions such as soccer ($N = 40$; Edwards, Wetzel, & Winer, 2006), Japanese chess ($N = 90$; Hasegawa, Toda, & Morimoto, 2008), or wrestling ($N = 15$; Elias, 1981). In some studies, T reactivity was shown to be higher in winners compared to losers of sports competitions, which has been termed the "winner effect". Although there is a mixture of positive and negative findings, a recent meta-analysis found overall evidence for larger T increases in winners relative to losers ($k = 60$ effect sizes, overall $N > 2500$ men and women; Geniole, Bird, Ruddick, & Carré, 2017).

Another setting where rapid increases in men's T levels have been shown are mating opportunities, or so-called "female exposures". A simple interaction with an attractive woman as short as five minutes can suffice to raise T in men (free T usually measured from saliva; Fiers et al., 2014). For example, a significant increase in T was shown in men ($N = 37$) after engaging in a short conversation with a young woman (Roney, Mahler, & Maestriperieri, 2003). No increase was detected in a control condition, in which the participants interacted with a male experimenter. In a similar study (van der Meij, Buunk, van de Sande, & Salvador, 2008), a T increase was found in men ($N = 30$) after interacting with a female confederate, and T reactivity was stronger in those men with a more aggressively dominant personality. Thus, it seems, short informal interactions with opposite-sex members are sufficient to reliably elicit a T response, at least in men (see also Roney, Lukaszewski, & Simmons, 2007; van der Meij, Almela, Buunk, Fawcett, & Salvador, 2012). Importantly, these T increases in response to mating opportunities and competitive interactions, as suggested by the challenge hypothesis, are assumed to be mediated or moderated by both individual differences

(personality and cognitive variables such as aggressive dominance, one's involvement and perceived control; Casto & Edwards, 2016a; Salvador, 2005) and situational characteristics (e.g., one's opponent's self-efficacy; van der Meij, Buunk, Almela, & Salvador, 2010). One particularly important situational characteristic, the outcome of a competition, can be more generally framed as status gains or losses, which have been linked to T levels in the biosocial model of status (Mazur 1985, 2015; Mazur, Welker, & Peng, 2015). This model has been derived from research in male rhesus monkeys (e.g., Rose, Berstein, & Gordon, 1975) and predicts T increases after status gains and declines in T after status losses, such as in competitive interactions, as mentioned above. Thus, the evidence reviewed so far shows that not only does T fluctuate in anticipation of, during and after competitions in humans, but also in accordance with variations in contextual factors such as rank, relationship status or parenthood.

1.2. Testosterone and personality

Extant findings have been interpreted in such a way that T regulates cognition, behaviour and related physiological processes along a unidimensional competition versus nurturance dimension (steroid/peptide theory of social bonds; van Anders, Goldey, & Kuo, 2011), a mating versus parenting dimension (Muller, 2017), or, as a theoretical framework embedded into life history theory, a trade-off between mating versus survival effort (Roney, 2016). These theoretical dimensions' endpoints largely overlap with the two main dimensions of another theoretical model of personality and social behaviour, the interpersonal circumplex model (Wiggins, 1982; Leary, 1957; Figure 1), which has been shown to be widely relevant in human social behaviour (e.g., for a review on the role of the two main axes in social perception, see Fiske, Cuddy, & Glick, 2007). The interpersonal circumplex' two main axes, Dominance and Love, correspond to the dimensions of competition/mating and nurturance/parenting/survival, respectively. A dominant behavioural strategy supports especially males in both intrasexual competition and mating (Kordsmeyer, Hunt, Puts, Ostner, & Penke, 2018), whereas the circumplex' main axis love captures nurturance, interpersonal warmth and connecting with others (Turan et al., 2014), which are inherent components of parenting, caring romantic

relationships, and may promote survival (Costa Jr., Terracciano, & McCrae, 2001). However, the circumplex model's main axes are defined as being orthogonal to each other, rather than unidimensional (as the T-mediated trade-offs outlined above). This provokes the question of how T reactivity is related to state changes in circumplex personality facets. On a trait level, Turan, Guo, Boggiano, and Bedgood (2014) investigated the association of baseline T with the interpersonal circumplex personality traits (based on the Interpersonal Adjective Scales-Revised, IAS-R; $N = 85$ men). A positive correlation with agency/disconnectedness (facet BC, a blend of dominance and coldheartedness; Fig. 1) and a negative relationship with submissiveness/communion (facet JK, a blend of nurturance and low dominance; Fig. 1) was found. In addition, Sellers, Mehl, and Josephs (2007) found positive, small-to-medium sized correlations of baseline T with self-rated dominance ($r = .25$; corresponding to the circumplex model's main axis Dominance) in both men and women ($N = 69$). They argue, also based on their additional finding of high temporal stability of T levels (across five days), that T may well serve as a biological marker of inter-individual differences in dominance. However, it needs to be acknowledged that a meta-analysis found a somewhat smaller effect of baseline T on dominance (conceptualized as over-ranking oneself; $k = 13$ samples, overall $N = 2437$; weighted $r = .13$; Archer, 2006), hence the finding of Sellers and colleagues (2007) needs to be replicated. Thus, some first findings, mostly from small samples, indicate baseline T may be related to self-perceptions of dominance, and negatively to submissiveness (both facets forming the circumplex model's vertical main axis; Figure 1) on a trait level, corresponding to the life history trade-off between parenting and mating (Muehlenbein & Bribiescas, 2005; Muller, 2017; Zilioli & Bird, 2017).

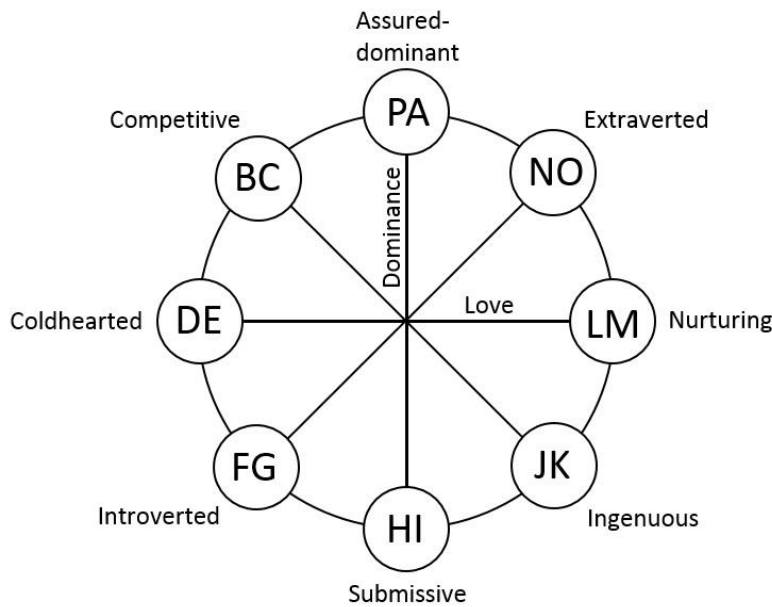


Figure 1. The interpersonal circumplex model with its eight personality facets (adapted from Wiggins, Trapnell, & Phillips, 1988).

1.3. Testosterone and intra-individual differences

Many of the inter-individual associations between T and behaviour in the literature appear to be rather weak. A meta-analysis pinpointed the average correlation between baseline T and aggression in humans to be $r = .08$ (weighted by sample size; $k = 42$ samples; overall $N = 9760$; Archer, Graham-Kevan, & Davis, 2005; update of an earlier meta-analysis by Book, Starzyk, & Quinsey, 2001).

Somewhat larger, but still small, overall associations have been found between baseline T and dominance (see above; Archer, 2006). Regarding potential explanations for these weak associations, firstly, it has been proposed that individual differences in these behaviours and traits may be more reliably related to acute fluctuations in T than to baseline T, hence on an intra- rather than an inter-individual level (Carré & Olmstead, 2015). Across three studies (total $N = 224$ men), antagonistic behaviour during a competitive interaction (Point Subtraction Aggression Paradigm, PSAP) was associated positively with T reactivity, but not baseline T, in men (Carré, Putnam, & McCormick, 2009; Geniole, Carré, & McCormick, 2011; Geniole, Busseri, & McCormick, 2013). In another study (Carré, Baird-Rowe, & Hariri, 2014), men's ($n = 42$, but not women's, $n = 41$) decreased trust ratings

of emotionally neutral faces were predicted by their T increases, but not baseline T, after having engaged in the PSAP. In one of the first studies on the effects of competition-induced T dynamics on behaviour (Mehta & Josephs, 2006), T changes in males ($N = 57$) after having engaged in a rigged one-on-one competition predicted the motivation to compete again, rather than doing a cooperative task after the initial competition, in losers, but not winners. The authors interpreted the findings as losers trying to regain status (after a loss), which is mediated by changes in T. In a similar study (Carré et al., 2009), both male and female participants performed a (same-sex) dyadic competition ($N = 77$). Afterwards, the PSAP was employed to measure reactive aggressive behaviour. While baseline T turned out to be unrelated to aggressiveness, T increases (from baseline T) predicted subsequent aggressive behaviour in male, but not female, losers. Additionally, the interaction of T increases and trait dominance was related to aggressiveness in male winners only (see also Carré & Archer, 2018 for a review). Eisenegger and colleagues (2017) had men ($N = 172$) engage in a mathematical skills-based task, and found positive associations between baseline T and competitiveness (choosing a competitive over a non-competitive payment scheme; but see Torrance, Hahn, Kandrik, DeBruine, & Jones, 2018 for a null-finding on men's baseline T and self-reported intrasexual competitiveness, $N = 59$), as well as between T reactivity during the competition and confidence in one's own performance. Finally, in an extensive review, Carré & Olmstead (2015) proposed competition-induced T fluctuations to be modulating aggressive behaviour, competitive motivation and performance, social cognition (e.g., trust, empathy, or moral decision-making) and mate-seeking behaviour (courtship displays), as well as increasing risk-taking (see Carré, Ruddick, Moreau, & Bird, 2017 for a review; Vermeer, Riečanský, & Eisenegger, 2016). Thus, behaviours and personality traits relevant in an intrasexually competitive context may well show stronger links with acute T fluctuations than with baseline levels, which will be investigated further in this study. However, the question remains how state changes along interpersonal circumplex personality facets are associated with acute T increases in an intrasexually competitive context.

1.4. The dual-hormone hypothesis

As a second explanation for weak links of T with behaviour and personality, an endocrinological interaction has been proposed. Associations between observer-rated trait dominance and baseline T in two studies (study 1: $N = 94$ men and women; study 2: $N = 57$ men; Mehta & Josephs, 2010) depended on the levels of another hormone, the glucocorticoid cortisol (C). A positive association between T and dominance became apparent only if baseline C was low (for men and women together in study 1, non-significant if analysed separately). When baseline C was relatively high, the associations were non-significant (study 1) or even partly reversed (study 2). The authors suggested the interaction of two neuroendocrine axes, the hypothalamus pituitary gonadal (HPG) and hypothalamus pituitary adrenal (HPA) axes, to be at work in regulating dominance, and coined this the dual-hormone hypothesis (Mehta & Josephs, 2010; see also Popma et al., 2007). Since then, a range of studies has examined associations with various kinds of behavioural tendencies and personality traits, which could be subsumed as status-seeking and -maintaining behaviours, and found support for the dual-hormone hypothesis (e.g., on risk-taking: Mehta, Welker, Zilioli, & Carré, 2015; status-attainment: Sherman, Lerner, Josephs, Renshon, & Gross, 2016; reactive aggression: Geniole et al., 2011; see Mehta & Prasad, 2015 for a review). Since multiple studies have already shown effects of an interaction between T and C in a fairly consistent way, it seems warranted to include baseline C as a potential moderator in our study.

1.5. Testosterone and observer perceptions

Beyond associations of T with people's behavioural propensities and self-reported personality traits, T has been suggested to be related to men's secondary sexual traits, and hence to modulate social signalling to both same-sex (e.g., threatening rivals) and opposite-sex (attracting potential mates) members (Puts, 2010). Examples of traits and behaviours which are ontogenetically or proximately linked to T and play a role in social signalling include men's muscularity (Frederick & Haselton, 2007), risk-taking (Mehta et al., 2015) and mating behaviour (van der Meij et al., 2012; see also Andersson, 1994). In intrasexually competitive and mating contexts, it is crucial for both rivals and potential

mates to perceive these signals accurately as cues to good condition (Sell, Cosmides, Tooby, Sznycer, von Rueden, & Gurven, 2009). However, there are only few studies on whether and to what extent such T-mediated behavioural changes (in line with the challenge hypothesis) are perceived by rivals and potential mates. After interacting with a female, but not a male confederate, men with larger T reactivity were perceived by female observers as engaging in stronger self-presentation, and as showing more interest in the confederate and more positive facial cues ($N = 82$; van der Meij et al., 2012). These results are similar to Roney and colleagues' study (2003), in which men's T reactivity after interacting with a female confederate was positively correlated with the confederate's rating of how much the males tried to impress her. Regarding evidence for associations of baseline T and trait-dominance with observer-perceived dominant behaviour, Slatcher, Mehta, and Josephs (2011) had men ($N = 76$) engage in a mate competition for the attention of an attractive female confederate against another male participant. For those high in self-reported trait-dominance, a positive association of baseline T with their observable dominant behaviour during the mate competition, as judged from video recordings, and a negative link with their opponent's observable dominant behaviour were found. While the extant studies have focussed on a specific aspect of men's behaviour in relation to either baseline T levels or T reactivity, there is a gap in the literature regarding associations between acute T fluctuations and a more comprehensive assessment of observer-perceived personality states.

1.6. This study: Aims and hypotheses

The current study aims to replicate and further investigate the reactivity of T in men in response to exposure to a potential mate and an intrasexually competitive situation (challenge hypothesis; Archer, 2006). In particular, pairs of men were asked to engage in a dyadic competition (mixture of four cognitive and physical disciplines), while being supervised by an attractive female confederate. Before (*pre*) and after (*post*) the competition, saliva samples were taken to assess baseline T and T reactivity, as well as baseline C. Men completed a state version of an interpersonal circumplex personality questionnaire. In order to capture behavioural changes and observer-perceptions of

these, our male participants were video-recorded both before (in a calm state, with baseline T levels) and after (in an aroused state, purportedly with elevated T) engaging in a dyadic male competition. In particular, participants were asked to present themselves describing their personal strengths within a short time frame (1 min.), thus engaging a somewhat challenging task (Study 1). Male and female observers subsequently judged these video recordings for personality states (also using the interpersonal circumplex; Study 2) and self-created “social impression” items (Study 3). We created items within three domains, which we believe are especially relevant in the context of both intrasexual competition and female mate choice. The domain “cooperativeness” should correspond to the interpersonal circumplex model’s Love main axis (e.g., Wiggins, 1982; see also the findings of van der Meij et al., 2012 on T reactivity and affiliative behaviour), “self-display” should tap into behaviour signalling to both male rivals and female potential mates (e.g., Roney et al., 2007) and “self-assurance” should capture perceptions of a man’s strength and confidence. These three domains are supposed to directly map onto the T-mediated trade-off between mating/competition (self-display and self-assurance) and parenting/nurturance (cooperativeness; Muehlenbein & Bribiescas, 2005; Muller, 2017; Roney, 2016; van Anders et al., 2011). Previous research has shown that personality traits can be reliably inferred by observers after viewing short recordings of behaviour, which are referred to as “thin slices of behaviour” (Ambady & Rosenthal, 1992; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004). We hence tested influences of T changes on both self-reported and observer-rated personality states, including observer-perceptions of social impressions relevant in men’s signalling behaviour (e.g., Puts, 2010). Furthermore, we examined the interaction of baseline C with T reactivity in association with personality state changes (dual-hormone hypothesis; Mehta & Josephs, 2010). For robustness checks, analyses in Study 1 were additionally performed including the following preregistered control variables, which have been associated with T before: participants’s age, BMI, relationship status, recent sexual experience, self-reported positive/negative affect, stress, and self-esteem (Keevil et al., 2017; Roney et al., 2003; Schultheiss & Stanton, 2009; van Anders & Watson, 2006; van der Meij et al., 2008; Vermeulen,

Goemaere, & Kaufman, 1999). Analyses in Studies 2 and 3 were re-run including the preregistered control variables participant age, relationship status and sexual identity (Keevil et al., 2017; van Anders & Watson, 2006). The methods and hypotheses of all three studies were preregistered on the Open Science Framework (Study 1: osf.io/8n7ev; 2 and 3: osf.io/uhzf3). Specifically, the following hypotheses were tested:

1. Salivary T increases relative to baseline in male participants after a dyadic intrasexual competition under female exposure in the experimental group, but not in a control group in which participants do not compete and are supervised by a male experimenter the whole time.
2. a) We hypothesize larger *pre-post* changes (before to after the competitive interaction) in the experimental than in the control group in the following self-reported personality states along the interpersonal circumplex (which are assumed to be implicated in a competitive mating situation; Roney, 2016; van Anders et al., 2011): increases in dominance (circumplex facet PA), assertiveness (NO), and competitiveness (BC), decreases in nurturance (LM) and introversion (FG).¹

b) We predict *pre-post* changes in the following observer-rated personality states along the interpersonal circumplex: increases in dominance (PA), and competitiveness (BC), decreases in submissiveness (HI), and ingenuousness (JK); and in the following social impression dimensions: increases in self-assurance and self-display, decrease in cooperativeness.
3. a) In the experimental group only, we hypothesize associations of T reactivity with self-reported personality state changes: Positive for dominance (PA), assertiveness (NO), competitiveness (BC), negative for nurturance (LM) and introversion (FG).

b) Further, we hypothesize associations of T reactivity with observer-rated personality states

¹ We did not explicitly predict differences in preregistered hypotheses for specific facets between self-reports and observer-ratings. Instead, differences can be ascribed to the long time frame in which preregistrations for Study 1 versus 2 and 3 were formulated, and resulting rethinking of specific hypotheses (also based on results of preliminary versions of these studies; see Tables S49-S52, S56-S58).

and social impressions changes in the experimental group only: Positive for dominance (PA), cold-heartedness (DE), competitiveness (BC), self-assurance, and self-display, negative for nurturance (LM), submissiveness (HI), and cooperativeness.

4. a) In the experimental group we predict a relationship of T reactivity with self-reported personality state changes to be stronger when baseline C is lower (dual-hormone hypothesis, Mehta & Josephs, 2010). Specifically, we predicted larger increases for dominance (PA), assertiveness (NO), and competitiveness (BC), as well as larger decreases for nurturance (LM) and introversion (FG) with higher T reactivity and low baseline C.
- b) Similarly, we predict baseline C to attenuate the relationship between T reactivity and the following observer-rated personality state and social impression changes: (positive) dominance (PA), competitiveness (BC), self-assurance, and self-display, (negative) submissiveness (HI), ingenuitousness (JK) and cooperativeness.

2. Study 1

We investigated men's T reactivity (Hypothesis 1) and changes in self-reported personality states (Hypothesis 2) after a dyadic intrasexual competitive situation, as well as associations of personality state changes with T reactivity (Hypothesis 3) and the TxC interaction (Hypothesis 4).

2.1. Methods

2.1.1. Participants. We recruited 165 male heterosexual young adults with no hormonal disorders. There were 125 participants in the experimental and 40 in the control group² (age: $M = 24.3$ years, $SD = 3.2$, range 18-34; experimental group: $M = 24.1$ years, $SD = 3.3$, control group: $M = 24.9$ years, $SD = 2.9$). The sample size in the experimental group ($n = 125$) had sufficient power ($> .80$) to detect

² The main purpose of the control group was to provide an opportunity to compare mean level changes (T reactivity, personality state changes) with the experimental group, not to compare correlational results in both conditions. For the former, we consider a sample size of $N = 40$ as appropriate. We had originally preregistered a sample size of $N = 20$ for the control group due to anticipated financial constraints. During data collection, we decided to increase the sample size to $N = 40$ for a more appropriate statistical power. This decision was not influenced by intermediate statistical analyses.

effect sizes of Pearson's $r > .24$ (Cohen, 1992). In the experimental group, 59 indicated to be single, 66 in a relationship (10 open, 50 committed, four engaged, two married, none divorced or widowed; control group: 21 single, one open, 16 committed relationship, two married, none engaged, divorced or widowed). In the experimental and control group 90.4% and 82.5% were students, respectively (of which only two were enrolled as psychology students). On the 7-point Kinsey scale of sexual identity (1 = exclusively heterosexual to 7 = exclusively homosexual; Kinsey, Pomeroy & Martin, 1948), the mean was 1.19 ($SD = 0.46$). One participant indicated a bisexual orientation (Kinsey score = 4) and an unusually high number of sexual partners in the 12 months previous to the study, hence robustness analyses were conducted excluding him and any differences are reported. All procedures received ethics approval from the Georg-Elias-Müller-Institute of Psychology's Ethics Committee (no. 111).

2.1.2. Procedure and Measures. To control for circadian variation in participants' hormonal reactivity, all testing was conducted between 2pm and 6pm (Idris, Wan, Zhang, & Punyadeera, 2017; Schultheiss & Stanton, 2009). The study was divided into two parts, a pre-session and a main session, with the latter happening a few days after the former.

Pre-session. During the pre-session, led by a male experimenter, participants provided informed consent, self-reports on interpersonal circumplex personality traits (using the Interpersonal Adjective List, Jacobs & Scholl, 2005), their sexual history and mating success (such as their relationship satisfaction (Sander & Böcker, 1993) and number of recent sexual partners; Penke & Asendorpf, 2008) were assessed, and body height and weight (to calculate BMI) were measured. A first saliva sample was taken approximately 20 minutes after arriving at the lab (to allow participants to calm down), in order to get a first measure of baseline T levels. Further measures not relevant to this study were also taken (see preregistrations). The pre-session was scheduled on a separate day to familiarize participants with the laboratory setting to avoid artificially increased hormonal levels during the main session (see Fales, Gildersleeve, & Haselton, 2014).

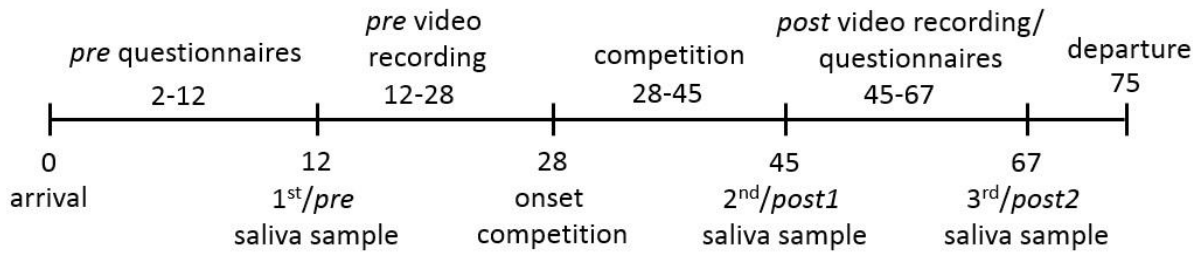


Figure 2. Timeline (in minutes) of the administration of the main session in the experimental group.

Main session. The main session's design included a *pre*- and a *post*-part (Figure 2), identical for the experimental and the control group. In the experimental group, two participants reported to the lab at a time, without meeting each other until the onset of the competition. First, participants filled out questionnaires on a computer. Personality states were assessed with a state version of the Interpersonal Adjective List (IAL; Jacobs & Scholl, 2005). Due to time constraints, the IAL was shortened to five out of eight items per facet (based on factor loadings and fit for the context of this study, 40 items total; see Table S41 for a list of items, and S1 for descriptive statistics). Participants also completed the following state questionnaires, which had been preregistered as control variables, next to age, BMI, recent sexual experience and relationship status: positive and negative affect (German version of the PANAS-X, 10 items each; R  cke & Gr  hn, 2003; Watson & Clark, 1994), stress (STAI, 6 items chosen from the full 20-item version; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and state self-esteem (RSES, 4-item short version; Nezlek & Plesko, 2003).

Approximately 12 to 15 minutes after arriving in the lab they provided a first saliva sample for baseline T and cortisol (C) measures. Afterwards, they were escorted into the video laboratory separately, one after another, to complete the first video recording. Then participants met each other and the female confederate, and engaged in the dyadic competition. Right after the competition, hence approximately 18-20 minutes after onset, a first *post*-saliva sample was taken, after which participants alternately completed further questionnaires including their personality states and the control variables (see above), and the second part of the video recordings. Finally, participants provided a second *post*-saliva sample (Figure 2). Two *post*-saliva samples were taken

since it is not clear when exactly hormonal reactivity is highest and when changes are best detected in saliva. A delay of 15-20 minutes has been suggested for T responses (Casto & Edwards, 2016a; Schultheiss, Schiepe, & Rawolle, 2012). Moreover, Schultheiss and colleagues (2012) recommend to spread out multiple *post*-samples for measuring reactivity by at least 10 minutes, in order to leave time for the hormones passing into saliva after salivary glands have been filled up again; our second *post*-sample was taken 20 minutes (on average, range ca. 18-30mins) after the first *post*-sample. Thus, the two *post*-samples can be interpreted as follows: The first can be seen as a measure of anticipatory reactivity (Marler, Oyegbile, Plavicki, & Trainor, 2005) and a T increase during the competition's first minutes. The second *post*-sample can be interpreted as representing T reactivity during the full competition phase and especially the two later disciplines (arm wrestling and turn-taking verbal fluency game, see below).

For the control group, the procedure was very similar, except that participants completed the main session individually and instead of engaging in a competition watched a documentary video on Canada's sustainable forests (SFM Canada, 2013), which was supposed to be neutral, free of social content, and not challenging, in order not to elicit a T response. The video had a duration of twelve minutes, thus roughly equivalent to the competition. Participants' perception of the video documentary was assessed as a manipulation check, to see if the video was actually seen as neutral and non-challenging by the participants. Items were chosen to tap upon aspects which have been associated with T and C increases in previous studies (Goldey & van Anders, 2016; Hellhammer, Hubert, & Schürmeyer, 1985). The video was rated to be informative and below-average disquieting and stressful, and average in excitement, boredom and challenge (Table S42). Moreover, in the control group there was no female confederate present, the whole procedure was led by a male experimenter instead. After the second *post*-saliva sample, participants were debriefed about the study's objective.

2.1.3. Hormonal assessments. Participants were asked to refrain from drinking alcohol, exercising, taking recreational or non-prescribed clinical drugs on the day of the study, from

ingesting caffeine (coffee, tea, coke) or sleeping three hours before the study, and from eating, drinking (except for water), smoking or brushing teeth one hour before their scheduled appointment (Geniole et al., 2013; Lopez, Hay, Conklin, 2009). To check participants' adherence to these instructions and to assess further potential influences on the saliva samples and hormonal levels, a screening questionnaire was administered at the beginning of the session (Schultheiss & Stanton, 2009). None of the participants indicated to be taking hormonal medication or supplements.

For all saliva samples, participants provided at least 2ml of saliva via unstimulated passive drool through a straw (following the procedural guidelines provided in Schultheiss et al., 2012; Fiers et al., 2014). The samples were immediately transported to an ultra-low temperature freezer (-80°C), where salivary T is stable for at least 36 months (Granger, Shirtcliff, Booth, Kivlighan, & Schwartz, 2004). At the end of data collection, saliva samples were shipped on dry ice to the Technical University of Dresden, where they were analysed using chemiluminescence-immunoassays with high sensitivity (IBL International, Hamburg, Germany). The intra- and inter-assay coefficients (CVs) for C are below 8% and for T below 11%. Outliers were winsorized to 3 SDs ($n = 8$ in the experimental, $n = 1$ in the control group, in accordance with Mehta et al., 2015; see also Pollet & van der Meij, 2017, for an extensive discussion of the influence of hormone outlier handling on significance testing). All T and C measures appeared to be positively skewed and to violate the assumption of normality (Shapiro-Wilk test statistics $< .94$, $ps < .001$). Consequently, all four variables were log₁₀-transformed (see e.g., Mehta et al., 2015). One participant in the experimental condition had missing data for baseline T and C (from the main session), hence we could not calculate T reactivity and the TxC interaction (decreasing the sample size for these measures to $n = 124$). Thirty-one participants reported either recent gum bleedings or oral infections, which can lead to elevated steroid hormone concentrations (Schultheiss & Stanton, 2009). Testosterone and Cortisol levels were compared for these as a group with the remaining participants and no differences were detected (all unsigned $ts < 1.58$, $ps > .11$).

To ease interpretation and comparison with other studies, hormonal values in these tables are reported in untransformed values (for T in pg/ml, C in nmol/l). The correlations amongst the two T *post*-measures were high (experimental/control group: $r = .75/.89$, $ps < .001$), suggesting moderate-to-high stability, comparable to previous results (Sellers et al., 2007; Turan et al., 2014). As expected, baseline T (from the main session) inversely predicted T reactivity for both *post*-saliva samples in the experimental group ($r = -.44$ and $r = -.40$, respectively, $ps < .001$; Roney, Simmons, & Lukaszewski, 2010).

2.1.4. Competition. Immediately before competition onset, participants filled out a short questionnaire assessing their motivation and expectation to win (Costa, Serrano, & Salvador, 2016). Participants competed in dyads in four disciplines, under supervision of the attractive female confederate, with the aim of eliciting a T response (e.g., Roney et al., 2007; Salvador & Costa, 2009). To increase participants' engagement in the competition, the winner of each discipline received an additional immediate monetary compensation of 2€, presented as a coin on the table at each discipline's conclusion. For the four disciplines, a mixture of physical and cognitive tasks was chosen in order to increase the chances that the outcome of the competition remained undecided for longer (assuming a heterogeneity in talents): (1) a table pinball soccer game (played until one contestant had scored five goals), (2) a snatching game (where participants had to solve figural reasoning tasks, inferring which one out of five objects did not match two objects on cards in shape or color, and grasp the right object from the table quicker than the opponent, until one participant had won five rounds), (3) arm wrestling (best of three, alternating arms) and (4) a turn-taking verbal fluency task (where participants took turns naming words belonging to a certain category and starting with a specific letter, e.g., "occupations starting with M"; best of three). For all disciplines, see illustrations in the online supplementary material (Figure S1).

During all four disciplines, the female experimenter was told to interact naturally with the participants, while providing some verbal encouragement. The confederate had been carefully chosen for above-average physical attractiveness and communicative skills, heterosexual

orientation, and age comparable to participants'. Her above-average physical attractiveness was confirmed in a pilot rating study (13 independent male raters unacquainted with the female confederate judged a face and a full-body photo on four 7-point Likert scales from 1 = not at all attractive to 7 = extremely attractive; facial attractiveness: $M = 5.15$, $SE = 0.27$; bodily attractiveness: $M = 5.46$, $SE = 0.27$; overall short-term attractiveness: $M = 5.69$, $SE = 0.31$; overall long-term attractiveness: $M = 4.77$, $SE = 0.47$).

2.1.5. Video recordings. During the *pre* and *post* parts of both the experimental and the control group, self-presentation video recordings of participants were taken. *Pre* recordings were supervised by the male experimenter, and *post* recordings by the female confederate, both from a separate video control room. Each participant was first told that the question he should answer within a one-minute time limit was, "What do you think, right now, is great about yourself?". Then he was presented with one of two sets of eight terms about "life domains" (Table S43) and instructed to choose three, which he would subsequently talk about. The life domains of the two sets were matched for equivalent meaning and presented in counterbalanced order, one in the *pre* and one in the *post* part (e.g., "humour" and "creativity"). The participants were given these terms as hints what to talk about and in order to ensure that they talked about of a variety of different, but roughly comparable things when presenting themselves. The three chosen domains were placed next to the camera, with the participant standing roughly four meters from the camera (to have a full-body view). Participants could start to speak whenever they felt like and gently reminded when they passed the time limit, but not stopped abruptly.

2.1.6. Statistical analyses. For personality state changes and T reactivity, residual scores were calculated by regressing *post*- on *pre*-personality states for all IAL octants and *post*- on *pre*-T levels (van der Meij et al., 2012; using the baseline saliva sample obtained on the day of the main session; for both *post*-competition samples separately), respectively (see Burt & Obradovic, 2013 for a detailed discussion of difference scores versus residuals). For all analyses described below, separate tests and models were employed for the two *post*-competition T measures. Since our

studies were preregistered, we decided to use one-sided tests for directional Hypotheses 1 to 4, marked with “one-tailed” below (Cho & Abe, 2013; Lakens, 2016). For robustness checks, all models (as described below in the results section) were again run including the preregistered control variables: age, BMI, relationship status coded as a binary variable (single versus partnered, the latter included those in an open or committed relationship, as well as engaged and married participants), recent sexual experience (binary, within previous 1 month) and *pre-to-post* changes in state positive and negative affect, stress, and self-esteem. Analyses were performed using R (R Core Team, 2015), computerized versions of questionnaires were administered using formr.org (Arslan & Tata, 2017).

2.1.7. Data availability. The data and analysis scripts associated with this research are available at osf.io/8n7ev.

2.2. Results

Descriptive statistics for all main variables, and bivariate correlations between personality state changes and T reactivity can be found in Tables S1 and S2. Internal consistencies (Cronbach’s α) for the eight IAL facets (*pre* and *post* separately) ranged between .60-.85/.45-.89 for the experimental/control group (Table S1).

2.2.1. Hypothesis 1: To assess T reactivity in the experimental compared to the control group, univariate linear regression models were employed, predicting T reactivity from the dummy-coded variable condition (0 = control, 1 = experimental condition). For the first *post*-saliva sample (taken directly after the competition), but not the second (taken on average 20 mins after the first sample), T reactivity was higher in the experimental than control group (1st: $\beta = 0.45$, $p < .01$ (one-tailed), *partial* $\eta^2 = 0.04$; 2nd: $\beta = 0.22$, $p = .11$ (one-tailed), *partial* $\eta^2 = 0.01$). When including the preregistered control variables results remained virtually unchanged (Table S2a).

2.2.2. Hypothesis 2: For personality state changes from before to after the competition (or watching the video in the control condition), comparing the two conditions, univariate linear regression models were run, with personality state changes (for IAL octants separately) as the dependent variable, predicted by condition (0 = control, 1 = experimental condition). Larger IAL

personality state increases in the experimental relative to the control group were found for competitiveness (BC; $\beta = 0.46$, $p < .01$ (one-tailed), $partial \eta^2 = 0.04$; Table S3) and coldheartedness (DE; $\beta = 0.40$, $p = .03$, $partial \eta^2 = 0.03$), while ingenuousness (JK; $\beta = -0.44$, $p = .02$, $partial \eta^2 = 0.04$) decreased more in the experimental group (for the remaining facets, $ps > .08$). When including the eight control variables, the changes in competitiveness and coldheartedness became non-significant ($ps > .05$; Table S4).

2.2.3. Hypothesis 3: To test the association between T reactivity and personality state changes, the latter were predicted by the former, in the experimental group only. A positive association between T reactivity and personality state changes in competitiveness (BC) was detected for the first, but not the second *post-saliva* sample (1st: $\beta = 0.15$, $p = .0496$ (one-tailed), $partial \eta^2 = 0.02$; 2nd: $\beta = -0.02$, $p = .80$, $partial \eta^2 = 0.00$; see Table 1; Figure 3). When including the control variables results remained unchanged (Table S9).

Table 1

Results from linear models predicting personality state changes from T reactivity (Study 1, Hypothesis 3)

IAL state changes	1 st T reactivity (β)	SE	P	Partial η^2	2 nd T reactivity (β)	SE	p	Partial η^2
Δ assured-dominant (PA)	.09	.09	.15 ^p	0.01	.11	.09	.12	0.01
Δ competitive (BC)	.15	.09	.049 ^p	0.02	-.02	.09	.80	0.00
Δ coldhearted (DE)	.12	.09	.20	0.01	-.01	.09	-.15	0.00
Δ introverted (FG)	.05	.09	.29 ^p	0.00	.01	.09	.47 ^p	0.00
Δ submissive (HI)	.01	.09	.85	0.00	.15	.09	.11	0.02
Δ ingenuous (JK)	-.12	.09	.18	0.01	-.11	.09	.22	0.01

Δ nurturing (LM)	-.07	.09	.22 ^p	0.00	-.04	.09	.32 ^p	0.00
Δ extraverted (NO)	.01	.09	.46 ^p	0.00	.13	.09	.08 ^p	0.02

Note. IAL = interpersonal adjective list; *SE* = standard error; partial η^2 = partial eta-squared effect size;
^p = one-tailed *p*-value due to preregistered hypothesis.

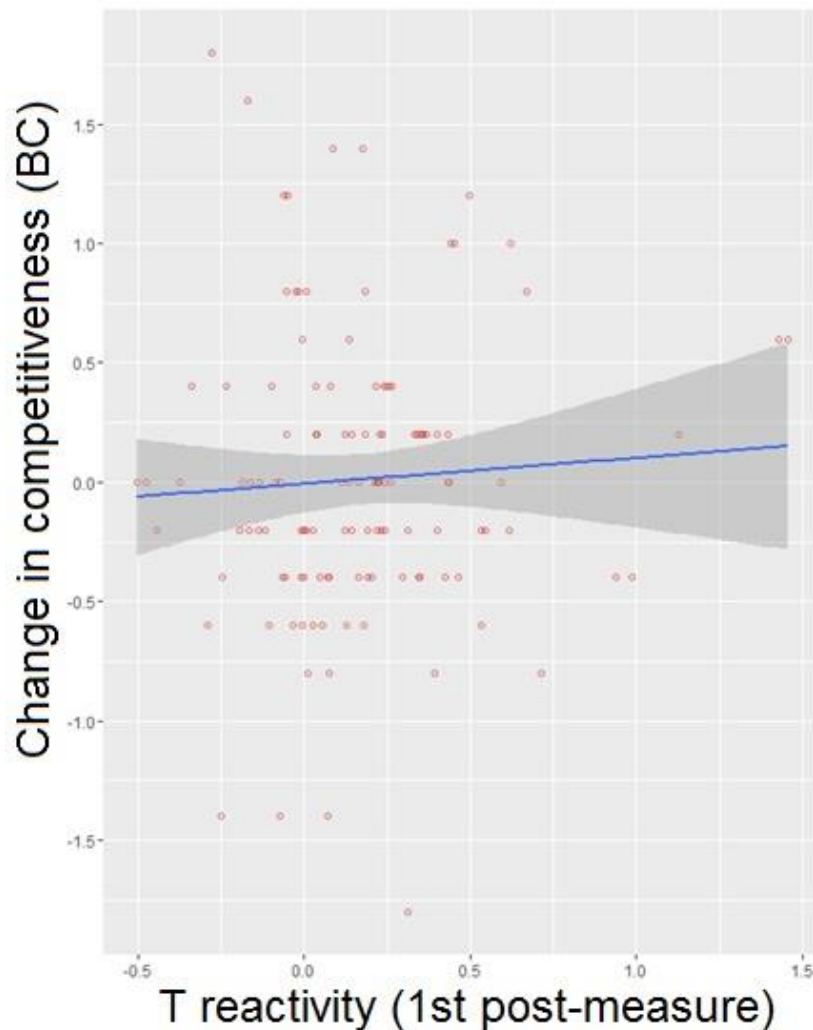


Figure 3. Pre-post changes in competitiveness (facet BC) predicted by the first T reactivity measure in the experimental group.

2.2.4. Hypothesis 4: The interaction between T reactivity and baseline C (TxC; using the baseline C measure obtained on the day of the main session) was added, besides T reactivity and baseline C, to the regression model predicting personality state changes to test for moderating effects of baseline C in the experimental group. A significant moderating effect of baseline C on the association between T reactivity and personality state changes in competitiveness (BC; 1st sample: 6

= -0.43, $p < .01$ (one-tailed), $partial \eta^2 = 0.06$; 2nd: $\beta = -0.36$, $p < .01$ (one-tailed), $partial \eta^2 = 0.05$) and in dominance (PA; 1st sample: $\beta = -0.35$, $p = .01$ (one-tailed), $partial \eta^2 = 0.04$; 2nd: $\beta = -0.26$, $p = .04$ (one-tailed), $partial \eta^2 = 0.03$; for the other facets, $ps > .06$; Table S10) was found. In both cases, there was a positive relationship between T reactivity and change in competitiveness when baseline C was low, but a negative link when baseline C was high (Figure 4). When adding the control variables, results were unchanged for competitiveness, but the significant TxC interactions for dominance faded ($ps > .27$ (one-tailed); Table S11).

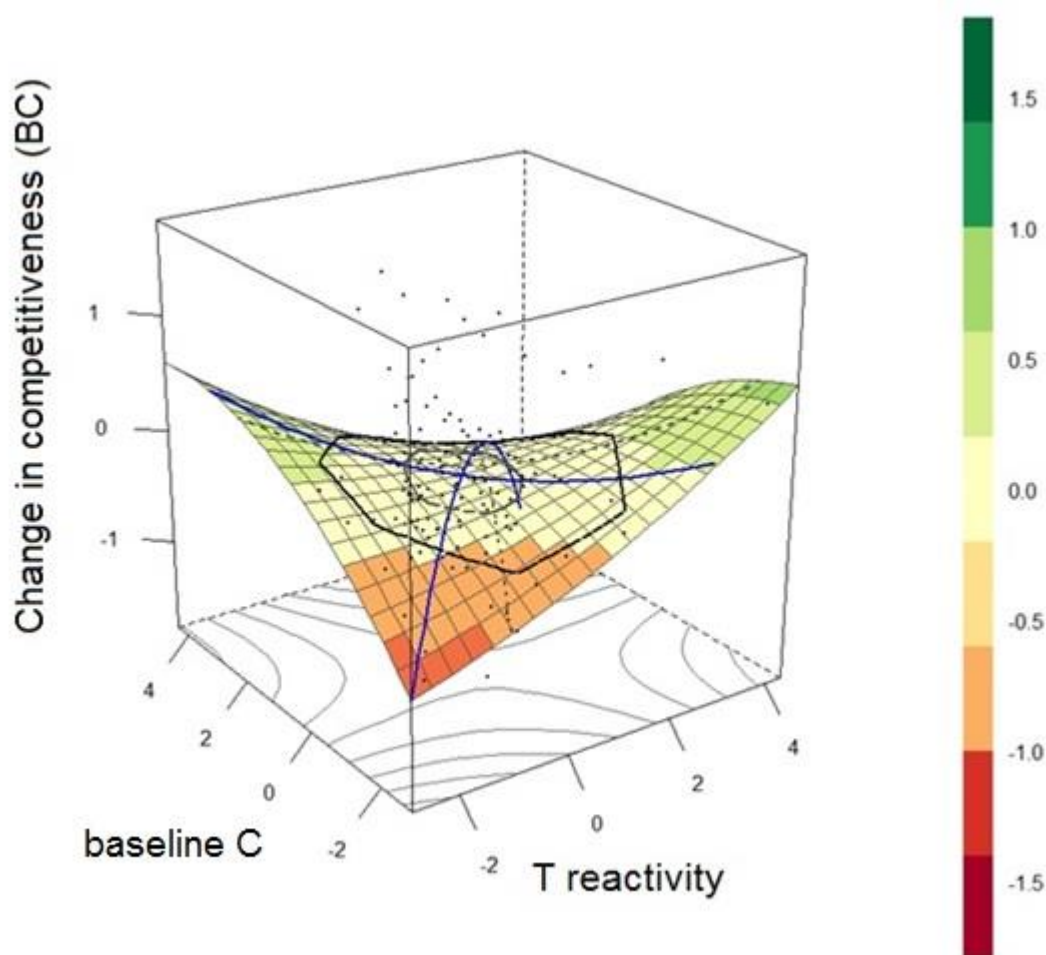


Figure 4. Interaction between T reactivity (first *post*-sample) and baseline C predicting the self-reported *pre-post* change in competitiveness (BC).

To conclude, we found a T reactivity which was larger in the experimental group, relative to the control group, for the first, but not the second sample. Participants in the experimental group rated themselves to be more competitive (BC) and coldhearted (DE) and less ingenuous (JK) *post*

compared to *pre*, relative to changes in the control group. The former change in the experimental group was positively predicted by participants' T reactivity, but only for the first *post*-saliva sample, not the second. For both *post*-saliva samples, the relationship between T reactivity and change in competitiveness was attenuated by high baseline C. We additionally found a TxC interaction on changes in dominance (PA; for the first *post*-saliva sample only). T reactivity was positively associated with increases in competitiveness and dominance, only if baseline C was low, in agreement with the dual-hormone hypothesis. Thus, all hypotheses received at least partial support.

2.2.5. Actor-partner effects. Due to the dyadic nature of the participants' interaction, we ran actor-partner interdependence models (APIM; Kenny, Kashy, & Cook, 2006) using the AMOS 23 statistics programme. Here, the data of both competitors are analysed simultaneously to control for a potential dependence amongst them, grouped into winners and losers of the competition. Effects on a focal participant's personality state changes by the respective opponent's T reactivity were investigated (Hypothesis 3), motivated by earlier findings on an association between baseline T levels and opponent's dominant behaviour during a male dyadic mate competition (Slatcher, Mehta, & Josephs, 2011). Since in previous analyses we detected an effect of the first *post*-measure T reactivity on changes in self-reported competitiveness (BC), we will focus on these. The APIM models included both competitors' T reactivities and changes in competitiveness. No partner effects were detected (winners' T reactivity on losers' BC change: $\beta = -.03$, $SE = 0.23$, $p = .85$; losers' T reactivity on winners' BC change: $\beta = -.07$, $SE = 0.24$, $p = .60$; see Figure S2). Hence, it can be concluded that for the association between personality state changes in competitiveness and T reactivity, no dependence amongst the two competitors was detected that could bias the results on an individual level, as reported above.

2.2.6. Further replication analyses and preregistered hypotheses. We attempted to replicate previous findings of a T increase in winners and a T decrease in losers ("winner effect", for a recent meta-analysis showing an average effect size of $d = 0.20$, $k = 60$ samples, $N > 2500$, see Geniole et al., 2017), a range of moderating and mediating effects on T reactivity, which have been

reported earlier (e.g., female confederate's ratings of men's behaviour during a mating competition; Roney et al., 2003; Slatcher et al., 2011; effects of men's recent sexual activity; Roney et al., 2003; van der Meij et al., 2008; effects of trait aggressive dominance; van der Meij et al., 2008; associations with relationship status; van der Meij et al., 2008), and reported associations between circumplex personality traits and baseline T (Turan et al., 2014). Results can be found in the supplementary (Tables S21-S40). Finally, the results of further preregistered analyses (mainly on *pre-post* changes in additional personality states and their associations with baseline T, baseline C, and T reactivity) can be found in the online supplementary (Tables S39-S57).

2.2.7. Discussion. In Study 1 we demonstrate significant T increases in men after engaging in a dyadic intrasexual, female-led competition (replicating earlier findings of an acute T rise in the face of challenges such as mating opportunities and intrasexual contest; Archer, 2006; van der Meij et al., 2008). In addition, we show that these hormonal fluctuations (T reactivity, relationships partly attenuated by high baseline C) are associated with personality state changes in dimensions relevant in such a competitive interaction (i.e., competitiveness and dominance), underlining previous suggestions of T modulating men's competitive behaviour (Carré & Olmstead, 2015). However, these associations were significant mostly only for the first, but not the second T reactivity measure (except for the TxC interaction on competitiveness; Table 2), and only for one or two (Hypotheses 3 and 4, respectively) of the five preregistered circumplex facets (Table 2). Accordingly, they should be treated with care until further replication strengthens their robustness. In the following studies, we examine if these hormone-mediated personality state changes are detectable by naive observers, in order to test if hormonal responses trigger perceptible behavioural changes that can potentially function as social signals.

3. Studies 2 and 3³

³ Additional preliminary versions of these two studies were conducted earlier. Details and results can be found in the supplementary (Tables S48-S58).

In preregistered Studies 2 and 3, we tested if men's personality state changes from before to after an intrasexual competitive situation can be detected by observers based on thin slices of behaviour (Hypothesis 2), and if these state changes are associated with target men's hormonal changes (T reactivity, Hypothesis 3, and TxC interaction, Hypothesis 4). The thin slices are based on the video recordings from Study 1. In Study 2, target men's personality states were rated by females using the Interpersonal Adjective List (IAL; Jacobs & Scholl, 2005). Study 3 employed self-created social impression items (3 facets: self-assurance, cooperativeness, self-display) and both male and female raters. Both studies involved target men from Study 1's experimental and control group, to be able to test differential personality state changes between these two conditions, since simple *pre-post* changes could at least partly be attributed to practice effects (target men speaking more fluently and feeling more confident in the *post* than in the *pre* video recordings and hence being judged differently). The comparison of *pre-post* changes with those in the control group theoretically enables us to partial out practice effects, as these should be present equally in both conditions.

3.1. Study 2

3.1.1. Methods

3.1.1.1. Participants. Participants were 400 females (age: $M = 23.7$, $SD = 4.8$, range 16-56 years), recruited via a local participant database. Raters' mean age was comparable to target men's mean age ($M = 24.3$ years).

3.1.1.2. Video-stimuli and procedure. Video-recorded self-presentations from Study 1 were used in this rating study (see above for details). All videos were cut to a maximum length of one minute. The videos of five participants in the experimental and two in the control group were removed from the stimuli sample due to audio problems, leaving a final stimulus set of *pre-* and *post-*videos each from $N = 158$ target men ($n = 120$ each for the experimental, $n = 38$ for the control group; length $M = 53$, range 10-62 sec.). Videos were distributed in a way so that the two videos of each target man never appeared together in the same set, to avoid direct contrast effects. Ratings

were conducted in a computer laboratory on 24" screens using the software MediaLab v2014 (Empirisoft Corporation). Videos were presented in a randomized order.

3.1.1.3. Stimuli and items. The video stimuli were divided into six sets of 40 videos and two sets of 38 videos each, of which half were *pre*- and the other half *post*-videos. Each video was rated by ten independent female raters. Three-hundred and twenty of the raters viewed 40 videos and the remaining 80 rated 38 videos. Participants were randomly assigned to one of the eight video groups and to one of five item groups. The German version of the IAL was employed (five items per facet as in Study 1; Table S1); raters judged the target men on one of the five items per facet (a total of eight items per rater; each rater used the same eight items for the 38/40 videos). Overall target men were rated on 40 IAL items on a 5-point Likert scale (1 = "disagree completely" to 5 = "agree completely").

3.1.1.4. Statistical analyses. Analyses equalled those of Study 1, only substituting observer-ratings for self-reports (see the results section for a description of the models).

3.1.2. Results

Bivariate Pearson correlations between observer-rated personality state changes and T reactivity, as well as descriptive statistics for all variables can be found in the online supplementary (Tables S8 & S47). Internal consistencies (Cronbach's α) for the eight IAL facets (*pre* and *post* separately) ranged between .81-.95/.73-.95 for the experimental/control group (Table S8), and interrater agreements (Cronbach's α) for the eight facets were satisfactory to good (*pre*: α = .85-.95, *post*: α = .89-.96, changes: α = .83-.91; Table S48).

3.1.2.1. Hypothesis 2: Univariate linear regression models were run, with observer-rated personality state changes as the dependent variable, predicted by condition (0 = control, 1 = experimental condition). We detected larger *pre-post* increases in the experimental compared to the control group for dominance (PA; β = 0.49, p < .01 (one-tailed), η^2_p = 0.04; Table S9) and competitiveness (BC; β = 0.37, p = .02 (one-tailed), η^2_p = 0.02), and larger decreases for submissiveness (HI; β = -0.55, p = .01 (one-tailed), η^2_p = 0.06) and ingenuousness (JK; β = -0.42, p =

.01 (one-tailed), $\eta^2_p = 0.03$). When adding the preregistered control variables age, relationship status and sexual orientation to the latter linear regression models results were unchanged (Table S10).

3.1.2.1. Hypothesis 3: Observer-rated personality state changes were predicted by T reactivity in the experimental group only. No significant associations were found for any of the two hormonal *post*-samples ($ps > .05$; Table 2 & S11).

Table 2

Results from linear models predicting observer-rated personality state changes from T reactivity (Study 2, Hypothesis 3)

IAL state changes	1 st T reactivity (β)	SE	p	Partial η^2	2 nd T reactivity (β)	SE	p	Partial η^2
Δ assured-dominant (PA)	.14	.09	.07 ^p	0.02	.15	.09	.06 ^p	0.02
Δ competitive (BC)	.06	.09	.25 ^p	0.00	.10	.09	.15 ^p	0.01
Δ coldhearted (DE)	.07	.09	.48	0.00	.06	.09	.54	0.00
Δ introverted (FG)	-.04	.09	.70	0.00	.00	.09	.97	0.00
Δ submissive (HI)	-.04	.09	.34 ^p	0.00	.00	.09	.50 ^p	0.00
Δ ingenuous (JK)	-.04	.09	.32 ^p	0.00	-.04	.09	.32 ^p	0.00
Δ nurturing (LM)	-.05	.09	.62	0.00	.03	.09	.75	0.00
Δ extraverted (NO)	.08	.09	.39	0.01	.08	.09	.36	0.01

Note. IAL = interpersonal adjective list; SE = standard error; partial η^2 = partial eta-squared effect size;

^p = one-tailed p -value due to preregistered hypothesis.

3.1.2.2. Hypothesis 4: The interaction between T reactivity and baseline C (TxC) was added, besides T reactivity and baseline C, to the regression model predicting observer-rated personality state changes in the experimental group. No significant TxC interaction effects were detected (for the first/second T *post*-sample: unsigned β s < .19/.19, $ps > .23$; Tables S12-S13).

3.1.2.3. Discussion. We show that naïve female observers attribute personality state changes to men from before to after engaging in an intrasexual competition. Men were perceived to

increase in dominance and competitiveness, and decrease in submissiveness and ingenuousness. These were not linked to target men's T reactivity (nor to the TxC interaction), however. Still, we provide evidence that post-competition/female exposure personality state changes appear to be detectable by naive observers based on video-recorded thin slices of behaviour, suggesting that engaging in intrasexual competition under female exposure triggers behavioural changes, which may be functional in social signalling. We employed only female observers, since we initially planned to focus on intersexual signalling effects (dominance- and competition-related behaviour) of T reactivity and associated personality state changes. In Study 3, we envisaged to examine changes in observer-perceptions in terms of more behaviourally phrased social impression items (self-display, self-assurance, cooperativeness). Since these social impression dimensions directly tap into facets implicated in men's intrasexual competition, mate attraction and affiliative behaviour (both intra- and intersexual signalling), we recruited male and female observers to investigate *pre-post* changes in social impressions and associations with T reactivity (and a TxC interaction).

3.2. Study 3

3.2.1. Methods

3.2.1.1. Participants. One-hundred and sixty raters in (80 females; age: $M = 24.5$, $SD = 4.9$, range 16-53 years) were recruited via a local participant database. Again, raters' mean age was comparable to target men's average age ($M = 24.3$ years).

3.2.1.2. Stimuli and procedure. Stimuli and procedure were the same as in Study 2. One-hundred twenty-eight raters watched 40 videos and the remaining 32 saw 38 videos.

3.2.1.3. Items. Ten male and female participants rated target men on three dimensions (self-display, cooperativeness, self-assurance) with three items each. Two positive items and one inversed item were employed on a 5-point Likert scale (1 = "disagree completely" to 5 = "agree completely"; see Table S38 for a full list of items), plus a question if the target was recognized (same as in Study 2).

3.2.1.4. Statistical analyses. Analyses equalled those of Studies 1 and 2 (see below).

3.2.2. Results

Descriptive statistics for all variables and bivariate Pearson correlations between observer-rated social impression changes and T reactivity can be found in the online supplementary (Tables S53-S54). Internal consistencies (Cronbach's α) for the three facets (*pre* and *post* separately) ranged between .85-.97/.77-.98 for the experimental/control group (Table S53). Interrater agreements (Cronbach's α) for the three facets were satisfactory to good (*pre*: α = .69-.89, *post*: α = .72-.88, changes: α = .69-.82; Table S55). Since we employed male and female raters (as explained above), at first we assessed whether rater sex had a significant effect on the observer-ratings (Hypotheses 2-4). Rater sex was added as a covariate, and its interactions with condition, T reactivity and with the TxC interaction were investigated. We found no significant main effects of or interactions with rater sex (all unsigned t s < 1.39). Since we detected no effect of the sex of the raters, we will present results with observer ratings aggregated across male and female raters.

3.2.2.1. Hypothesis 2: Univariate linear regression models were run, with observer-rated social impression changes as the dependent variable, predicted by condition (0 = control, 1 = experimental condition). We found a higher increase in both self-display (β = 0.31, p = .046 (one-tailed), η^2_p = 0.02; Table S15) and self-assurance (β = 0.37, p = .02 (one-tailed), η^2_p = 0.02), but not cooperativeness (p = .45 (one-tailed)), in the experimental relative to the control group. When adding the preregistered control variables age, relationship status and sexual orientation, the significant increase in self-display faded (p = .06 (one-tailed); Table S16).

3.2.2.2. Hypothesis 3: Observer-rated social impression changes were predicted by T reactivity in the experimental group only. We detected a positive association for the first hormonal *post*-sample (β = 0.20, p = .01 (one-tailed), η^2_p = 0.04), but not the second (β = 0.14, p = .07 (one-tailed)); for self-display and cooperativeness, p s > .06 (one-tailed); Table S17). This effect was robust when adding the preregistered control variables (Table S18).

3.2.2.3. Hypothesis 4: The interaction between T reactivity and baseline C (TxC) was added, besides T reactivity and baseline C, to the regression model predicting observer-rated social

impression changes in the experimental group. No significant interaction effect, hence no moderation of the association between T reactivity and changes in observer-ratings by baseline C, was detected (for the first/second T *post*-sample, $ps > .05$; Tables S19 and S20).

3.2.2.4. Discussion. Target men were perceived to increase more in self-display and self-assurance by naïve observers after, relative to before, engaging in an intrasexual competition (experimental vs. control group). In the experimental group, the observer-rated increase in self-assurance was higher for target men showing a larger T reactivity (no association with TxC interaction). Hence, we demonstrate T-modulated changes in social signalling in terms of more concretely phrased (compared to interpersonal circumplex items) social impression items, in domains relevant in men's inter- and intrasexual competition.

Table 3

Overview of results for preregistered hypotheses for Studies 1, 2 and 3

Hypotheses	Self-reports	Observer-ratings	
	IAL	IAL	Social impressions
1) T reactivity*	1st, not 2nd T <i>post</i> -sample		
2) personality state changes*	competitiveness (BC), coldheartedness (DE)***	dominance (PA), competitiveness (BC), submissiveness (HI), ingenuousness (JK)	self-display, self-assurance
Not supported for:	dominance (PA), extraversion (NO), nurturance (LM), introversion (FG)	-	cooperativeness
3) personality state changes & T reactivity**	competitiveness (BC; 1st, not 2nd T <i>post</i> -sample)	-	self-assurance (1st, not 2nd T <i>post</i> -sample)
Not supported for:	dominance (PA), extraversion (NO), nurturance (LM), introversion (FG)	dominance (PA), competitiveness (BC), submissiveness (HI), ingenuousness (JK)	cooperativeness, self-display
4) personality state changes & TxC**	competitiveness (BC; 1st & 2nd T <i>post</i> -sample), dominance (PA; 1st T <i>post</i> -sample only)	-	-
Not supported for:	dominance (PA), extraversion (NO), nurturance (LM), introversion (FG)	dominance (PA), competitiveness (BC), submissiveness (HI), ingenuousness (JK)	cooperativeness, self-display, self-assurance

Note. T = testosterone, TxC = T reactivity x baseline C interaction, IAL = interpersonal adjective list,

*relative changes, experimental versus control group, **experimental group only, ***hypothesis not preregistered.

4. General Discussion

Across three preregistered study parts, we investigated the association between self-reported and observer-rated personality state changes and hormonal reactivity in men in an intrasexually competitive context. Several interesting findings regarding the interplay of personality and hormones emerged. Firstly, an increase in testosterone (T) was detected from before to after competing against another male participant while being supervised by an attractive female confederate. The increase was partly (for the 1st, but not the 2nd *post* T sample) significantly higher than in a control group, in which men only watched a neutral documentary and were supervised by a male experimenter. Secondly, in Study 1, *pre-post* increases in self-reported personality state facets competitiveness (BC)⁴ and coldheartedness (DE) of the Interpersonal Circumplex (Wiggins, 1982; Figure 1), and decreases in ingenuousness (FG) were found (experimental relative to control group). Concerning observer-ratings on the interpersonal circumplex (Study 2), larger increases in the experimental compared to the control group for dominance (PA) and competitiveness (BC), and larger decreases for submissiveness (HI) ingenuousness (JK) were found. Regarding the three social impressions (Study 3), *pre-post* increases in observer-perceived self-display and self-assurance, but not cooperativeness, were larger in the experimental than in the control group. Note that concerning the association between T reactivity and self-reported personality state changes, a positive link emerged of T reactivity with changes in competitiveness for the first, but not the second *post* T sample. For observer-rated social impressions, increases in self-assurance were linked to a higher T reactivity for the first, but not second *post* T sample. We will discuss this pattern below. An interaction between T reactivity and baseline C on changes in self-reported competitiveness (for both *post* T samples) and dominance (for the first sample only) was found in the experimental group. That is, associations between T reactivity and changes in competitiveness and dominance were

⁴ The facet BC is originally called “arrogant-calculating” (Horowitz et al., 2006). Since we selected five out of the overall eight items of this facet, with the final set of items including “competitive”, “provocative” and “belligerent”, and given our study’s intrasexually competitive context, we decided to re-label the facet to “competitiveness”. This is, of course, only descriptive, and our interpretation concerning this facet would equally apply when using the label “arrogant-calculating”.

attenuated by high baseline C. No TxC interaction emerged for the observer-perceptions, for neither circumplex personality states nor social impressions. Finally, dyadic effects between both participants' T reactivity and self-reported personality state changes were investigated employing actor-partner interdependence models (APIM; Kenny, Kashy, & Cook, 2006). No effects on a focal participant's personality state changes by the respective opponent's T reactivity were revealed, thus questioning potential partner effects (as reported for dominance behaviours by Slatcher, Mehta & Josephs, 2011).

Thus, employing a relatively large sample of men ($N = 165$), we show preregistered associations of post-competition T reactivity with self-reported personality state changes, and social impression changes perceived by naive observers. The T increase, which was partly larger in the experimental group following a competitive interaction than in the control group, is in line with predictions derived from the challenge hypothesis (Archer, 2006; Wingfield et al., 1990), replicating previous studies in the realms of intrasexual competition and female exposure (e.g., Roney et al., 2007; van der Meij et al., 2010).

Moreover, personality state changes were detected in our intrasexually competitive context, which mostly support our preregistered hypotheses (as outlined above; see Table 3). Regarding the interpersonal circumplex, changes in self-reports and observer-ratings overlap for competitiveness and ingenuousness. Interestingly, observer-perceptions also changed on both endpoints of the Dominance main axis (dominance and submissiveness), whereas self-reports increased *pre-post* on the negative endpoint of the Love axis (coldheartedness). Thus, we see somewhat diverging changes for participants' own reports and observer-ratings. These might be meaningful, in that while (changes in) dominant personality states are more perceptible externally (related to boldness and self-assurance - for the latter social impression facet we also saw changes in observer-ratings), changes in coldheartedness may be more an internal process, with no clear associated differences in mimics or gestures, which would be observable by raters. There is some evidence that certain personality facets, such as extraversion, are being judged with higher accuracy than more internal

facets, such as openness to experience (Ambady, Bernieri, & Richeson, 2000). However, importantly, in our case we are not primarily interested in accuracy (in terms of overlap between target and informant), only in changes from baseline to a hormonally aroused state. Thus, so far we can only say that changes in coldheartedness seem to be more salient to oneself, and changes in dominance and submissiveness more to observers. Changes in these personality facets may well be adaptive in competitive situations (Dall, Houston, & McNamara, 2004), and map onto a behavioural spectrum of competition versus nurturance, which has been suggested for effects of T (van Anders et al., 2011).

4.1. Implications for the role of T in mate acquisition and intrasexual competition

Hence, in our study we showed that T indeed rose in the face of an intrasexual competition combined with female exposure. T reactivity also appeared to be linked to personality state changes in domains relevant to this intrasexual competitive context. Since these personality dimensions relate to aspects of social status (Eisenegger, Haushofer, & Fehr, 2011; Mattan, Kubota, & Cloutier, 2017), our findings converge with predictions from the biosocial model of status (Mazur 1985, 2015; Mazur, Welker, & Peng, 2015), according to which fluctuations in status should be linked to T levels. Moreover, these associations fit well into previous claims of acute T fluctuations playing a crucial role in men's mating efforts and intrasexually competitive behaviour, by hinting at a potential trade-off between competitive versus nurturing behaviour, which seems to be reflected in changes in interpersonal personality states. In addition, these T-modulated personality state changes were not only reported by oneself, but also recognized by naive observers. The latter fact may mean that these personality state changes associated with T fluctuations function as an intrasexual competitive signal detectable by observers, further supporting status competition and/or mate acquisition. On an evolutionary functional level (Tinbergen, 1963), this shows that both personality state changes and hormonal reactivity might play a crucial role in supporting important aspects of men's striving for a high reproductive success. This trade-off surrounding T variability in men can be embedded in a larger set of trade-offs in the realm of human reproduction. According to life history theory, an individual faces a range of trade-offs of allocating effort (especially time, energy and resources) to

tasks and traits in the pursuit of optimal fitness (Del Giudice, Gangestad & Kaplan, 2015). One such trade-off, which has been suggested to be regulated by acute T levels, is between mating and parental effort (Muehlenbein & Bribiescas, 2005) and should translate into competitive versus nurturing behaviours, with high T being related to the former (e.g., status acquisition) and low T to the latter (e.g., pair bonding) (van Anders et al., 2011). The interaction of T and C (a buffering of the association between T reactivity and increases in competitiveness and dominance by high baseline C) can be interpreted in light of life history theory. In particular, in times of high stress, C levels tend to be elevated, and reproductive effort decreases (Del Giudice, Ellis, & Shirtcliff, 2011).

Consequently, the status-seeking effects of increased T levels are attenuated in times of high stress, to limit an individual's extensive spending of resources and risky behaviour, in order to ensure survival. Thus, T reactivity in response to intrasexual challenges and mating opportunities may be one of many mechanisms in the calibration of immediate personality and behaviour, depending on contextual cues and the availability of resources, to achieve a high inclusive fitness, particularly in men (for an extensive review, see Gray, McHale, & Carré, 2017).

4.2. Observer-perceptions of T-mediated behavioural changes

We additionally demonstrated personality state changes perceptible by naïve observers based on short video recordings (thin slices of behaviour; Borkenau et al., 2004). The observers showed a moderate to high interrater agreement for both *pre* and *post* videos, and consensus was only slightly lower for *pre-post* changes. So far, it was unclear whether and how hormone-mediated behavioural changes are perceived by male and female observers outside the immediate competitive context at all (Mattan, Kubota, & Cloutier, 2017). Some previous studies have coded and rated behavioural facets and analysed these in association with baseline T and/or T reactivity. In previous female exposure studies, men's T reactivity was found to be related to female-perceived self-presentation behaviour (van der Meij et al., 2012) and female confederate's rating of how much the males tried to impress her (Roney et al., 2003). Our results regarding the link between T reactivity and *pre-post* changes in observer-perceived self-assurance strongly support these, only that we were not looking

at absolute behaviours, but behavioural changes. Moreover, these findings support previous interpretations of T increases being linked to status-seeking and -maintaining behaviours in a competitive context in men (Anderson & Kilduff, 2009; Eisenegger et al., 2011; Mehta & Josephs, 2010). We thus provide clear evidence that on the dimension self-assurance in particular, T-modulated behavioural changes are noticed by observers outside the original competitive context. Cooperativeness, however, appears to be rather orthogonal to the other two dimensions, at least in the interpersonal circumplex model (Wiggins, 1982). Consequently, it may be that T is only related to self-assurance, but unrelated to cooperativeness. Alternatively, since T reactivity was associated with self-reported changes in the facet between the Dominance axis and negative endpoint of the Love axis (i.e., competitiveness), it may well be that indeed state changes on a facet closely related to cooperativeness happened, but were not perceived accurately by observers. Similar associations between warmth and nurturance, and low T have been proposed earlier (van Anders et al., 2011). Finally, in some contexts, cooperativeness has been positively linked to T levels (e.g., in-group cooperation during inter-group competition; Reimers & Diekhoff, 2015). Since cooperativeness has been linked to both high and low T values depending on contextual variables, the null finding of our Study 3 is not surprising. Target men were not instructed to behave in a conflicting way. Consequently, they did not seem to have emitted specific signals, which were perceived and interpreted accordingly by observers. Thus, it would be interesting to study how the behaviour of target men would change in different settings, for example a competitive group task, and how this would be judged by observers.

Effects for self-reports in association with T reactivity were found not for any of the two interpersonal circumplex model's main axes, Dominance and Love, but for the facet in between Dominance and the negative pole of Love, competitiveness. For observer-perceptions, in contrast, no associations were found for any of the interpersonal circumplex facet, only for social impressions of self-assurance. These still seem complementary, since all fit well in an intrasexually competitive context. The differential findings for self-reports and observer-perceptions could be interpreted in

such a way that T fluctuations are related more to self-reported personality aspects of status-seeking (competitiveness), and observer-ratings more to confidence and hence current status (self-assurance, but not self-display, which would more fit into attempts of status-seeking; Hays & Bendersky, 2015). Of course, replications are called for, to see whether this slight differentiation regarding self-reports and observer-perceptions holds. Overall it can be concluded that most of the effects were located on the Dominance main axis and the competitiveness facet (as well as related social impressions), but not so much on the Love main axis, and not at all on the extraversion-introversion axis. This further corroborates the relevance of a T-modulated trade-off between mating and parental effort (e.g., Muehlenbein & Bribiescas, 2005; Muller, 2017; Zilioli & Bird, 2017), and associations along a competition-nurturance dimension (van Anders et al., 2011).

4.3. Testosterone x Cortisol interaction

Associations between men's T reactivity and increase in self-reported competitiveness and dominance were attenuated by high baseline C. This finding of a TxC interaction further corroborates previous reports that C may inhibit effects of T on status-related behaviours (e.g., Mehta et al., 2015; Mehta & Prasad, 2015; Sherman et al., 2016). Especially the finding regarding changes in dominance is exactly in line with the original study proposing the dual-hormone hypothesis (Mehta & Josephs, 2010), in that the positive relationship between dominance and T was only significant with simultaneously low C levels, which we showed for personality state changes and hormonal reactivity. Since we found a TxC interaction on changes in self-reports, but not observer-perceptions, it appears the buffering of T effects by baseline C is not related to perceptible state changes, in contrast to effects associated with T reactivity (changes in self-assurance). The moderation of the link of increases in self-reported competitiveness and dominance with T reactivity by baseline C can be interpreted in such a way that a T increase, as experienced in a competitive situation, is only converted into stronger status-seeking personality states when there is no shortage of available resources (i.e., low stress; Sherman et al., 2016). Hence, baseline C functions as a regulator between competition-induced T fluctuations and personality state changes. However, it is not a complete

gatekeeper, since we found main effects of T reactivity on changes in competitiveness as well (see Hamilton, Carré, Mehta, Olmstead, & Whitaker, 2015). Even when including baseline C without the TxC interaction in the model predicting changes in competitiveness, the effects of T reactivity (for the first *post*-sample) prevailed (no main effect of T reactivity was found for dominance). This shows that even though the effects of T reactivity were stronger when baseline C was low, activation of the hypothalamus pituitary gonadal (HPG) axis was still related to these personality state changes when controlling for baseline hypothalamus pituitary adrenal (HPA) axis activation (van Anders, Steiger, & Goldey, 2015). Thus, we provide further evidence for the dual-hormone hypothesis, at least for self-reports, and in terms of changes in competitive and dominant personality states.

4.4. Trait activation in a competitive context

The increases in competitiveness-related personality states we found can also be interpreted in terms of trait activation, since these personality dimensions fit well with the competitive context we created in the lab. According to trait activation theory, individuals express their personality traits when confronted with situational cues relevant to these traits (Tett & Burnett, 2003). In our study, aspects of the situation like being challenged by the competition, having the opportunity to win over another man and earn a monetary reward, and the presence of the attractive female may have functioned as primarily task-related and social cues (Tett & Burnett, 2003). Here, personality state changes can be seen as the primary manifestations of trait activation, with T reactivity as the underlying physiological mechanism. In the current study, we primarily wanted to make sure that T reactivity was triggered by our experimental manipulation of the lab situation, so we purposely combined several situational features that in the literature had been shown to trigger a T response, including a competitive intrasexual challenge with no immediately clear winner and exposure to a potential mate. To further disentangle which particular aspects may have triggered the personality state changes and T reactivity, future studies could employ more fine-grained control groups, such as staging a competition without involving a female confederate, or a non-competitive interaction between two men, simply substituting the female confederate by a male experimenter, or varying

the female confederate's attractiveness. This could also shed some light upon which aspects of the competitive situation activate which personality dimensions and have the most influence on T reactivity exactly (cf. Edelstein, Yim, & Quas, 2010; Roney, 2016).

Most of our significant associations between personality state changes and T reactivity were detected for the first, but not the second *post*-sample (Table 3). We chose to employ two *post* measures, since from previous studies it was not entirely clear when T reactivity was highest (Schultheiss et al., 2012), and previous studies were heterogeneous concerning the timing of reactivity measures (from immediately to 1 hour after a competition's end; Casto, Elliott, & Edwards, 2014; Trumble et al., 2012). Based on claims of a delay of 15-20 minutes for hormonal reactivity to be detectable in saliva (Schultheiss et al.), our findings could be interpreted as follows: changes in self-reported competitiveness and in observer-rated self-assurance are linked to anticipatory reactivity (Marler, Oyegbile, Plavicki, & Trainor, 2005) and a T increase during the competition's first minutes. Increases in observer-perceived dominance, in turn, relate to T reactivity during the full competition phase and especially the later disciplines. Of course, these interpretations should be treated with care, since there is large intra- and inter-individual variations in hormonal levels (especially diurnal declines; Schultheiss & Stanton, 2009), so further studies are required to see if these represent meaningful differences (Casto & Edwards, 2016a,b). From our findings we could also conclude that the first *post*-sample was timed better for detecting relationships with personality state changes, and future studies may follow the protocol of assessing T reactivity slightly earlier.

4.5. Strengths and limitations

Compared to the social endocrinology literature and considering the rather complex design, we recruited a large sample, providing relatively high statistical power to detect significant effects. We also recruited a very large number of male and female raters for the video ratings and replicated results for the observer ratings across multiple rater groups and operationalizations (adjective and social impression ratings). Moreover, we used a multi-method approach including physiological measures, self-reported and observer-judged personality states based on questionnaires and video

recordings (Vazire, 2006), allowing us to comprehensively study the association between hormones and personality from different angles. We considered not only isolated effects of a single hormone (T), but also the interaction with a second hormone (C), since often it is co-released endocrine signals that affect behaviour and personality (Roney, 2016). Furthermore, we implemented a control group to check for changes in perceived personality states and social impressions due to practice effects from *pre* to *post* video recordings. Our dyadic competition was composed in such a way that it presumably was at least somewhat relevant to all participants, since we employed a mixture of cognitive and more physical disciplines. Men with different skill sets should have perceived similar chances to win the competition. We selected disciplines in which participants could be expected to not have too much experience (e.g., table pinball game rather than simple foosball, since the former is less common). We designed our competition to be as competitive as possible, with opponents being seated directly opposite each other, and presumably being motivated by the monetary incentives and presence of the attractive female (van der Meij et al., 2010). The T reactivity and personality state changes we found relative to the control group indicate that our manipulation was successful. Finally, results concerning the effects of our competition can be seen as at least as or even more generalizable than previous findings, since we employed a more realistic and natural kind of competition (including a classic “male” competitive discipline, arm wrestling), compared to computerized tasks such as the Point Subtraction Aggression Paradigm (Carré et al., 2009). In addition, our competition outcome emerged naturally (being more credible for the participants than rigged competitions; e.g., Geniole et al., 2013). Still, our study took place in a laboratory setting, which often shows a limited but satisfactory generalizability (Sherman et al., 2016), so that replications in more natural contexts are required.

Regarding limitations, it has been questioned to what extent salivary T is a good estimate of free unbound T as measured in serum, thus potentially limiting the validity of our findings. However, Fiers and colleagues (2014) recently pointed out measurement bias may be less influential in men than in women, due to lower T levels in the latter. Moreover, a superior validity of salivary

measurement using passive drooling (as in our study) was shown compared to salivettes. We conducted the saliva collection with great care following a strict procedure (e.g., Granger, Shirtcliff, Booth, Kivlighan, & Schwartz, 2004), subsequently storing the samples at -80°C (where hormone levels are stable for years; Granger et al.), and assessed a wide range of confounding variables (Schultheiss & Stanton, 2009). Besides the significant T reactivity for both *post*-samples in the experimental group (relative to baseline T), we also detected a significant T increase in the control condition for the second (but not first) sample. It is rather unlikely this can be explained by the participants' watching the documentary video, since the documentary was rated to be informative, but average on items such as challenging, exciting, boring, and significantly below the midpoint of the scale for disquieting and stressful. Rather, the T increase may be ascribed to participants being stressed and challenged by the video-recorded self-presentation, especially since the question was framed in a challenging way ("What is great about yourself?"). This is corroborated by the fact that these participants also showed a C increase. Since greater T and personality state reactivity was shown in the experimental compared to the control group and further analyses involving T reactivity focussed on the experimental group, the control group's T reactivity does not constitute a serious issue. Still, we show how relatively little manipulation is necessary to induce a hormonal response, especially in the control group, and only exactly why T increased here remains inconclusive. It also needs to be acknowledged that the experimental and control groups differed considerably in sample size. This was done purposely, because the main aim of the control group was to examine whether T increases and/or personality state changes were also found without exposure to the two main experimental manipulations (dyadic competition and female presence), for which we see a sample size of 40 as sufficient. Analyses on associations between T reactivity and personality state changes, or moderation analyses involving baseline C, were not planned for and conducted in the control group.

For some self-reported personality state facets we have to acknowledge only moderate internal consistency (especially competitiveness and ingenuousness; Table S4a). This can likely be

ascribed to item selection, since we selected five instead of the originally eight items per facet (Jacobs & Scholl, 2005) we selected five, mainly for reasons of brevity and to reduce strain on the raters. Moreover, internal consistency may have been lower because we employed a scale designed for the assessment of personality traits (e.g., Turan et al., 2014) to measure states. Results for the facets with low internal consistencies might thus be underestimates, attenuated by the low reliabilities. Finally, it is important to acknowledge that analyses of correlated changes, such as in this study on personality state changes and hormonal reactivity, are statistically difficult. Despite comparably large sample size, power to detect such effects was only moderate. Consequently, even though our results are theoretically sound, they need to be taken with care until further replication.

4.6. Future research

There is an abundance of possible pathways to further elucidate hormonal and personality state changes as well as their associations in intrasexually competitive contexts. First of all, this study was restricted deliberately in terms of age (focussing on participants presumably active on the mating market and hence especially responsive to an attractive female confederate) and sex (only males). To assess the generalizability of our findings beyond these, similar studies could be conducted in different samples. Even though T supposedly plays a more minor role in females than in males, and T levels are considerably lower in the former, future research is needed to corroborate whether hormone-personality associations in an intrasexual competitive situation are similar across the sexes (Mehta et al., 2015). Especially the role of competitiveness would be interesting to investigate in women, who are generally found to be lower on this trait than men, presumably since over human evolution women purportedly competed less for mates and other resources than men did (Owens, 2017). A first study (Hahn, Fisher, Cobey, DeBruine, & Jones, 2016) has shown a positive association between baseline T and self-reported intrasexual competitiveness ($N = 136$ women). It would hence be interesting to examine whether such associations can be replicated involving competition-induced T fluctuations in women. Further studies could investigate effects of competitive interactions in other age groups, such as adolescents, a phase marked by increased aggression and

risk-taking. These behaviours would be interesting to examine on a state basis in association with hormonal reactivity (Gray et al., 2017). Additionally, the effects of hormonal reactivity in intrasexual competition could be studied in older male adults, at an age when T levels declined considerably, compared to early adulthood (Keevil et al., 2017). A particularly interesting sample to disentangle age and relationship status effects on the results we found could be middle-aged and older men that transitioned from stable romantic relationships back to singlehood, as re-entering the mating market should shift their life history priorities back from parental and nurturing effort to mating and status competition. Since our findings are based on a western sample mainly from the student population, replications in non-student samples as well as different, non-western cultures are necessary, the latter to assess cross-cultural consistency of our findings. For example, previous research showed intercultural differences in baseline T (Bribiescas, 1998), potentially due to variation in men's reproductive effort (Alvergne, Faurie, & Raymond, 2009).

Moreover, for an evolutionarily complete account of the association between hormonal reactivity and personality variability in an intrasexually competitive context, one needs to take into account further levels of analysis, besides the overly proximate mechanisms considered in the present article (Simpson, Griskevicius, & Kim, 2011; Tinbergen, 1963). A longitudinal follow-up study could investigate functional consequences of acute T reactivity and personality state changes, and hence associations with men's mating success and ultimately reproductive fitness (ideally towards the end of men's reproductive period and hence their lifespan; Pollet, Cobey, & van der Meij, 2013). This would provide insights into the adaptiveness and potential adaptation (e.g. informing about mechanisms of sexual selection; Puts, 2016) of extant human individual differences (e.g., Dall, Houston, & McNamara, 2004; Muehlenbein, 2006). Finally, to further elaborate on phylogenetic mechanisms, comparative studies in related species, such as nonhuman primates, are required (Eisenegger et al., 2011; Fuxjager, Trainor, & Marler, 2017). Testing the challenge hypothesis, several studies have shown T increases in competitive mating contexts in species such as chimpanzees (Muller & Wrangham, 2004). Ostner, Heistermann, and Schülke (2011) found a positive relationship

between aggressiveness and fecal androgens (in particular, immunoreactive epiandrosterone (iEA), a major metabolite of testosterone in macaque feces) in male Assamese macaques over a 16-month period. Further studies could investigate associations between competition-induced changes in T and personality states in nonhuman primate species to shed further light on the phylogeny and evolutionary basis of this study's findings.

A crucial question we could not fully address is that of causality. We detected larger increases in self-reported competitiveness with stronger T reactivity in the experimental group. Since these are correlated changes, it remains unclear whether T reactivity caused these personality state changes (Carré & Olmstead, 2015). It remains possible that the T response did not have any causal influence at all, and the personality change could be ascribed to a third variable. Moreover, this study was not designed to disentangle which particular aspect of the experimentally manipulated context (intrasexual competition, female exposure, or their combination) had the largest impact on state and hormonal changes. To be able to make causal interpretations, one way would be to administer T and subsequently measure its effects on personality states (McCall & Singer, 2012).

Regarding diverging self-reported and observed personality state changes, it would be interesting to analyse more objectively which changes in mimics and gestures mediated observable personality state changes. Objective behaviours such as gaze direction, smiling/laughing, illustrators (communicative gestures) and adaptors (non-illustrative hand movements) (Penke & Asendorpf, 2008) could be coded and analysed in conjunction with personality and hormonal changes. This would provide further insights into how competition-induced personality state changes, partly mediated by T increases, facilitate social signalling.

Geniole and colleagues (2011) note that hormonal effects may to a large extent depend on contextual factors (e.g., opponent's psychological state such as self-efficacy and dominance; van der Meij et al., 2010; social inclusion vs. exclusion, stable vs. unstable hierarchies; Knight & Mehta, 2014), and that these have not been acknowledged sufficiently in the extant literature. Further

research on contextual factor is called for, to further examine the robustness of our and previous findings, and the strength of a “winner effect”, if it exists at all. Future studies could explicitly vary the context in which hormone-personality/behaviour interactions are being examined (Gleason, Fuxjager, Oyegbile, & Marler, 2009). The effect of an audience varying in sex ratios and responsivity (Ronay & von Hippel, 2010), the kind of competition (more sportive or mating-related), or men’s behaviour in an actual mating situation could be assessed. In the latter, a group of single males and females interacting freely in an externally valid dating context could be observed in order to study the conjunction of hormonal and personality variables as well as mating outcomes. Such studies would provide further insight into the complex nature of interactions between contextual factors and hormonal associations with behaviour in competitive situations, specifically, and social interactions, more generally (McCall & Singer, 2012).

5. Conclusion

In this preregistered study, we demonstrated how hormonal and personality state changes co-occurred in men engaging in an intrasexual competition. We showed a T response after the competition, in line with the challenge hypothesis (Archer, 2006), as well as changes in personality states, recognized not only by the men themselves, but also by outside male and female observers. The larger the T reactivity, the higher increases in self-reported competitiveness, as well as observer-perceived self-assurance were found. Furthermore, as predicted by the dual-hormone hypothesis (Mehta & Josephs, 2010), an interaction between T reactivity and baseline C on self-reported personality state changes in competitiveness and dominance was found. Our results stress the importance of considering T reactivity during social interactions as a key modulator of personality state changes and social behaviour (Carré et al., 2014). These findings are also in line with previous accounts of T modulating a life history trade-off between mating and parental effort (e.g., Muehlenbein & Bribiescas, 2005; Muller, 2017; Zilioli & Bird, 2017), and with the steroid/peptide theory of social bonds, according to which high T is related to competitive and low T to nurturing behaviours (van Anders et al., 2011). In addition, outside observers were able to detect personality

state changes in dominance and self-assurance that co-occurred with T reactivity based on thin slices of men's behaviour, suggesting that T may indeed be functional in social signalling towards rivals and potential mates (Puts, 2010).

References

- Abad-Tortosa, D., Alacreu-Crespo, A., Costa, R., Salvador, A., & Serrano, M. Á. (2017). Sex differences in autonomic response and situational appraisal of a competitive situation in young adults. *Biological Psychology*, 126, 61-70. <http://doi.org/10.1016/j.biopsycho.2017.04.008>
- Alvergne, A., Faurie, C., & Raymond, M. (2009). Variation in testosterone levels and male reproductive effort: Insight from a polygynous human population. *Hormones and Behavior*, 56, 491-497. <http://doi.org/10.1016/j.yhbeh.2009.07.013>

- Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, 32, 201-271. [http://doi.org/10.1016/S0065-2601\(00\)80006-4](http://doi.org/10.1016/S0065-2601(00)80006-4)
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111, 256-274. <http://doi.org/10.1037/0033-2909.111.2.256>
- Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology*, 96, 491-503. <http://doi.org/10.1037/a0014201>
- Andersson, M. B. (1994). *Sexual selection*. Princeton, NJ: Princeton University Press.
- Apicella, C. L., Dreber, A., Gray, P. B., Hoffman, M., Little, A. C., & Campbell, B. C. (2011). Androgens and competitiveness in men. *Journal of Neuroscience, Psychology, and Economics*, 4, 54-62. <http://doi.org/10.1037/a0021979>
- Archer, J. (1998). Problems with the concept of dominance and lack of empirical support for a testosterone-dominance link. *Behavioral and Brain Sciences*, 21, 363. <http://doi.org/10.1017/S0140525X98221224>
- Archer, J. (2006). Testosterone and human aggression: An evaluation of the challenge hypothesis. *Neuroscience & Biobehavioral Reviews*, 30, 319-345. <http://doi.org/10.1016/j.neubiorev.2004.12.007>
- Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences*, 32, 249-266. <http://doi.org/10.1017/S0140525X09990951>
- Archer, J., Graham-Kevan, N., & Davies, M. (2005). Testosterone and aggression: A reanalysis of Book, Starzyk, and Quinsey's (2001) study. *Aggression and Violent Behavior*, 10, 241-261. <http://doi.org/10.1016/j.avb.2004.01.001>
- Arslan, R.C., & Tata, C.S. (2017). *formr.org* [Survey Software] (Version v0.16.12). <http://doi.org/10.5281/zenodo.823627>
- Bhasin, S., Woodhouse, L., & Storer, T. W. (2001). Proof of the effect of testosterone on skeletal muscle. *Journal of Endocrinology*, 170, 27-38. <http://doi.org/10.1677/joe.0.1700027>

- Book, A. S., Starzyk, K. B., & Quinsey, V. L. (2001). The relationship between testosterone and aggression: A meta-analysis. *Aggression and Violent Behavior, 6*, 579-599. [http://doi.org/10.1016/S1359-1789\(00\)00032-X](http://doi.org/10.1016/S1359-1789(00)00032-X)
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599-614. <http://doi.org/10.1037/0022-3514.86.4.599>
- Bribiescas, R. G. (1998). Testosterone and dominance: Between-population variance and male energetics. *Behavioral and Brain Sciences, 21*, 364-365. <http://doi.org/10.1017/S0140525X98241227>
- Burt, K. B., & Obradović, J. (2013). The construct of psychophysiological reactivity: Statistical and psychometric issues. *Developmental Review, 33*, 29-57. <http://doi.org/10.1016/j.dr.2012.10.002>
- Buunk, A. P., & Fisher, M. (2009). Individual differences in intrasexual competition. *Journal of Evolutionary Psychology, 7*, 37-48. <http://doi.org/10.1556/JEP.7.2009.1.5>
- Carré, J. M., & Archer, J. (2018). Testosterone and Human Behavior: The role of individual and contextual variables. *Current Opinion in Psychology, 19*, 149-153. <http://doi.org/10.1016/j.copsyc.2017.03.021>
- 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. <http://doi.org/10.1016/j.psyneuen.2014.06.011>
- 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*, 2034-2041. <http://doi.org/10.1016/j.psyneuen.2013.03.008>
- Carré, J. M., Iselin, A. M. R., Welker, K. M., Hariri, A. R., & Dodge, K. A. (2014). Testosterone reactivity to provocation mediates the effect of early intervention on aggressive behavior. *Psychological Science, 25*, 1140-1146. <http://doi.org/10.1177/09567979614525642>
- 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*, 403-409. <http://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. <http://doi.org/10.1016/j.neuroscience.2014.11.029>

- Carré, J. M., Putnam, S. K., & McCormick, C. M. (2009). Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. *Psychoneuroendocrinology*, 34, 561-570.
<http://doi.org/10.1016/j.psyneuen.2008.10.018>
- Carré, J. M., Ruddick, E. L., Moreau, B. J., & Bird, B. M. (2017). Testosterone and human aggression. In P. Sturmeij (Ed.), *The Wiley Handbook of Violence and Aggression (Vol. I: Definition, Conception, and Development)*. Hoboken, New Jersey: John Wiley & Sons.
- Casto, K. V., & Edwards, D. A. (2016a). Testosterone, cortisol, and human competition. *Hormones and Behavior*, 82, 21-37. <http://doi.org/10.1016/j.yhbeh.2016.04.004>
- Casto, K. V., & Edwards, D. A. (2016b). Before, during, and after: How phases of competition differentially affect testosterone, cortisol, and estradiol levels in women athletes. *Adaptive Human Behavior and Physiology*, 2, 11-25. <http://doi.org/10.1007/s40750-015-0028-2>
- Casto, K. V., Elliott, C., & Edwards, D. A. (2014). Intercollegiate cross country competition: Effects of warm-up and racing on salivary levels of cortisol and testosterone. *International Journal of Exercise Science*, 7, 318-328.
- Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research*, 66, 1261-1266. <http://doi.org/10.1016/j.jbusres.2012.02.023>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. <http://doi.org/10.1037/0033-2909.112.1.155>
- Costa, R., Serrano, M. A., & Salvador, A. (2016). Importance of self-efficacy in psychoendocrine responses to competition and performance in women. *Psicothema*, 28, 66-70.
<http://doi.org/10.7334/psicothema2015.166>
- Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology*, 81, 322-331.
<http://doi.org/10.1037/0022-3514.81.2.322>
- Dabbs, J. M. (1997). Testosterone, smiling, and facial appearance. *Journal of Nonverbal Behavior*, 21, 45-55.
<http://doi.org/10.1023/A:1024947801843>
- Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. *Ecology Letters*, 7, 734-739.
<http://doi.org/10.1111/j.1461-0248.2004.00618.x>

- Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The adaptive calibration model of stress responsivity. *Neuroscience and Biobehavioral Reviews*, 35, 1562-1592.
<http://doi.org/10.1016/j.neubiorev.2010.11.007>
- Del Giudice, M., Gangestad, S. W., & Kaplan, H. S. (2015). Life history theory and evolutionary psychology. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology (Vol. 1: Foundations)*; pp. 88-114). Hoboken, New Jersey: John Wiley & Sons.
- DeSoto, M. C., Hitlan, R. T., Deol, R.-S. S., & McAdams, D. (2010). Testosterone fluctuations in young men: The difference between interacting with like and not-like others. *Evolutionary Psychology*, 8, 173-188. <http://doi.org/10.1177/147470491000800203>
- Dettenborn, L., Kirschbaum, C., Gao, W., Spitzer, C., Roepke, S., Otte, C., & Wingenfeld, K. (2016). Increased hair testosterone but unaltered hair cortisol in female patients with borderline personality disorder. *Psychoneuroendocrinology*, 71, 176-179. <http://doi.org/10.1016/j.psyneuen.2016.05.026>
- Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: Animal personality meets individual plasticity. *Trends in Ecology & Evolution*, 25, 81-89.
<http://doi.org/10.1016/j.tree.2009.07.013>
- Edelstein, R. S., Chopik, W. J., & Kean, E. L. (2011). Sociosexuality moderates the association between testosterone and relationship status in men and women. *Hormones and Behavior*, 60, 248-255.
<http://doi.org/10.1016/j.yhbeh.2011.05.007>
- Edelstein, R. S., Yim, I. S., & Quas, J. A. (2010). Narcissism predicts heightened cortisol reactivity to a psychosocial stressor in men. *Journal of Research in Personality*, 44, 565-572.
<http://doi.org/10.1016/j.jrp.2010.06.008>
- Edwards, D. A., Wetzel, K., & Wyner, D. R. (2006). Intercollegiate soccer: Saliva cortisol and testosterone are elevated during competition, and testosterone is related to status and social connectedness with teammates. *Physiology and Behavior*, 87, 135-143. <http://doi.org/10.1016/j.physbeh.2005.09.007>
- Eisenegger, C., Haushofer, J., & Fehr, E. (2011). The role of testosterone in social interaction. *Trends in Cognitive Sciences*, 15, 263-271. <http://doi.org/10.1016/j.tics.2011.04.008>
- Eisenegger, C., Kumsta, R., Naef, M., Gromoll, J., & Heinrichs, M. (2017). Testosterone and androgen receptor gene polymorphism are associated with confidence and competitiveness in men. *Hormones and Behavior*, 92, 93-102. <http://doi.org/10.1016/j.yhbeh.2016.09.011>

- Elias, M. (1981). Serum cortisol, testosterone, and testosterone-binding globulin responses to competitive fighting in human males. *Aggressive Behavior*, 7, 215-224. [http://doi.org/10.1002/1098-2337\(1981\)7:3<215::AID-AB2480070305>3.0.CO;2-M](http://doi.org/10.1002/1098-2337(1981)7:3<215::AID-AB2480070305>3.0.CO;2-M)
- Exton, M. S., Krüger, T. H., Bursch, N., Haake, P., Knapp, W., Schedlowski, M., & Hartmann, U. (2001). Endocrine response to masturbation-induced orgasm in healthy men following a 3-week sexual abstinence. *World Journal of Urology*, 19, 377-382. <http://doi.org/10.1007/s003450100222>
- Fales, M. R., Gildersleeve, K. A., & Haselton, M. G. (2014). Exposure to perceived male rivals raises men's testosterone on fertile relative to nonfertile days of their partner's ovulatory cycle. *Hormones and Behavior*, 65, 454-460. <http://doi.org/10.1016/j.yhbeh.2014.04.002>
- Farrelly, D., Owens, R., Elliott, H. R., Walden, H. R., & Wetherell, M. A. (2015). The effects of being in a “new relationship” on levels of testosterone in men. *Evolutionary Psychology*, 13, 250-261. <http://doi.org/10.1177/147470491501300116>
- Fiers, T., Delanghe, J., T'Sjoen, G., Van Caenegem, E., Wierckx, K. & Kaufman, J. M. (2014). A critical evaluation of salivary testosterone as a method for the assessment of serum testosterone. *Steroids*, 86, 5-9. <http://doi.org/10.1016/j.steroids.2014.04.013>
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11, 77-83. <http://doi.org/10.1016/j.tics.2006.11.005>
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80, 1011-1027. <http://doi.org/10.1037/0022-3514.80.6.1011>
- Foo, Y. Z., Nakagawa, S., Rhodes, G., & Simmons, L. W. (2017). The effects of sex hormones on immune function: A meta-analysis. *Biological Reviews*, 92, 551-571. <http://doi.org/10.1111/brv.12243>
- Fothergill, M., Wolfson, S., & Neave, N. (2017). Testosterone and cortisol responses in male soccer players: The effect of home and away venues. *Physiology & Behavior*, 177, 215-220. <http://doi.org/10.1016/j.physbeh.2017.04.021>
- Frederick, D. A., & Haselton, M. G. (2007). Why is muscularity sexy? Tests of the fitness indicator hypothesis. *Personality and Social Psychology Bulletin*, 33, 1167-1183. <http://doi.org/10.1177/0146167207303022>

- Fuxjager, M. J., Mast, G., Becker, E. A., & Marler, C. A. (2009). The 'home advantage' is necessary for a full winner effect and changes in post-encounter testosterone. *Hormones and Behavior*, 56, 214-219. <http://doi.org/10.1016/j.yhbeh.2009.04.009>
- 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. <http://doi.org/10.1016/j.yhbeh.2016.10.002>
- Geniole, S. N., Busseri, M. A., & McCormick, C. M. (2013). Testosterone dynamics and psychopathic personality traits independently predict antagonistic behavior towards the perceived loser of a competitive interaction. *Hormones and Behavior*, 64, 790-798. <http://doi.org/10.1016/j.yhbeh.2013.09.005>
- Geniole, S. N., Carré, J. M., & McCormick, C. M. (2011). State, not trait, neuroendocrine function predicts costly reactive aggression in men after social exclusion and inclusion. *Biological Psychology*, 87, 137-145. <http://doi.org/10.1016/j.biopsycho.2011.02.020>
- Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C., & Back, M. D. (2017). Trait personality and state variability: Predicting individual differences in within- and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, 69, 124-138. <http://doi.org/10.1016/j.jrp.2016.06.003>
- 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, 16194-16199. <http://doi.org/10.1073/pnas.1105403108>
- Gleason, E. D., Fuxjager, M. J., Oyegbile, T. O., & Marler, C. A. (2009). Testosterone release and social context: When it occurs and why. *Frontiers in Neuroendocrinology*, 30, 460-469. <http://doi.org/10.1016/j.yfrne.2009.04.009>
- Goldey, K. L., & van Anders, S. M. (2016). Identification with stimuli moderates women's affective and testosterone responses to self-chosen erotica. *Archives of Sexual Behavior*, 45, 2155-2171. <http://doi.org/10.1007/s10508-015-0612-3>
- Granger, D. A., Shirtcliff, E. A., Booth, A., Kivlighan, K. T., & Schwartz, E. B. (2004). The "trouble" with salivary testosterone. *Psychoneuroendocrinology*, 29, 1229-1240. <http://doi.org/10.1016/j.psyneuen.2004.02.005>

- Gray, P. B., McHale, T. S., & Carré, J. M. (2017). A review of human male field studies of hormones and behavioral reproductive effort. *Hormones and Behavior*, 91, 52-67.
<http://doi.org/10.1016/j.yhbeh.2016.07.004>
- Hahn, A. C., Fisher, C. I., Cobey, K. D., DeBruine, L. M., & Jones, B. C. (2016). A longitudinal analysis of women's salivary testosterone and intrasexual competitiveness. *Psychoneuroendocrinology*, 64, 117-122. <http://doi.org/10.1016/j.psyneuen.2015.11.014>
- Hamilton, L. D., Carré, J. M., Mehta, P. H., Olmstead, N., & Whitaker, J. D. (2015). Social neuroendocrinology of status: A review and future directions. *Adaptive Human Behavior and Physiology*, 1, 202-230.
<http://doi.org/10.1007/s40750-015-0025-5>
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, 7, 17-52. [http://doi.org/10.1016/0022-5193\(64\)90039-6](http://doi.org/10.1016/0022-5193(64)90039-6)
- Hasegawa, M., Toda, M., & Morimoto, K. (2008). Changes in salivary physiological stress markers associated with winning and losing. *Biomedical Research*, 29, 43-46. <http://doi.org/10.2220/biomedres.29.43>
- Hays, N. A., & Bendersky, C. (2015). Not all inequality is created equal: Effects of status versus power hierarchies on competition for upward mobility. *Journal of Personality and Social Psychology*, 108, 867-882. <http://doi.org/10.1037/pspo0000017>
- Hellhammer, D. H., Hubert, W., & Schürmeyer, T. (1985). Changes in saliva testosterone after psychological stimulation in men. *Psychoneuroendocrinology*, 10, 77-81. [http://doi.org/10.1016/0306-4530\(85\)90041-1](http://doi.org/10.1016/0306-4530(85)90041-1)
- Horowitz, L. M., Wilson, K. R., Turan, B., Zolotsev, P., Constantino, M. J., & Henderson, L. (2006). How interpersonal motives clarify the meaning of interpersonal behavior: A revised circumplex model. *Personality and Social Psychology Review*, 10, 67-86. http://doi.org/10.1207/s15327957pspr1001_4
- Idris, F. P., Wan, Y., Zhang, X., & Punyadeera, C. (2017). Within-day baseline variation in salivary biomarkers in healthy men. *OMICS: A Journal of Integrative Biology*, 21, 74-80.
<http://doi.org/10.1089/omi.2016.0168>
- Jacobs, I., & Scholl, W. (2005). Interpersonale Adjektivliste (IAL). *Diagnostica*, 51, 145-155.
<http://doi.org/10.1026/0012-1924.51.3.145>
- Keevil, B. G., Clifton, S., Tanton, C., Macdowall, W., Copas, A. J., Lee, D., ... & Mercer, C. H. (2017). Distribution of salivary testosterone in men and women in a British general population-based

sample: The third national survey of sexual attitudes and lifestyles (Natsal-3). *Journal of the Endocrine Society*, 1, 14-25. <http://doi.org/10.1210/js.2016-1029>

Kenny, D. A., Kashy, D. A., & Cook, W. (2006). *Dyadic data analysis*. New York: Guilford Press.

Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia: W. B. Saunders.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' – A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28, 76-81. <http://doi.org/10.1159/000119004>

Kivlighan, K. T., Granger, D. A., & Booth, A. (2005). Gender differences in testosterone and cortisol response to competition. *Psychoneuroendocrinology*, 30, 58-71. <http://doi.org/10.1016/j.psyneuen.2004.05.009>

Knight, E. L., & Mehta, P. H. (2014). Hormones and hierarchies. In J. T. Cheng, J. L. Tracy, & C. Anderson (Eds.), *The Psychology of Social Status* (pp. 269-301). New York, NY: Springer International Publishing.

Kordsmeyer, T. L., Hunt, J., Puts, D. A., Ostner, J., & Penke, L. (2018). The relative importance of intra-and intersexual selection on human male sexually dimorphic traits. *Evolution and Human Behavior*, 39, 424-436. <http://doi.org/10.1016/j.evolhumbehav.2018.03.008>

Lakens, D. (2016, March 17). *One-sided tests: Efficient and underused* [Blog post]. Retrieved from <http://daniellakens.blogspot.de/2016/03/one-sided-tests-efficient-and-underused.html>

Leary, T. (1957). *Interpersonal diagnosis of personality*. New York: Ronald.

Lopez, H.H., Hay, A.C. & Conklin, P.H. (2009). Attractive men induce testosterone and cortisol release in women. *Hormones and Behavior*, 56, 84-92. <http://doi.org/10.1016/j.yhbeh.2009.03.004>

Lorenz, T. K., Heiman, J. R., & Demas, G. E. (2017). Testosterone and immune-reproductive tradeoffs in healthy women. *Hormones and Behavior*, 88, 122-130. <http://doi.org/10.1016/j.yhbeh.2016.11.009>

Määttänen, I., Jokela, M., Hintsala, T., Firtser, S., Kähönen, M., Jula, A., ... & Keltikangas-Järvinen, L. (2013). Testosterone and temperament traits in men: Longitudinal analysis. *Psychoneuroendocrinology*, 38, 2243-2248. <http://doi.org/10.1016/j.psyneuen.2013.04.009>

- Marler, C. A., Oyegbile, T. O., Plavicki, J., & Trainor, B. C. (2005). Response to Wingfield's commentary on "A continuing saga: The role of testosterone in aggression". *Hormones and Behavior*, 48, 256-258.
<http://doi.org/10.1016/j.yhbeh.2005.05.010>
- Matas, D., & Koren, L. (2016, May 31st). A simple method for measuring long-term integrated testosterone levels in men. Retrieved from peerj.com/preprints/2069
- Mattan, B. D., Kubota, J. T., & Cloutier, J. (2017). How social status shapes person perception and evaluation: A social neuroscience perspective. *Perspectives on Psychological Science*, 12, 468-507.
<http://doi.org/10.1177/1745691616677828>
- Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces*, 64, 377-402.
<http://doi.org/10.1093/sf/64.2.377>
- Mazur, A. (2005). *Biosociology of dominance and deference*. Lanham, Maryland: Rowman & Littlefield Publishers.
- Mazur, A. (2015). A biosocial model of status in face-to-face groups. In V. Zeigler-Hill, L. L. M. Welling, & T. K. Shackelford (Eds.), *Evolutionary Perspectives on Social Psychology* (pp. 303-315). New York, NY: Springer International Publishing.
- Mazur, A., Welker, K. M., & Peng, B. (2015). Does the biosocial model explain the emergence of status differences in conversations among unacquainted men? *PloS One*, 10, e0142941.
<http://doi.org/10.1371/journal.pone.0142941>
- McCall, C., & Singer, T. (2012). The animal and human neuroendocrinology of social cognition, motivation and behavior. *Nature Neuroscience*, 15, 681-688. <http://doi.org/10.1038/nn.3084>
- McHale, T. S., Zava, D. T., Hales, D., & Gray, P. B. (2016). Physical competition increases dehydroepiandrosterone (DHEA) and androstenedione rather than testosterone among juvenile boy soccer players. *Adaptive Human Behavior and Physiology*, 2, 44-56. <http://doi.org/10.1007/s40750-015-0030-8>
- Mehta, P. H., & Josephs, R. A. (2006). Testosterone change after losing predicts the decision to compete again. *Hormones and Behavior*, 50, 684-692. <http://doi.org/10.1016/j.yhbeh.2006.07.001>
- Mehta, P. H., & Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior*, 58, 898-906.
<http://doi.org/10.1016/j.yhbeh.2010.08.020>

- Mehta, P. H., & Prasad, S. (2015). The dual-hormone hypothesis: A brief review and future research agenda. *Current opinion in Behavioral Sciences*, 3, 163-168. <http://doi.org/10.1016/j.coheba.2015.04.008>
- Mehta, P. H., Snyder, N. A., Knight, E. L., & Lassetter, B. (2015). Close versus decisive victory moderates the effect of testosterone change on competitive decisions and task enjoyment. *Adaptive Human Behavior and Physiology*, 1, 291-311. <http://doi.org/10.1007/s40750-014-0014-0>
- Mehta, P. H., Welker, K. M., Zilioli, S., & Carré, J. M. (2015). Testosterone and cortisol jointly modulate risk-taking. *Psychoneuroendocrinology*, 56, 88-99. <http://doi.org/10.1016/j.psyneuen.2015.02.023>
- Miller, G. (2000). The mating mind: How sexual selection shaped the evolution of human nature. New York: Doubleday.
- Muehlenbein, M. P. (2006). Adaptive variation in testosterone levels in response to immune activation: Empirical and theoretical perspectives. *Social Biology*, 53, 13-23. <http://doi.org/10.1080/19485565.2006.9989113>
- Muehlenbein, M. P., & Bribiescas, R. G. (2005). Testosterone-mediated immune functions and male life histories. *American Journal of Human Biology*, 17, 527-558. <http://doi.org/10.1002/ajhb.20419>
- Muller, M. N. (2017). Testosterone and reproductive effort in male primates. *Hormones and Behavior*, 91, 36-51. <http://doi.org/10.1016/j.yhbeh.2016.09.001>
- Muller, M. N., & Wrangham, R. W. (2004). Dominance, aggression and testosterone in wild chimpanzees: A test of the 'challenge hypothesis'. *Animal Behaviour*, 67, 113-123. <http://doi.org/10.1016/j.anbehav.2003.03.013>
- Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.
- Nave, G., Nadler, A., Zava, D., & Camerer, C. (2017). Single dose testosterone administration impairs cognitive reflection in men. *Psychological Science*, 28, 1398-1407. <http://doi.org/10.1177/0956797617709592>
- Nezlek, J. B., & Plesko, R. M. (2003). Affect- and self-based models of relationships between daily events and daily well-being. *Personality and Social Psychology Bulletin*, 29, 584-596. <http://doi.org/10.1177/0146167203029005004>
- Ostner, J., Heistermann, M., & Schülke, O. (2011). Male competition and its hormonal correlates in Assamese macaques (*Macaca assamensis*). *Hormones and Behavior*, 59, 105-113. <http://doi.org/10.1016/j.yhbeh.2010.10.017>

- Owens, R. (2017). *The role of life history variables in male competitive behaviour* (Doctoral dissertation, University of Sunderland). Retrieved from <http://sure.sunderland.ac.uk/7031/>
- Penke, L. & Asendorpf, J.B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, 95, 1113-1135. <http://doi.org/10.1037/0022-3514.95.5.1113>
- Pollet, T. V., Cobey, K. D., & van der Meij, L. (2013). Testosterone levels are negatively associated with childlessness in males, but positively related to offspring count in fathers. *PloS One*, 8, e60018. <http://doi.org/10.1371/journal.pone.0060018>
- Pollet, T. V., & van der Meij, L. (2017). To remove or not to remove: The impact of outlier handling on significance testing in testosterone data. *Adaptive Human Behavior and Physiology*, 3, 43-60. <http://doi.org/10.1007/s40750-016-0050-z>
- Popma, A., Vermeiren, R., Geluk, C. A., Rinne, T., van den Brink, W., Knol, D. L., ... & Doreleijers, T. A. (2007). Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. *Biological Psychiatry*, 61, 405-411. <http://doi.org/10.1016/j.biopsych.2006.06.006>
- Puts, D. A. (2010). Beauty and the beast: Mechanisms of sexual selection in humans. *Evolution and Human Behavior*, 31, 157-175. <http://doi.org/10.1016/j.evolhumbehav.2010.02.005>
- Puts, D. (2016). Human sexual selection. *Current Opinion in Psychology*, 7, 28-32. <http://doi.org/10.1016/j.copsyc.2015.07.011>
- Puts, D. A., Pope, L. E., Hill, A. K., Cárdenas, R. A., Welling, L. L., Wheatley, J. R., & Breedlove, S. M. (2015). Fulfilling desire: Evidence for negative feedback between men's testosterone, sociosexual psychology, and sexual partner number. *Hormones and Behavior*, 70, 14-21. <http://doi.org/10.1016/j.yhbeh.2015.01.006>
- R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org>
- Reimers, L., & Diekhof, E. K. (2015). Testosterone is associated with cooperation during intergroup competition by enhancing parochial altruism. *Frontiers in Neuroscience*, 9, 183. <http://doi.org/10.3389/fnins.2015.00183>
- Röcke, C., & Grühn, D. (2003). *German translation of the PANAS-X*. Unpublished manuscript, Free University Berlin, Germany.

- Ronay, R., & von Hippel, W. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Social Psychological and Personality Science*, 1, 57-64.
<http://doi.org/10.1177/1948550609352807>
- Roney, J. R. (2016). Theoretical frameworks for human behavioral endocrinology. *Hormones and Behavior*, 84, 97-110. <http://doi.org/10.1016/j.yhbeh.2016.06.004>
- Roney, J.R., Mahler, S.V., & Maestripieri, D. (2003). Behavioral and hormonal responses of men to brief interactions with women. *Evolution and Human Behavior*, 24, 365-375.
[http://doi.org/10.1016/S1090-5138\(03\)00053-9](http://doi.org/10.1016/S1090-5138(03)00053-9)
- Roney, J. R., Lukaszewski, A. W. & Simmons, Z. L. (2007). Rapid endocrine responses of young men to social interactions with young women. *Hormones and Behavior*, 52, 326-333.
<http://doi.org/10.1016/j.yhbeh.2007.05.008>
- Roney, J. R., Simmons, Z. L., & Lukaszewski, A. W. (2010). Androgen receptor gene sequence and basal cortisol concentrations predict men's hormonal responses to potential mates. *Proceedings of the Royal Society of London B: Biological Sciences*, 277, 57-63. <http://doi.org/10.1098/rspb.2009.1538>
- Rose, R. M., Bernstein, I. S., & Gordon, T. P. (1975). Consequences of social conflict on plasma testosterone levels in rhesus monkeys. *Psychosomatic Medicine*, 37, 50-61.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36. <http://doi.org/10.18637/jss.v048.i02>
- Salvador, A. (2005). Coping with competitive situations in humans. *Neuroscience & Biobehavioral Reviews*, 29, 195-205. <http://doi.org/10.1016/j.neubiorev.2004.07.004>
- Salvador, A. & Costa, R. (2009). Coping with competition: Neuroendocrine responses and cognitive variables. *Neuroscience and Biobehavioral Reviews*, 33, 160-170.
<http://doi.org/10.1016/j.neubiorev.2008.09.005>
- Sander, J., & Böcker, S. (1993). Die deutsche Form der Relationship Assessment Scale (RAS): Eine kurze Skala zur Messung der Zufriedenheit in einer Partnerschaft [The German version of the Relationship Assessment Scale (RAS): A short scale for measuring satisfaction in a dyadic relationship]. *Diagnostica*, 39, 55-62.
- Schultheiss, O. C., Schiepe, A., & Rawolle, M. (2012). Hormone assays. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *Handbook of Research Methods in Psychology (Vol. 1:*

Foundations, planning, measures, and psychometrics) (pp. 489-500). Washington DC: American Psychological Association.

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones, & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 17-44). New York, NY: Guilford Press.

Seery, M. D. (2011). Challenge or threat? Cardiovascular indexes of resilience and vulnerability to potential stress in humans. *Neuroscience & Biobehavioral Reviews*, *35*, 1603-1610.
<http://doi.org/10.1016/j.neubiorev.2011.03.003>

Seery, M. D., Weisbuch, M., & Blascovich, J. (2009). Something to gain, something to lose: The cardiovascular consequences of outcome framing. *International Journal of Psychophysiology*, *73*, 308-312.
<http://doi.org/10.1016/j.ijpsycho.2009.05.006>

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*, 575-584. <http://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*, 126-138.
<http://doi.org/10.1016/j.jrp.2006.02.004>

SFM Canada. (2013, October 4). Kanadas nachhaltige Wälder: Themen wie Holz, die biologische Vielfalt und den borealen Wald [video file]. Retrieved from <http://www.youtube.com/watch?v=qBQYWDJg3SM>

Sherman, G. D., Lerner, J. S., Josephs, R. A., Renshon, J., & Gross, J. J. (2016). The interaction of testosterone and cortisol is associated with attained status in male executives. *Journal of Personality and Social Psychology*, *110*, 921-929. <http://doi.org/10.1037/pspp0000063>

Simmons, Z. L., & Roney, J. R. (2011). Variation in CAG repeat length of the androgen receptor gene predicts variables associated with intrasexual competitiveness in human males. *Hormones and Behavior*, *60*, 306-312. <http://doi.org/10.1016/j.yhbeh.2011.06.006>

Simpson, J. A., Griskevicius, V., & Kim, J. S. (2011). Evolution, life history theory, and personality. In L. M. Horowitz & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions* (pp. 75-89). Hoboken, New Jersey: John Wiley & Sons.

Slatcher, R. B., Mehta, P. H., & Josephs, R. A. (2011). Testosterone and self-reported dominance interact to influence human mating behavior. *Social Psychological and Personality Science*, *2*, 531-539.
<http://doi.org/10.1177/1948550611400099>

- Smither, R. D., & Houston, J. M. (1992). The nature of competitiveness: The development and validation of the competitiveness index. *Educational and Psychological Measurement*, 52, 407-418.
<http://doi.org/10.1177/0013164492052002016>
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory STAI*. Palo Alto, CA: Consulting Psychologists Press, Inc.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88, 500-517. <http://doi.org/10.1037/0021-9010.88.3.500>
- Tett, R. P., Simonet, D. V., Walser, B., & Brown, C. (2013). Trait activation theory. In Christiansen, N.D.; Tett, R.P. *Handbook of personality at work* (pp. 71-100). New York, NY: Routledge.
- Tinbergen, N. (1963). On aims and methods of ethology. *Ethology*, 20, 410-433.
<http://doi.org/10.1111/j.1439-0310.1963.tb01161.x>
- Torrance, J. S., Hahn, A. C., Kandrik, M., DeBruine, L. M., & Jones, B. C. (2018). No evidence for associations between men's salivary testosterone and responses on the Intrasexual Competitiveness Scale. *Adaptive Human Behavior and Physiology*, 4, 321-327. <http://doi.org/10.1007/s40750-018-0095-2>
- Trumble, B. C., Cummings, D., von Rueden, C., O'Connor, K. A., Smith, E. A., Gurven, M., & Kaplan, H. (2012). Physical competition increases testosterone among Amazonian forager-horticulturalists: A test of the 'challenge hypothesis'. *Proceedings of the Royal Society B: Biological Sciences*, 279, 2907-2912.
<http://doi.org/10.1098/rspb.2012.0455>
- 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. <http://doi.org/10.1016/j.jrp.2014.03.008>
- van Anders, S. M., Goldey, K. L., & Kuo, P. X. (2011). The Steroid/Peptide Theory of Social Bonds: Integrating testosterone and peptide responses for classifying social behavioral contexts. *Psychoneuroendocrinology*, 36, 1265-1275. <http://doi.org/10.1016/j.psyneuen.2011.06.001>
- van Anders, S. M., Steiger, J., & Goldey, K. L. (2015). Effects of gendered behavior on testosterone in women and men. *Proceedings of the National Academy of Sciences*, 112, 13805-13810.
<http://doi.org/10.1037/pnas.1509591112>
- van Anders, S. M., & Watson, N. V. (2006). Social neuroendocrinology. *Human Nature*, 17, 212-237.
<http://doi.org/10.1007/s12110-006-1018-7>

- van der Meij, L., Almela, M., Buunk, A. P., Fawcett, T. W., & Salvador, A. (2012). Men with elevated testosterone levels show more affiliative behaviours during interactions with women. *Proceedings of the Royal Society B: Biological Sciences*, 279, 202-208. <http://doi.org/10.1098/rspb.2011.0764>
- van der Meij, L., Buunk, A. P., Almela, M., & Salvador, A. (2010). Testosterone responses to competition: The opponent's psychological state makes it challenging. *Biological Psychology*, 84, 330-335. <http://doi.org/10.1016/j.biopsycho.2010.03.017>
- van der Meij, L., Buunk, A. P., van de Sande, J. P., & Salvador, A. (2008). The presence of a woman increases testosterone in aggressive dominant men. *Hormones and Behavior*, 54, 640-644. <http://doi.org/10.1016/j.yhbeh.2008.07.001>
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, 40, 472-481. <http://doi.org/10.1016/j.jrp.2005.03.003>
- Vermeer, A. L., Riečanský, I., & Eisenegger, C. (2016). Competition, testosterone, and adult neurobehavioral plasticity. *Progress in Brain Research*, 229, 213-238. <http://doi.org/10.1016/bs.pbr.2016.05.004>
- Vermeulen, A., Goemaere, S., & Kaufman, J. M. (1999). Testosterone, body composition and aging. *Journal of Endocrinological Investigation*, 22, 110-116.
- Walker, F. R., Pflingst, K., Carnevali, L., Sgoifo, A., & Nalivaiko, E. (2017). In the search for integrative biomarker of resilience to psychological stress. *Neuroscience & Biobehavioral Reviews*, 74, 310-320. <http://doi.org/10.1016/j.neubiorev.2016.05.003>
- Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. Unpublished manuscript. University of Iowa, Iowa City.
- Welker, K. M., & Carré, J. M. (2015). Individual differences in testosterone predict persistence in men. *European Journal of Personality*, 29, 83-89. <http://doi.org/10.1002/per.1958>
- Welker, K. M., Lassetter, B., Brandes, C. M., Prasad, S., Koop, D. R., & Mehta, P. H. (2016). A comparison of salivary testosterone measurement using immunoassays and tandem mass spectrometry. *Psychoneuroendocrinology*, 71, 180-188. <http://doi.org/10.1016/j.psyneuen.2016.05.022>
- 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. <http://doi.org/10.1016/j.psyneuen.2015.11.016>

- Wiggins, J. S. (1982). Circumplex models of interpersonal behavior in clinical psychology. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of Research Methods in Clinical Psychology* (pp. 183-221). New York: Wiley.
- Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research*, 23, 517-530.
http://doi.org/10.1207/s15327906mbr2304_8
- Wingfield, J. C. (2017). The challenge hypothesis: Where it began and relevance to humans. *Hormones and Behavior*, 92, 9-12. <http://doi.org/10.1016/j.yhbeh.2016.11.008>
- 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. *American Naturalist*, 136, 829-846. <http://doi.org/10.1086/285134>
- Zilioli, S., & Bird, B. M. (2017). Functional significance of men's testosterone reactivity to social stimuli. *Frontiers in Neuroendocrinology*, 47, 1-18. <http://doi.org/10.1016/j.yfrne.2017.06.002>

Online supplementary

Tables for results of main analyses

Table S1

Descriptive statistics for all main variables measured in Study 1

Experimental/control	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	Reliability
Age	24.1/24.9	3.3/2.9	18/21	34/32	-
BMI	24.08/23.68	4.02/3.10	15.87/18.94	41.58/34.82	-
Relationship status	59/21 single, 10/1 open, 50/16 committed relationship, 4/0 engaged, 2/2 married				
Baseline C	2.98/2.84	1.85/1.46	0.55/0.26	10.42/6.20	-
Baseline T	78.47/73.78	30.93/30.45	24.10/28.80	180.37/171.50	-
1 st <i>post</i> T measure	88.63/75.78	33.63/33.96	30.40/30.70	193.38/223.14	-
2 nd <i>post</i> T measure	93.99/83.50	37.94/35.09	28.20/40.80	216.47/230.51	-
PA pre	3.34/3.35	0.57/0.47	1.80/2.40	4.80/4.40	.77/.65
HI pre	2.23/2.00	0.61/0.44	1.00/1.00	3.80/2.80	.72/.53
BC pre	2.36/2.28	0.70/0.60	1.00/1.00	4.20/3.60	.69/.61
JK pre	2.70/2.42	0.67/0.67	1.40/1.20	1.60/4.00	.72/.75
DE pre	1.85/1.64	0.70/0.58	1.00/1.00	4.20/3.20	.79/.76
LM pre	3.40/3.22	0.61/0.65	1.60/1.80	5.00/4.80	.77/.80
FG pre	2.42/2.27	0.67/0.57	1.00/1.00	4.20/3.80	.75/.79
NO pre	3.49/3.37	0.68/0.74	1.60/1.80	5.00/4.60	.85/.89
PA post	3.46/3.47	0.59/0.49	2.00/2.60	5.00/4.60	.78/.68
HI post	2.06/1.95	0.60/0.39	1.00/1.20	3.80/2.80	.72/.45
BC post	2.37/2.09	0.65/0.63	1.00/1.00	4.00/3.20	.60/.65
JK post	2.51/2.51	0.66/0.63	1.00/1.00	4.00/4.20	.71/.67
DE post	1.86/1.54	0.67/0.49	1.00/1.00	4.00/2.60	.79/.75
LM post	3.43/3.35	0.63/0.75	1.80/1.20	5.00/4.80	.80/.86
FG post	2.17/2.11	0.64/0.55	1.00/1.00	4.00/3.20	.82/.65
NO post	3.75/3.56	0.61/0.68	2.20/2.00	5.00/4.80	.83/.65
Stress pre	2.36/2.25	0.70/0.58	1.00/1.17	4.00/3.50	.80/.72
Positive affect pre	3.26/3.13	0.54/0.50	1.80/2.30	4.70/4.20	.83/.80
Negative affect pre	1.94/1.84	0.63/0.49	1.00/1.00	3.67/3.00	.85/.75
Self-esteem pre	3.81/3.98	0.85/0.69	1.50/1.75	5.00/5.00	.85/.82
Stress post	2.57/2.21	0.65/0.55	1.00/1.17	4.00/3.50	.76/.72
Positive affect post	3.61/3.33	0.52/0.62	2.40/2.10	4.90/4.50	.83/.89
Negative affect post	1.91/1.76	0.62/0.47	1.00/1.00	3.44/3.11	.84/.81
Self-esteem post	4.05/4.18	0.74/0.66	1.50/2.00	5.00/5.00	.86/.85

Notes. Descriptive statistics separately for the experimental ($n=124$ -125)/control group ($n=40$).

Reliabilities: Cronbach's α for *pre* and *post* personality states. Unit for Testosterone values: pg/ml; for Cortisol values: nmol/l.

Table S2

Bivariate Pearson correlations between self-reported personality state changes and testosterone reactivity in Study 1

	Baseline C	Baseline T	1 st T reactivity	2 nd T reactivity	PA change	HI change	BC change	JK change	DE change	LM change	FG change	NO change
Baseline C	-	.11	.12	.07	-.04	-.28	-.08	.03	-.05	-.07	-.11	.14
Baseline T	.29***	-	.07	-.03	.28	-.13	.18	.08	.04	.25	.01	.26
1 st T reactivity	.06	-.04	-	.62***	.26	-.28	.30	-.27	-.12	-.01	-.25	.28
2 nd T reactivity	-.14	.00	.56***	-	-.09	-.08	.13	-.21	.03	-.22	-.10	-.13
PA change	-.13	-.02	.09	.11	-	-.49**	.19	-.30	.01	.11	-.55***	.54***
HI change	.11	.10	.02	.15	-.47***	-	.00	.36*	.25	-.07	.42**	-.42**
BC change	.03	-.06	.15	-.02	.21*	.01	-	-.19	.30	.20	-.09	-.07
JK change	.05	.07	-.12	-.11	-.30**	.43***	-.06	-	-.07	-.07	.23	.00
DE change	.06	-.08	.12	-.01	.08	.02	.46***	-.05	-	-.21	.10	.04
LM change	.04	.13	-.07	-.04	.13	.05	-.11	.18*	-.33***	-	-.26	.14
FG change	.25**	.04	.05	.01	-.32***	.47***	.11	.22*	.28**	-.13	-	-.46**
NO change	-.17	-.02	.01	.13	.31***	-.15	-.01	-.07	-.25**	.31***	-.51***	-

Note. Baseline C/T = 2nd baseline cortisol/testosterone measure (main session, before competition);

1st/2nd T reactivity = residuals from 1st/2nd *post*-sample regressed on *pre*-sample; separately for the experimental (bottom-left; $n=124-125$) and the control group (top-right; $n=40$); * $p<.05$, ** $p<.01$, *** $p<.001$.

Table 2a

T reactivity (Hypothesis 1), experimental versus control group (effects of condition), including all eight control variables

IV: condition	β	SE	t	p	Partial η^2
1 st T reactivity	.39	.19	2.08	.02 ^p	.03
2 nd T reactivity	.14	.19	0.73	.23 ^p	.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd *post*-sample regressed on *pre*-sample; ^p = one-tailed p -value due to preregistered hypothesis.

Table S3

Self-reported pre-post personality changes in Study 1 (Hypothesis 2)

IV: condition	β	SE	t	p	Partial η^2
Δ assured-dominant (PA)	-.01	.18	-0.06	.48 ^p	0.00
Δ competitive (BC)	.46	.18	2.58	<.001 ^p	0.04
Δ coldhearted (DE)	.40	.18	-2.23	.03	0.03
Δ introverted (FG)	-.07	.18	-0.37	.35 ^p	0.00
Δ submissive (HI)	-.11	.18	-0.58	.56	0.00
Δ ingenuous (JK)	-.44	.18	-2.46	.02	0.04
Δ nurturing (LM)	-.17	.18	-0.94	.18 ^p	0.01

Δ extraverted (NO)	.26	.18	1.44	.08 ^p	0.01
---------------------------	-----	-----	------	------------------	------

Note: IV = independent variable; condition: 0=control, 1=experimental group; β =standardized regression coefficient; SE=standard error; *Partial η^2* =partial eta-squared effect sizes; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S4

Self-reported pre-post personality changes in Study 1 (Hypothesis 2), including all eight control variables

IV: condition	β	SE	<i>t</i>	<i>p</i>	<i>Partial η^2</i>
Δ assured-dominant (PA)	-.25	.12	-1.99	.02 ^p	0.03
Δ competitive (BC)	.26	.16	1.60	.055 ^p	0.02
Δ coldhearted (DE)	.28	.17	1.68	.09	0.02
Δ introverted (FG)	.16	.16	0.99	.16 ^p	0.01
Δ submissive (HI)	-.11	.17	-0.65	.52	0.00
Δ ingenuous (JK)	-.37	.19	-2.01	.046	0.03
Δ nurturing (LM)	-.22	.18	-1.21	.12	0.01
Δ extraverted (NO)	.02	.15	0.15	.44 ^p	0.00

Note: IV = independent variable; condition: 0=control, 1=experimental group; β =standardized regression coefficient; SE=standard error; *Partial η^2* =partial eta-squared effect sizes; control variables: age, BMI, relationship status (binary), sexual partners within previous 1 month (binary yes/no), changes in stress (STAI), self-esteem (RSES), positive and negative affect (PANAS); ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S5

Results from linear models predicting personality state changes from T reactivity in Study 1 (Hypothesis 3), including all eight control variables

IAL state changes	1 st T reactivity (β)	SE	<i>P</i>	<i>Partial η^2</i>	2 nd T reactivity (β)	SE	<i>p</i>	<i>Partial η^2</i>
Δ assured-dominant (PA)	.08	.06	.11 ^p	0.01	.06	.06	.18 ^p	0.01
Δ competitive (BC)	.13	.08	.04 ^p	0.03	-.05	.08	.55	0.00
Δ coldhearted (DE)	.05	.08	.52	0.00	-.06	.08	.46	0.00
Δ introverted (FG)	.08	.08	.34	0.01	.06	.08	.48	0.00
Δ submissive (HI)	.01	.08	.91	0.00	.17	.08	.04	0.04
Δ ingenuous (JK)	-.12	.09	.20	0.01	-.09	.09	.32	0.01
Δ nurturing (LM)	-.04	.09	.63 ^p	0.00	-.03	.09	.37 ^p	0.00
Δ extraverted (NO)	-.03	.08	.67	0.00	.05	.08	.53	0.00

Note. IAL = interpersonal adjective list; SE = standard error; *partial η^2* = partial eta-squared effect size;

^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S6

Self-reported personality state changes and interaction between testosterone reactivity and baseline cortisol in Study 1 (Hypothesis 4)

	for 1 st T reactivity				for 2 nd T reactivity			
	TxC (β)	SE	<i>t</i>	Partial η^2	TxC (β)	SE	<i>t</i>	Partial η^2
Δ assured-dominant (PA)	-.35	.15	-2.30* ^p	0.04	-.26	.15	-1.76* ^p	0.03
Δ competitive (BC)	-.43	.15	-2.82** ^p	0.06	-.36	.15	-2.40** ^p	0.05
Δ coldhearted (DE)	-.24	.16	-1.55	0.02	-.16	.15	-1.06	0.01
Δ introverted (FG)	.24	.15	1.59	0.02	.21	.15	1.44	0.02
Δ submissive (HI)	.24	.16	1.53	0.02	.07	.15	0.45	0.00
Δ ingenuous (JK)	.23	.16	1.48	0.02	.18	.15	1.19	0.01
Δ nurturing (LM)	-.01	.16	-0.07	0.00	-.16	.15	-1.08	0.01
Δ extraverted (NO)	-.16	.16	-0.99	0.01	-.14	.15	-0.92	0.01

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; C=baseline cortisol (2nd sample); ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S7

Self-reported personality state changes and interaction between testosterone reactivity and baseline cortisol in Study 1 (Hypothesis 4), including all eight control variables

	TxC for 1 st T reactivity				TxC for 1 st T reactivity			
	<i>B</i>	SE	<i>t</i>	Partial η^2	β	SE	<i>t</i>	Partial η^2
Δ assured-dominant (PA)	-.07	.11	-0.60	0.00	-.05	.10	-0.47	0.00
Δ competitive (BC)	-.28	.13	-2.12* ^p	0.04	-.25	.13	-1.94* ^p	0.03
Δ coldhearted (DE)	-.19	.15	-1.28	0.01	-.16	.14	-1.19	0.01
Δ introverted (FG)	.11	.14	0.81	0.01	.08	.13	0.60	0.00
Δ submissive (HI)	.17	.15	1.14	0.01	-.03	.14	-0.22	0.00
Δ ingenuous (JK)	.16	.16	0.96	0.01	.10	.15	0.68	0.00
Δ nurturing (LM)	.08	.16	0.52	0.00	-.09	.15	-0.61	0.00
Δ extraverted (NO)	.12	.14	0.88	0.01	.08	.13	0.66	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; C=baseline cortisol (2nd sample); ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S8

Descriptive statistics and internal consistencies in Study 2

Experimental/control	<i>M</i>	<i>SD</i>	Reliability
PA pre	3.19/3.27	0.63/0.65	.93/.94
BC pre	2.12/2.26	0.44/0.46	.82/.78
DE pre	1.78/1.86	0.32/0.33	.80/.78

FG pre	2.45/2.40	0.48/0.46	.85/.82
HI pre	2.66/2.54	0.64/0.61	.94/.93
JK pre	3.01/2.88	0.43/0.39	.81/.74
LM pre	3.48/3.67	0.45/0.47	.89/.88
NO pre	3.47/3.50	0.53/0.55	.90/.93
PA post	3.35/3.20	0.67/0.67	.94/.93
BC post	2.29/2.20	0.55/0.45	.88/.75
DE post	1.85/1.86	0.38/0.35	.84/.84
FG post	2.34/2.43	0.47/0.49	.84/.84
HI post	2.51/2.62	0.63/0.59	.94/.91
JK post	2.87/2.93	0.48/0.37	.85/.73
LM post	3.41/3.33	0.48/0.48	.90/.90
NO post	3.52/3.40	0.54/0.55	.91/.90

Note: PA=dominance, BC=competitiveness, DE=coldheartedness, FG=introversion,

HI=submissiveness, LM=nurturance, NO=extraversion; reliability = Cronbach's α internal consistency.

Table S9

Observer-rated pre-post changes in experimental relative to control group in Study 2 (Hypothesis 2)

	β	SE	t	P	Partial η^2
Δ assured-dominant (PA)	.49	.18	2.69	<.01 ^p	0.04
Δ competitive (BC)	.37	.18	1.99	.02 ^p	0.02
Δ coldhearted (DE)	.08	.19	0.44	.66	0.00
Δ introverted (FG)	-.35	.18	-1.92	.057	0.02
Δ submissive (HI)	-.55	.18	-3.02	<.01 ^p	0.06
Δ ingenuous (JK)	-.42	.18	-2.29	.01 ^p	0.03
Δ nurturing (LM)	.06	.19	0.35	.73	0.00
Δ extraverted (NO)	.34	.18	1.86	.06	0.02

Note: Partial η^2 =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S10

Observer-rated pre-post changes in experimental relative to control group in Study 2 (Hypothesis 2), including all eight control variables

	β	SE	t	p	Partial η^2
Δ assured-dominant (PA)	.50	.19	2.66	<.01 ^p	0.04
Δ competitive (BC)	.37	.19	1.97	.03 ^p	0.02
Δ coldhearted (DE)	.10	.19	0.51	.61	0.00
Δ introverted (FG)	-.33	.19	-1.77	.08	0.02
Δ submissive (HI)	-.53	.19	-2.87	<.02 ^p	0.05
Δ ingenuous (JK)	-.40	.19	-2.15	.02 ^p	0.03

Δ nurturing (LM)	.05	.19	0.26	.79	0.00
Δ extraverted (NO)	.32	.19	1.70	.09	0.02

Note: *Partial η^2* =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S11

Associations between testosterone reactivity and observer-rated pre-post changes in the experimental group in Study 2 (Hypothesis 3), including the control variables age, relationship status and sexual identity

IAL state changes	1 st T reactivity (β)	SE	<i>t</i>	<i>Partial η^2</i>	2 nd T reactivity (β)	SE	<i>t</i>	<i>Partial η^2</i>
Δ assured-dominant (PA)	.14	.09	1.45	0.02	.14	.09	1.53	0.02
Δ competitive (BC)	.07	.09	0.71	0.00	.10	.09	1.08	0.01
Δ coldhearted (DE)	.06	.10	0.64	0.00	.07	.09	0.76	0.01
Δ introverted (FG)	-.04	.09	-0.44	0.00	.01	.09	0.14	0.00
Δ submissive (HI)	-.04	.09	-0.44	0.00	.00	.09	-0.03	0.00
Δ ingenuous (JK)	-.04	.09	-0.42	0.00	-.04	.09	-0.45	0.00
Δ nurturing (LM)	-.03	.09	-0.31	0.00	.03	.09	0.33	0.00
Δ extraverted (NO)	.10	.09	1.02	0.01	.08	.09	0.89	0.01

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; *Partial η^2* =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S12

Associations between observer-rated pre-post changes and the interaction between testosterone reactivity and baseline cortisol in the experimental group in Study 2 (Hypothesis 4)

	TxC for 1 st T reactivity				TxC for 1 st T reactivity			
	β	SE	<i>t</i>	<i>Partial η^2</i>	β	SE	<i>t</i>	<i>Partial η^2</i>
Δ assured-dominant (PA)	.17	.16	1.07	0.01	.05	.15	0.32	0.00
Δ competitive (BC)	.01	.16	0.06	0.00	.06	.15	0.40	0.00
Δ coldhearted (DE)	.01	.16	0.09	0.00	.10	.15	0.67	0.00
Δ introverted (FG)	-.13	.16	-0.79	0.01	.19	.15	1.20	0.01
Δ submissive (HI)	-.19	.16	-1.14	0.01	-.01	.15	-0.03	0.00
Δ ingenuous (JK)	-.14	.16	-0.88	0.01	-.04	.15	-0.24	0.00
Δ nurturing (LM)	.07	.16	0.41	0.00	-.06	.15	-0.40	0.00
Δ extraverted (NO)	.10	.16	0.58	0.00	-.12	.15	-0.76	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; C=baseline cortisol (2nd sample); ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S13

Associations between observer-rated pre-post changes and the interaction between testosterone reactivity and baseline cortisol in the experimental group in Study 2 (Hypothesis 4), including the control variables age, relationship status and sexual identity

	TxC for 1 st T reactivity				TxC for 1 st T reactivity			
	β	SE	<i>t</i>	Partial η^2	β	SE	<i>t</i>	Partial η^2
Δ assured-dominant (PA)	.17	.16	1.05	0.01	.05	.16	0.32	0.00
Δ competitive (BC)	.03	.17	0.02	0.00	.05	.16	0.32	0.00
Δ coldhearted (DE)	-.01	.16	-0.06	0.00	.05	.15	0.35	0.00
Δ introverted (FG)	-.15	.17	-0.90	0.01	.15	.16	0.93	0.01
Δ submissive (HI)	-.19	.16	-0.66	0.01	-.09	.16	-0.05	0.00
Δ ingenuous (JK)	-.14	.16	-0.89	0.01	-.04	.15	-0.28	0.00
Δ nurturing (LM)	.08	.16	0.46	0.00	-.03	.16	-0.19	0.00
Δ extraverted (NO)	.11	.16	0.65	0.00	-.08	.16	-0.53	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; C=baseline cortisol (2nd sample); ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S14

Descriptive statistics and internal consistencies in Study 3

Experimental/control group	<i>M</i>	<i>SD</i>	Reliability
Self-assured pre	3.16/3.14	0.69/0.68	.97/.97
Cooperative pre	3.40/3.30	0.40/0.39	.85/.78
Self-displaying pre	2.78/2.89	0.63/0.64	.90/.94
Self-assured post	3.30/3.13	0.69/0.75	.97/.97
Cooperative post	3.34/3.29	0.44/0.45	.86/.92
Self-displaying post	2.94/2.85	0.68/0.68	.93/.92

Note: Reliability = Cronbach's α internal consistency.

Table S15

Observer-rated pre-post social impression changes in experimental relative to control group in Study 3 (Hypothesis 2)

	β	SE	<i>t</i>	<i>p</i>	Partial η^2
Δ self-assurance	.37	.18	1.99	.02 ^p	0.02
Δ cooperativeness	.02	.19	-0.01	.99	0.00
Δ self-display	.31	.19	1.69	.046 ^p	0.02

Note: Partial η^2 = partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S16

Observer-rated pre-post social impression changes in experimental relative to control group in Study 3 (Hypothesis 2), including the three control variables age, relationship status and sexual identity

	β	SE	<i>t</i>	<i>p</i>	Partial η^2
Δ self-assurance	.36	.19	1.92	.03 ^p	0.02
Δ cooperativeness	-.02	.19	-0.12	.90	0.00
Δ self-display	.30	.19	1.59	.06 ^p	0.02

Note: Partial η^2 = partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S17

Associations between testosterone reactivity and observer-rated pre-post social impression changes in the experimental group in Study 3 (Hypothesis 3)

	1 st T reactivity (β)	SE	<i>t</i>	Partial η^2	2 nd T reactivity (β)	SE	<i>t</i>	Partial η^2
Δ self-assurance	.20	.09	2.26* ^p	0.04	.14	.09	1.49	0.02
Δ cooperativeness	-.02	.09	-0.25	0.00	.04	.09	0.43	0.00
Δ self-display	.14	.09	1.52	0.02	.07	.09	0.75	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; Partial η^2 =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S18

Associations between testosterone reactivity and observer-rated pre-post social impression changes in the experimental group in Study 3 (Hypothesis 3), including the three control variables age, relationship status and sexual identity

	1 st T reactivity (β)	SE	<i>t</i>	Partial η^2	2 nd T reactivity (β)	SE	<i>t</i>	Partial η^2
Δ self-assurance	.21	.09	2.29* ^p	0.04	.13	.09	1.43	0.02
Δ cooperativeness	-.02	.09	-0.19	0.00	.03	.09	0.36	0.00
Δ self-display	.14	.09	1.52	0.02	.07	.09	0.72	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; Partial η^2 =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S19

Associations between observer-rated pre-post social impression changes and the interaction between testosterone reactivity and baseline cortisol in the experimental group in Study 3 (Hypothesis 4)

	TxC for 1 st T reactivity				TxC for 2 nd T reactivity			
	β	SE	t	Partial η^2	β	SE	T	Partial η^2
Δ self-assurance	.31	.16	1.93	0.03	-.12	.15	-0.79	0.01
Δ cooperativeness	-.01	.16	-0.06	0.00	-.12	.15	-0.75	0.00
Δ self-display	.21	.16	1.33	0.02	-.06	.15	-0.38	0.01

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; C = baseline cortisol (2nd sample); * p <.05, ** p <.01, *** p <.001.

Table S20

Associations between observer-rated pre-post social impression changes and the interaction between testosterone reactivity and baseline cortisol in the experimental group in Study 3 (Hypothesis 4), including the three control variables age, relationship status and sexual identity

	TxC for 1 st T reactivity				TxC for 2 nd T reactivity			
	β	SE	t	Partial η^2	β	SE	T	Partial η^2
Δ self-assurance	.31	.16	1.96	0.03	-.10	.16	-0.67	0.00
Δ cooperativeness	.01	.17	0.04	0.00	-.08	.16	-0.50	0.00
Δ self-display	.22	.16	1.33	0.02	-.05	.16	-0.32	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; C = baseline cortisol (2nd sample); * p <.05, ** p <.01, *** p <.001.

Tables for additional preregistered analyses Study 1 (osf.io/8n7ev)

Table S21

Self-reported personality state changes (Hypothesis 2), experimental versus control group (effects of Condition)

	β	SE	P
Extraversion (FFM)	.47	.18	<.01 ^p
Attractiveness	-.08	.18	.66
Self-esteem (RSES)	-.05	.18	.77
Masculinity	.00	.18	.98
Femininity	-.12	.18	.26 ^p
Narcissistic admiration (NARQ)	-.11	.18	.54
Narcissistic rivalry (NARQ)	.12	.18	.26 ^p

Note: FFM = Five Factor Model; RSES = Rosenberg Self-Esteem Scale; NARQ = Narcissistic Admiration and Rivalry Questionnaire; ^p = one-tailed p -value due to preregistered hypothesis.

Table S22

Associations between self-reported personality state changes and T reactivity (Hypothesis 3), experimental group only

	1 st T reactivity			2 nd T reactivity		
	β	SE	p	β	SE	P

Extraversion (FFM)	.08	.09	.36	.00	.09	.45 ^p
Attractiveness	.14	.09	.12	.03	.09	.74
Self-esteem (RSES)	.02	.09	.80	.00	.09	.48
Masculinity	.08	.09	.36	.12	.09	.18
Femininity	-.09	.09	.32	.02	.09	.85
Narcissistic admiration (NARQ)	.16	.09	.08	.16	.09	.08
Narcissistic rivalry (NARQ)	.16	.09	.08	.08	.09	.40

Note: FFM = Five Factor Model; RSES = Rosenberg Self-Esteem Scale; NARQ = Narcissistic Admiration and Rivalry Questionnaire; 1st/2nd T reactivity = 1st/2nd testosterone *post*-sample; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S23

Associations between self-reported personality state changes and the interaction of T reactivity and baseline C (Hypothesis 4), experimental group only

	1 st T reactivity x baseline C			2 nd T reactivity x baseline C		
	<i>β</i>	<i>SE</i>	<i>p</i>	<i>β</i>	<i>SE</i>	<i>p</i>
Extraversion (FFM)	-.01	.16	.93	-.07	.15	.63
Attractiveness	.06	.16	.72	.18	.15	.24
Self-esteem (RSES)	-.05	.16	.74	.03	.15	.87
Masculinity	-.19	.15	.22	-.17	.15	.24
Femininity	-.02	.16	.92	.00	.15	.98
Narcissistic admiration (NARQ)	.16	.16	.32	.18	.15	.23
Narcissistic rivalry (NARQ)	.05	.16	.74	.10	.15	.49

Note: FFM = Five Factor Model; RSES = Rosenberg Self-Esteem Scale; NARQ = Narcissistic Admiration and Rivalry Questionnaire; 1st/2nd T reactivity = residuals from 1st/2nd *post*-sample regressed on *pre*-sample; baseline C = 2nd baseline cortisol sample; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S24

C reactivity (Hypothesis 5), experimental versus control group (effects of Condition)

	<i>β</i>	<i>SE</i>	<i>p</i>
1 st C reactivity	.10	.18	.28 ^p
2 nd C reactivity	.94	.17	<.001 ^p

Note: 1st/2nd C reactivity = 1st/2nd cortisol *post*-sample; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S25

T reactivity predicted by baseline C (Hypothesis 6), experimental group only

	<i>β</i>	<i>SE</i>	<i>P</i>
1 st T reactivity	.06	.09	.53
2 nd T reactivity	-.14	.09	.07 ^p

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S26

Baseline T, T reactivity and trait-dominance, also moderated by baseline C (Hypothesis 7), experimental group only

DV: trait-dominance	β	SE	<i>p</i>
Baseline T1	-.07	.09	.44
Baseline T2	.02	.09	.40 ^p
Baseline T12	-.04	.09	.66
Baseline T1*baseline C1	.24	1.28	.85
Baseline T2*baseline C2	.64	.94	.50
Baseline T12*baseline C12	1.96	1.16	.10
1 st T reactivity	.02	.09	.40 ^p
2 nd T reactivity	-.01	.09	.40
1 st T reactivity*baseline C2	.40	.16	.01
2 nd T reactivity*baseline C2	.50	.14	<.001

Note: Trait-dominance measured by the interpersonal adjective list; Baseline T1/T2/T12 =

1st/2nd/mean of 1st and 2nd baseline testosterone measure; 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S27

Self-rated performance in the competition and T reactivity (Further predictions, Hypothesis 1), experimental group only

DV: self-rated performance	β	SE	<i>p</i>
1 st T reactivity	.03	.09	.39 ^p
2 nd T reactivity	-.05	.09	.61

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; ^p = one-tailed *p*-value due to preregistered hypothesis, including baseline T as a covariate.

Table S28

Baseline T, trait-dominance and self-reported dominant/competitive behaviour during the competition (Further predictions, Hypothesis 2), experimental group only

DV: dominant behaviour	β	SE	<i>p</i>
Trait-dominance*baseline T1	1.36	.99	.17
Trait-dominance*baseline T2	.61	1.03	.56
Trait-dominance*baseline T12	1.16	1.18	.33
DV: competitive behaviour			
Trait-dominance*baseline T1	-.42	1.02	.69
Trait-dominance*baseline T2	.52	1.10	.64
Trait-dominance*baseline T12	-.26	1.25	.83

Note: Trait-dominance measured by the interpersonal adjective list; Baseline T1/T2/T12 = 1st/2nd/mean of 1st and 2nd baseline testosterone measure; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S29

Baseline T, self-reported trait-dominance and opponent's dominant/competitive behaviour during the competition (Further predictions, Hypothesis 3), experimental group only

DV: dominant behaviour	<i>β</i>	<i>SE</i>	<i>P</i>
Trait-dominance*baseline T1	-.49	1.03	.64
Trait-dominance*baseline T2	-1.66	1.08	.13
Trait-dominance*baseline T12	-1.46	1.25	.24
DV: competitive behaviour			
Trait-dominance*baseline T1	.74	1.03	.48
Trait-dominance*baseline T2	.05	1.10	.96
Trait-dominance*baseline T12	.69	1.26	.59

Note: Trait-dominance measured by the interpersonal adjective list; Baseline T1/T2/T12 = 1st/2nd/mean of 1st and 2nd baseline testosterone measure; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S30

Association between T reactivity and self-reported change in competitiveness moderated by trait anxiety (BFI, neuroticism) (Further predictions, Hypothesis 4), experimental group only

	<i>β</i>	<i>SE</i>	<i>p</i>
1 st T reactivity	-.65	.35	.03 ^p
2 nd T reactivity	-.56	.35	.06 ^p

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; BFI = big five inventory; including baseline T as a covariate; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S31

Trait-neuroticism (BFI) and baseline C (Further predictions, Hypothesis 5a), both experimental and control group

	<i>β</i>	<i>SE</i>	<i>p</i>
baseline C1	-.01	.08	.93
baseline C2	.03	.08	.37 ^p
baseline C12	.01	.08	.44 ^p

Note: baseline C1/C2/C12 = 1st/2nd/mean of 1st and 2nd baseline cortisol measure; BFI = big five inventory; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S32

Trait-neuroticism (BFI) and C reactivity (Further predictions, Hypothesis 5b), experimental group only

	<i>β</i>	<i>SE</i>	<i>p</i>
1 st C reactivity	-.04	.09	.32 ^p
2 nd C reactivity	.01	.09	.92

Note: 1st/2nd C reactivity = 1st/2nd cortisol *post*-sample; BFI = big five inventory; including baseline C as a covariate; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S33

C reactivity predicted by interaction between baseline T and competition outcome (Further predictions, Hypothesis 6), experimental group only

	<i>β</i>	<i>SE</i>	<i>P</i>
1 st C reactivity	-.02	.19	.94
2 nd C reactivity	.20	.18	.28

Note: 0/1 = lost/won competition; 1st/2nd C reactivity = 1st/2nd cortisol *post*-sample; BFI = big five inventory; baseline T = 2nd baseline testosterone measure; including baseline C as a covariate; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S34

T reactivity in winners predicted by interaction between baseline T and baseline C (Further predictions, Hypothesis 7), experimental group only

	<i>β</i>	<i>SE</i>	<i>p</i>
1 st T reactivity	3.65	1.03	<.001
2 nd T reactivity	3.22	1.03	<.01

Note: 1st/2nd T reactivity = residuals from 1st/2nd *post*-sample regressed on *pre*-sample; baseline T/C = 2nd baseline testosterone/cortisol measure; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S35

Baseline T in interaction with baseline C predicting competitiveness/dominance/aggressiveness during the competition (rated by competitor) (Further predictions, Hypothesis 8a), experimental group only

	<i>β</i>	<i>SE</i>	<i>p</i>
Competitiveness	-.32	.94	.74
Dominance	.01	.94	.99
Aggressiveness	.03	.94	.97

Note: baseline T2/C2 = 2nd baseline cortisol/testosterone measure; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S36

T reactivity in interaction with baseline C predicting competitiveness/dominance/aggressiveness during the competition (rated by competitor) (Further predictions, Hypothesis 8b), experimental group only

	1 st T reactivity			2 nd T reactivity		
	<i>β</i>	<i>SE</i>	<i>p</i>	<i>β</i>	<i>SE</i>	<i>P</i>
Competitiveness	.06	.13	.66	.10	.13	.46
Dominance	.08	.13	.57	.03	.12	.80
Aggressiveness	.15	.12	.21	.15	.12	.20

Note: 1st/2nd T reactivity = residuals from 1st/2nd *post*-sample regressed on *pre*-sample; including baseline T as a covariate; baseline T/C = 2nd baseline cortisol/testosterone measure; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S36

T reactivity and willingness to compete again (vs. engaging in a cooperative task) after the competition (Further predictions, Hypothesis 9a), experimental group only

	<i>β</i>	<i>SE</i>	<i>p</i>
1 st T reactivity	-.02	.05	.33 ^p
2 nd T reactivity	-.09	.04	.03 ^p

Note: 1/2 = compete again/cooperate; 1st/2nd T reactivity = residuals from 1st/2nd *post*-sample regressed on *pre*-sample; including baseline testosterone (2nd measure) as a covariate; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S37

T reactivity and willingness to compete again (vs. engaging in a cooperative task) after the competition (Further predictions, Hypothesis 9b), losers experimental group only

	<i>B</i>	<i>SE</i>	<i>p</i>
1 st T reactivity	.01	.06	.42 ^p
2 nd T reactivity	-.03	.06	.70

Note: 1/2 = compete again/cooperate; 1st/2nd T reactivity = residuals from 1st/2nd *post*-sample regressed on *pre*-sample; including baseline testosterone (2nd measure) as a covariate; ^p = one-tailed *p*-value due to preregistered hypothesis.

Table S38

Baseline T and sociosexual orientation (Further predictions, Hypothesis 10), experimental and control group

	Baseline T1		Baseline T2		Baseline T12	
	<i>r</i>	<i>P</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>P</i>
Full sample						
SOI-R total score	.26	<.001 ^p	.12	.07 ^p	.21	<.01 ^p

SOI-R attitude	.26	<.001 ^p	.10	.11 ^p	.20	<.01 ^p
SOI-R desire	.19	<.01 ^p	.06	.21 ^p	.13	.04 ^p
Partnered men only						
SOI-R total score	.16	.07 ^p	.05	.34 ^p	.10	.35 ^p
SOI-R attitude	.20	.04 ^p	.06	.30 ^p	.11	.16 ^p
SOI-R desire	.08	.23 ^p	-.08	.48	-.01	.90

Note: r = bivariate Pearson correlation; SOI-R = revised sociosexual orientation inventory; baseline T1/T2/T12 = 1st/2nd/mean of 1st and 2nd baseline testosterone measure; partnered = open, committed relationship, engaged or married; one ^p = one-tailed p -value due to preregistered hypothesis.

Tables for further replication attempts

According to the “winner effect”, derived from the biosocial model of status, winners increase and losers decrease in their T levels after competitions (or at least the T increase is larger in winners, compared to losers; Casto & Edwards, 2016a). However, in our study we detected no differential associations by outcome, neither in terms of larger T increases in winners, nor consistent interactions of outcome with T reactivity on self-reported personality state changes (though some effects were found for observer-ratings; see below). Only associations with personality state changes and effect of how close or decisive the competition outcome was on T reactivity were found (see Mehta, Snyder, Knight, & Lassetter, 2015; Table S36). These are only partly in line with predictions from the biosocial model of status and contradict a recent meta-analysis on the winner effect by Geniole and colleagues (2017). Van der Meij and colleagues (2010) suggested rather than actual outcome, perceived outcome, which would depend on cognitive and contextual factors (e.g., self-efficacy; the competition’s setting, or the “home advantage”; Fuxjager, Mast, Becker, & Marler, 2009; but see Fothergill, Wolfson, & Neave, 2017), would more likely impact T reactivity. In our study, T reactivity was associated with how close or decisive the outcome was, in that T responses were larger after close than decisive outcomes. This could be explained by contestants having perceived the competition as more challenging and been more engaged when the level of skills within a dyad was similar across disciplines, which might have triggered a larger T reactivity. In further studies, contestants’ perceived degree of being challenged and engaged in a competition could be assessed, which, given our theorizing received support, would nicely corroborate the challenge hypothesis. Thus, while a competition outcome’s decisiveness may well influence subsequent T fluctuations, the outcome was more associated with personality state changes, but not T reactivity, in our study. As a consequence, one proposed key aspect of the social environment, objective competition outcome, may not be as influential as suggested by the biosocial model of status and previous studies. Instead, T reactivity more generally translates into more competitive self-perceptions and making a more

dominant and self-assured impressions on others in winners and losers alike (Carré & Olmstead, 2015).

Table S39

Associations between competition outcome and T reactivity, interaction between competition outcome and T reactivity on self-reported and observer-reported personality state changes

	<i>1st T reactivity</i>			<i>2nd T reactivity</i>		
	<i>β</i>	<i>SE</i>	<i>t</i>	<i>β</i>	<i>SE</i>	<i>t</i>
Δ outcome → T reactivity	-.01	.18	-0.07	.15	.18	0.84
Δ outcome x T reactivity → self-reported dominance	-.04	.13	-0.28	.03	.13	0.82
Δ outcome x T reactivity → self-reported competitiveness	.01	.14	0.05	.28	.14	1.97
Δ outcome x T reactivity → self-reported coldheartedness	-.07	.14	-0.49	-.15	.14	-1.07
Δ outcome x T reactivity → self-reported introversion	-.28	.13	-2.08*	-.22	.14	-1.60
Δ outcome x T reactivity → self-reported submissiveness	.26	.13	2.02*	.01	.13	0.05
Δ outcome x T reactivity → self-reported ingenuousness	-.03	.14	-0.22	.05	.14	0.34
Δ outcome x T reactivity → self-reported nurturance	-.16	.14	-1.14	-.07	.14	-0.51
Δ outcome x T reactivity → self-reported extraversion	.20	.14	1.46	.14	.14	1.02
Δ outcome x T reactivity → observer-rated dominance (2)	.17	.14	1.20	.26	.14	1.79
Δ outcome x T reactivity → observer-rated competitiveness (2)	.05	.14	0.32	.11	.15	0.76
Δ outcome x T reactivity → observer-rated coldheartedness (2)	.06	.14	0.39	.14	.15	0.95
Δ outcome x T reactivity → observer-rated introversion (2)	.04	.15	0.29	-.08	.15	-0.57
Δ outcome x T reactivity → observer-rated submissiveness (2)	-.12	.14	-0.86	-.25	.14	-1.74
Δ outcome x T reactivity → observer-rated ingenuousness (2)	.00	.14	.03	-.08	.14	-0.52
Δ outcome x T reactivity → observer-rated nurturance (2)	-.04	.14	-0.29	-.14	.15	-0.94
Δ outcome x T reactivity → observer-rated extraversion (2)	-.03	.14	-0.24	.00	.15	-0.03

Note: outcome = competition outcome (0=lost, 1=won); 1st/2nd T reactivity = residuals from 1st/2nd testosterone *post*-sample regressed on *pre*-sample; outcome x T reactivity = interaction between competition outcome and T reactivity; 2 = Study 2; **p*<.05, ***p*<.01, ****p*<.001.

Table S40

Pearson correlations of baseline testosterone and testosterone reactivity with moderators and mediators from previous studies

	1 st baseline T	2 nd baseline T	Mean 1 st & 2 nd baseline T	1 st T reactivity	2 nd T reactivity
female confederate's ratings of men's self-displaying during the competition ¹	.04	.01	.02	.02	.04
female confederate's ratings of men's dominance during the competition ¹	-.05	-.10	-.09	.04	.04
recent sexual activity within the previous one month ²	-.10	-.11	-.12	.03	.07
recent sexual activity within the previous six months ²	.00	-.02	-.02	-.03	.09
trait dominance (PA) ³	-.06	.08	.00	.02	-.01
trait competitiveness (BC) ³	.08	-.04	.02	.07	.02
trait competitiveness (DE) ³	-.05	-.04	-.06	.10	-.01
relationship status (partnered including open relationships) ³	-.16*	-.14	-.17*	-.12	-.02
relationship status (single including open relationships) ³	-.26***	-.19*	-.27***	-.11	-.04
relationship status (ordinal categories) ³	-.21**	-.18*	-.23**	-.13	-.04

Note: full sample ($N=164-165$) for baseline T, experimental group ($n=124-125$) for T reactivity

measures; 1st/2nd baseline T = 1st/2nd baseline testosterone measure; 1st/2nd T reactivity = residuals

from 1st/2nd *post*-sample regressed on *pre*-sample; recent sexual activity 0=no, 1=yes; relationship

status: 1=single, 2=partnered, ordinal categories: 1=single, 2=open relationship, 3=committed

relationship, 4=engaged or married; references: ¹Roney, Mahler, & Maestripieri, 2003; Slatcher,

Mehta, & Josephs, 2011; ²Roney, Mahler, & Maestripieri, 2003; van der Meij, Buunk, van de Sande, &

Salvador, 2008; ³van der Meij et al., 2008; * $p<.05$, ** $p<.01$, *** $p<.001$.

Tables/figures for additional background information and analyses

Table S41

List of interpersonal circumplex items used in Studies 1 and 2

Facet	Item
PA, assured-dominant	assertive, self-assured, self-conscious, dominant, proud
BC, competitive	cynical, provocative, egoistic, aggressive, competitive
DE, coldhearted	hostile, reckless, scheming, relentless, cold
FG, introverted	unsociable, withdrawn, reserved, aloof, uninterested
HI, submissive	shy, avoidant, hesitant, submissive, conflict avoidant
JK, ingenuous	tame, obedient, adjusted, cautious, compliant
LM, nurturing	empathetic, cordial, helpful, nurturing, harmony-seeking
NO, extraverted	outgoing, communicative, approachable, sociable, interested

Note: Items from the German version of the Interpersonal Adjective List (IAL; Jacobs & Scholl, 2005).

Table S42

Participants' perception of the video documentary in the control group (n = 40)

	<i>M</i>	<i>SE</i>	<i>t</i>
Exciting	2.58	.16	0.46
Boring	2.73	.17	1.31
Informative	3.80	.13	10.40***
Disquieting	1.98	.19	-2.75**
Challenging	2.28	.20	-1.11
Stressful	1.65	.14	-6.22***

Note: One-sample *t*-tests on differences from the 5-point scale's midpoint (=2.50); **p*<.05, ***p*<.01, ****p*<.001.

Table S43

The two sets of eight "life domains" each for the video-recorded self-presentations in Study 1; presented in counterbalanced order

1 st set	2 nd set
leisure time	sports
team	companionship/communion
party	social life
energy/strength	vigour
success with women	physical attractiveness
humour	creativity
friendship	family
success in studies/job	ambition

We correlated participants' trait IAL scores with their baseline T values in order to replicate the finding of Turan and colleagues (2014). Based on their results, we hypothesized a positive correlation of T with competitiveness (facet BC) and a negative correlation with the opposing facet ingenuous (JK). We also tested potential baseline TxC interaction effects on trait IAL facets, extending the replication of Turan et al. in line of the dual-hormone hypothesis (Mehta & Josephs, 2010). In our sample (hypotheses were preregistered only for the PA facet, based on earlier findings on associations between T and dominance; e.g., Sellers et al., 2007), no significant correlations between any IAL traits and baseline T were found (all *rs* < .11, *ps* > .18, *N* = 164; see Table S5), providing no support for Turan and colleagues' findings. Our non-replication of earlier findings of inter-individual associations between dominant, competitive or nurturing personality traits and baseline T (Turan et al., 2014; Sellers et al., 2007) supports Simmons and Roney's (2011) null-finding of a link between baseline T and three measures of intrasexual competitiveness (self-reported dominance and prestige as well as physical strength) in men (*N* = 149). To further clarify potential relationships on an inter-individual difference level, alternative sampling methods, such as from hair (Dettenborn et al., 2016)

or fingernails (Matas & Koren, 2016) could be employed. These can potentially provide more stable and long-term aggregated hormonal measures, and hence potentially more valid accounts of baseline T. So far, our results corroborate the idea that there may be no strong and consistent links between T and personality on a stable trait level, and that personality and behavioural effects of T reactivity are stronger, compared to baseline T (Carré & Olmstead, 2015). In support, we found that variability in competitive personality states was positively related to men's T response, hence on an intra-individual level.

Table S44

Pearson correlations between baseline testosterone and self-reported circumplex personality traits (replicating Turan, Guo, Boggiano & Bedgood, 2014)

	Mean 1 st & 2 nd	1 st	2 nd
assured-dominant (PA)	.00	.00	.08
submissive (HI)	.06	.06	.01
competitive (BC)	.02	.02	-.04
ingenuous (JK)	-.02	-.02	.03
coldhearted (DE)	-.06	-.06	-.04
nurturing (LM)	.10	.10	.07
introverted (FG)	-.06	-.06	-.10
extraverted (NO)	.04	.04	.05

Note. Mean 1st & 2nd/1st/2nd = mean of 1st and 2nd/1st/2nd baseline testosterone measure (1st = pre-session sample; 2nd = main session sample); circumplex personality traits assessed with Interpersonal Adjective List (IAL; Jacobs & Scholl, 2005); *N* = 164.

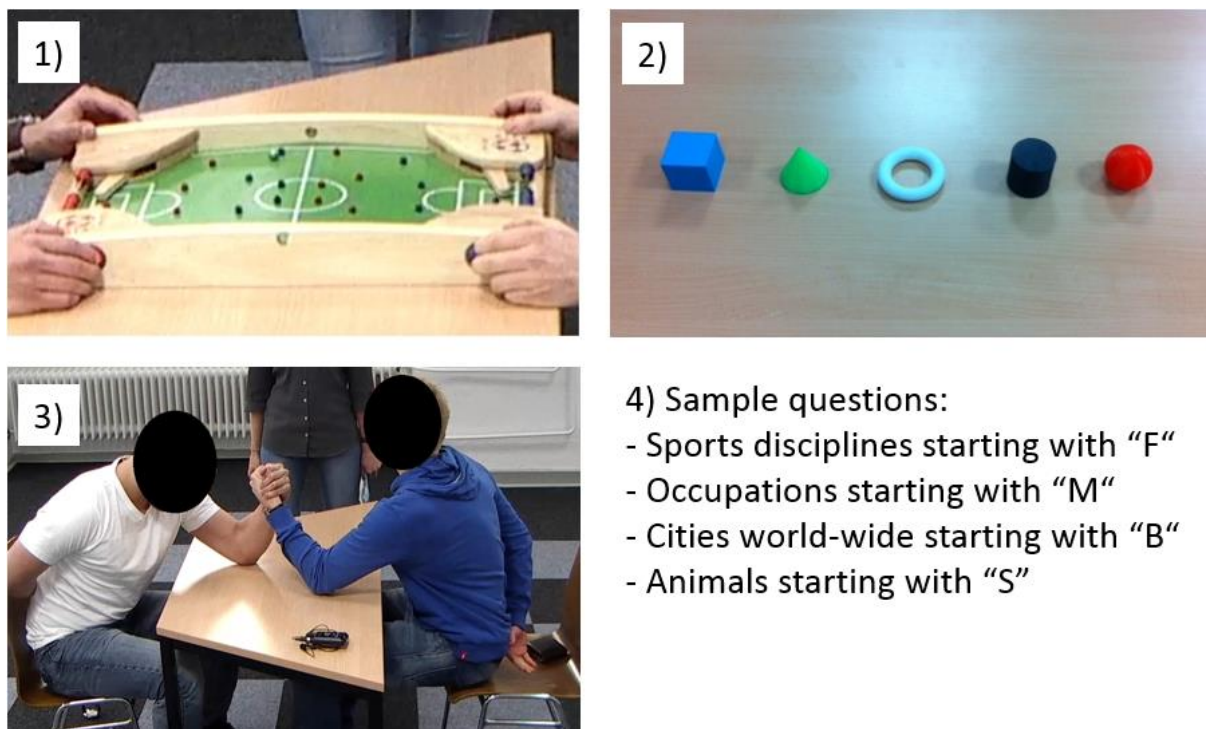


Figure S1. Illustrations of all four competition disciplines in chronological order; 1) = table pinball soccer game, 2) = snatching game, 3) = arm wrestling, 4) = turn-taking verbal fluency task. The winners and losers were determined by adding up the disciplines won. In case of a draw (i.e., if both participants had won two disciplines each), the final discipline (verbal fluency task) was counted twice, as had been announced to the participants after the third discipline.

Table S45

Social impression items for Studies 3a and 3

Study	Facet	Item
3a	Self-assurance	This man seems to be competent.
	Self-assurance	This man seems to be assertive.
	Self-assurance	This man seems to be self-confident.
	Cooperativeness	I would lend this man some money.
	Cooperativeness	I would like to work closely with this man.
	Cooperativeness	I would ask this man for a helping hand.
	Self-display	This man is trying to impress.
	Self-display	This man is showing a strong self-display.
	Self-display	This man is playing to the gallery.
3	Self-assurance	This man seems to be assertive.
	Self-assurance	This man seems to be self-confident.
	Self-assurance	*This man seems to be inhibited.
	Cooperativeness	I would enjoy working with this man.
	Cooperativeness	This man seems to me as someone who cares for other people.
	Cooperativeness	*This man does not seem to be trustworthy to me.
	Self-display	This man tries to impress.
	Self-display	This man is playing to the gallery.
	Self-display	*This man seems to be reserved.

Note: Inverted items in 3b are marked with an asterisk.

Table S46

Pearson correlations between self-reported and observer-rated personality states.

<i>r</i>	<i>pre</i>	<i>post</i>	changes
PA	.16*	.18	-.06
BC	.27***	.27**	.10
DE	.22**	.26**	.21*
FG	.14	.03	-.08
HI	.15	.03	-.04
JK	.17*	-.04	-.11
LM	.21**	.07	-.08
NO	.17*	-.04	-.11

Note: *Pre* full sample ($N=165$), *post*/changes experimental group only ($n=120$); * $p<.05$, ** $p<.01$, *** $p<.001$.

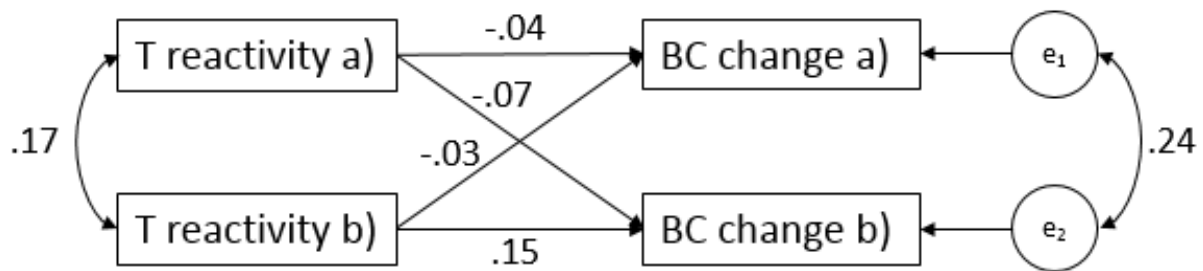


Figure S2. Standardized regression coefficients from the basic actor-partner interdependence model (APIM) for analysing the association between testosterone reactivity (1st *post*-measure) and personality state changes in competitiveness (BC facet) in Study 1; a) is the competition loser, b) the competition winner; e_1 and e_2 are error terms.

Preliminary versions of Studies 2 and 3 (in the following, Studies 2a and 3a, respectively)

2a and 3a only involved target men from Study 1's experimental group. In Study 3a we used only three items per IAL facet (from the originally eight items), and each rater was asked to judge 24 items for each video. Preregistrations can be found under osf.io/rp4qk and osf.io/76bwj for Studies 2a and 3a, respectively.

Study 2a

Methods

Participants. Participants were 120 females (age: $M = 23.8$, $SD = 2.6$, range 19-31 years), recruited via a local participant database. Raters' mean age was comparable to target men's mean age ($M = 24.3$ years).

Video-stimuli and procedure. Video-recorded self-presentations from Study 1 were used in this rating study (see above for details). All videos were cut to a maximum length of one minute. The videos of five participants in the experimental and two in the control group were removed from the stimuli sample due to audio problems, leaving a final stimulus set of *pre*- and *post*-videos each from $N = 158$ target men ($n = 120$ each for the experimental, $n = 38$ for the control group; length $M = 53$, range 10-62 sec.). Videos were distributed in a way so that the two videos of each target man never appeared together in the same set, to avoid direct contrast effects. Ratings were conducted in a computer laboratory on 24" screens using the software MediaLab v2014 (Empirisoft Corporation). Videos were presented in a randomized order.

Stimuli and items. The video stimuli were divided into eight sets of 30 videos (experimental group only), of which half were *pre*- and the other half *post*-videos each. Each video was rated by fifteen independent female raters; participants rated 30 videos each. Participants were randomly assigned to one of the eight video groups. After every video 24 items were rated by each participant. As in Study 1, the German version of the Interpersonal Adjective List (IAL; Jacobs & Scholl, 2005) was employed. For reasons of brevity we used three out of the five items per facet that we had chosen for the self-ratings in Study 1 (overall 24 items; Table S1), which were rated on a 7-point Likert scale (1 = "disagree completely" to 7 = "agree completely").

Statistical analyses. Analyses equalled those of Study 1, only substituting observer-ratings for self-reports. Only participants in the experimental group were included for Hypothesis 2.

Results

Internal consistencies (Cronbach's α) for the eight IAL facets (*pre* and *post* separately) ranged between .78-.98, and interrater agreements (Cronbach's α) for the eight facets were satisfactory to good (*pre*: $\alpha = .79$ -.92, *post*: $\alpha = .85$ -.94, changes: $\alpha = .63$ -.77; Table S48).

Hypothesis 2: *pre-post* changes in observer-rated personality. We detected a significant increase in observer-rated dominance (PA; $t = 2.73$, $p < .01$ (one-tailed), Cohen's $d = 0.17$) and extraversion (NO; $t = 2.80$, $p < .01$, Cohen's $d = 0.22$) and decreases in introversion (FG; $t = -3.27$, $p < .001$, Cohen's $d = 0.21$), submissiveness (HI; $t = -2.97$, $p < .01$, Cohen's $d = 0.19$) and ingenuousness (JK; $t = -3.04$, $p < .01$, Cohen's $d = 0.19$). When adding the preregistered control variables age, relationship status and sexual orientation to the latter linear regression models results were unchanged.

Hypothesis 3: *pre-post* personality changes & T reactivity. We found a significant positive link between change in dominance (PA) and T reactivity in the experimental group for the second hormonal *post*-sample ($\beta = 0.17$, $p = .04$ (one-tailed), $\eta^2_p = 0.03$), but not the first ($\beta = 0.12$, $p = .12$ (one-tailed), $\eta^2_p = 0.01$; for all others IAL facets, $ps > .08$; Table S49). Results remained virtually identical when adding the preregistered control variables (Tables S50).

Hypothesis 4: *pre-post* personality changes & TxC interaction. Testing for a potential moderation of the association between T reactivity and changes in observer-ratings by baseline C, no significant interactions were detected (for the first/second T *post*-sample: unsigned β s $< .15/.14$, $ps > .19$; Tables S51). Results were unchanged when including the control variables (Tables S52).

Study 3a

Methods

Participants. Eighty raters (40 females; age: $M = 24.1$, $SD = 2.9$, range 19-31 years) were recruited via a local participant database. Raters' mean age was comparable to target men's average age ($M = 24.3$ years).

Stimuli and procedure. Stimuli and procedure were the same as in Study 2a. There were eight video groups with 30 videos and each rater watched 60 videos. For each video, nine items plus a question if the target was recognized (same as in Study 2a) were rated for each target independently by ten male and ten female raters.

Items. Participants rated target men on three dimensions (self-display, cooperativeness, self-assurance) with three items each. Three positive items for each dimension were rated on a 7-point

Likert scale (1 = “disagree completely” to 7 = “agree completely”; see Table S45 for a full list of items).

Statistical analyses. Analyses equalled those of Studies 1 and 2a. Difference scores were calculated for the three social impression dimensions. Only participants in the experimental group were included for Hypothesis 2.

Results

Internal consistencies (Cronbach’s α) for the three facets (*pre* and *post* separately) ranged between .94-.99 (Table S53). Interrater agreements (Cronbach’s α) for the three facets were satisfactory to good (*pre*: α = .79-.92, *post*: α = .85-.94, changes: α = .61-.77; Table S55). Since we employed male and female raters (as explained above), at first we assessed whether rater sex had a significant effect on the observer-ratings (Hypotheses 2-4). Rater sex was added as a covariate. In addition, for Hypothesis 2, only looking at the experimental group, Cohen’s d effect sizes for *pre-post* changes for male and female raters were analysed for overlapping confidence intervals. We found no significant differences between male and female observer-ratings, neither regarding *pre-post* changes (all unsigned t s < 0.40; no overlap between males’ and females’ effect sizes in the experimental group), nor concerning T reactivity (all unsigned t s < 0.96) or the TxC interaction (all unsigned t s < 1.22). Since we detected no effect of the sex of the raters, we will present results with observer ratings aggregated across male and female raters.

Hypothesis 2: *pre-post* changes in observer-rated social impressions. In the experimental group, we found an increase in observer-rated self-display (t = 3.53, p < .001 (one-tailed), Cohen’s d = 0.28) and self-assurance (t = 3.92, p < .001, Cohen’s d = 0.22), but not cooperativeness (t = -0.22, p = .41 (one-tailed), Cohen’s d = .02). These results were robust when adding the preregistered control variables age, relationship status and sexual orientation.

Hypothesis 3: *pre-post* changes in social impressions & T reactivity. In the experimental group of both 3a and 3b, we found a positive association between changes in observer-rated self-assurance and T reactivity for the first hormonal *post*-sample (3a: β = 0.19, p = .03 (one-tailed) η^2_p =

0.03; 3b: $\beta = 0.19$, $p = .03$ (one-tailed), $\eta^2_p = 0.03$), but not the second ($\beta = 0.11$ and $\beta = 0.15$, respectively, $ps > .055$ (one-tailed); for self-display and cooperativeness, $ps > .08$ (one-tailed); Table S56). Results were virtually unchanged when adding the preregistered control variables age and relationship status (Table S57).

Hypothesis 4: pre-post changes in social impressions & TxC interaction. No significant interaction between T reactivity and baseline C, hence no moderation of the association between T reactivity and changes in observer-ratings by baseline C, was detected (for the first/second T post-sample, $ps > .051$; Table S58).

Table S47

Bivariate correlations between T reactivity and observer-rated personality state changes in Study 2

	1 st T reactivity	2 nd T reactivity
1 st T reactivity	-	.56***
PA change	.14	.15
BC change	.06	.10
DE change	.07	.06
FG change	-.04	.01
HI change	-.04	.00
JK change	-.04	-.04
LM change	-.05	.03
NO change	.08	.08

Note: 1st/2nd T reactivity = residuals from 1st/2nd testosterone post-sample regressed on pre-sample; experimental group ($n=120$); * $p<.05$, ** $p<.01$, *** $p<.001$.

Table S48

Interrater agreements in Studies 2a and 2

	Study 2a			Study 2		
	<i>Pre</i>	<i>post</i>	<i>changes</i>	<i>pre</i>	<i>post</i>	<i>changes</i>
dominant (PA)	.92	.94	.63	.81	.96	.81
competitive (BC)	.81	.88	.70	.83	.93	.83
coldhearted (DE)	.79	.85	.63	.78	.89	.78
introverted (FG)	.88	.87	.61	.67	.90	.67
submissive (HI)	.91	.92	.65	.73	.94	.73
ingenuous (JK)	.84	.87	.65	.73	.91	.73
nurturing (LM)	.85	.86	.77	.87	.93	.87
extraverted (NO)	.91	.90	.75	.80	.93	.80

Note: Groups of 10 female raters in both studies; interrater agreement = Cronbach's α .

Table S49

Associations between testosterone reactivity and observer-rated pre-post changes in the experimental group in Study 2a (Hypothesis 3)

IAL state changes	1 st T reactivity (β)	SE	<i>t</i>	Partial η^2	2 nd T reactivity (β)	SE	<i>t</i>	Partial η^2
Δ assured-dominant (PA)	.12	.10	1.17	0.01	.17	.10	1.75* ^p	0.03
Δ competitive (BC)	.11	.10	1.12	0.01	.12	.10	1.22	0.01
Δ coldhearted (DE)	.12	.09	1.40	0.02	.04	.09	0.51	0.00
Δ introverted (FG)	.02	.09	0.26	0.00	-.01	.09	-0.11	0.00
Δ submissive (HI)	-.04	.10	-0.42	0.00	-.12	.10	-1.25	0.01
Δ ingenuous (JK)	.00	.10	0.00	0.00	-.03	.09	-0.35	0.00
Δ nurturing (LM)	-.08	.08	-0.95	0.01	-.01	.08	-0.07	0.00
Δ extraverted (NO)	.01	.09	0.16	0.00	.01	.09	0.16	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample ; Partial η^2 =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S50

Associations between testosterone reactivity and observer-rated pre-post changes in the experimental group in Study 2a (Hypothesis 3), including the control variables age, relationship status and sexual identity

IAL state changes	1 st T reactivity (β)	SE	<i>t</i>	Partial η^2	2 nd T reactivity (β)	SE	<i>t</i>	Partial η^2
Δ assured-dominant (PA)	.11	.11	1.06	0.01	.17	.10	1.70* ^p	0.03
Δ competitive (BC)	.11	.10	1.03	0.01	.12	.10	1.22	0.01
Δ coldhearted (DE)	.11	.09	1.21	0.01	.04	.09	0.51	0.00
Δ introverted (FG)	.01	.10	0.10	0.00	-.02	.09	-0.17	0.00
Δ submissive (HI)	-.05	.10	-0.46	0.00	-.13	.10	-1.29	0.01
Δ ingenuous (JK)	.01	.10	-0.13	0.00	-.04	.09	-0.39	0.00
Δ nurturing (LM)	-.07	.09	-0.79	0.01	.00	.08	-0.02	0.00
Δ extraverted (NO)	.00	.09	-0.03	0.00	.00	.09	0.05	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample ; Partial η^2 =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S51

Associations between observer-rated pre-post changes and the interaction between testosterone reactivity and baseline cortisol in the experimental group in Study 2a (Hypothesis 4)

	TxC for 1 st T reactivity				TxC for 1 st T reactivity			
	β	SE	t	Partial η^2	β	SE	t	Partial η^2
Δ assured-dominant (PA)	.08	.14	0.59	0.00	-.05	.12	-0.43	0.00
Δ competitive (BC)	.05	.13	0.38	0.00	.13	.11	1.61	0.01
Δ coldhearted (DE)	.05	.12	0.42	0.00	.13	.10	1.29	0.01
Δ introverted (FG)	-.15	.12	-1.22	0.01	.00	.11	0.03	0.00
Δ submissive (HI)	-.14	.13	-1.10	0.01	-.01	.12	-0.11	0.00
Δ ingenuous (JK)	-.10	.13	-0.77	0.01	-.03	.11	-0.29	0.00
Δ nurturing (LM)	-.01	.11	-0.12	0.00	-.09	.10	-0.90	0.01
Δ extraverted (NO)	.07	.12	0.55	0.00	-.09	.11	-0.86	0.01

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample ; C=baseline cortisol (2nd sample); ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S52

Associations between observer-rated pre-post changes and the interaction between testosterone reactivity and baseline cortisol in the experimental group in Study 2a (Hypothesis 4), including the control variables age, relationship status and sexual identity

	TxC for 1 st T reactivity				TxC for 1 st T reactivity			
	β	SE	t	Partial η^2	β	SE	t	Partial η^2
Δ assured-dominant (PA)	.09	.14	0.66	0.00	-.03	.22	-0.16	0.00
Δ competitive (BC)	.04	.13	0.32	0.00	.12	.12	1.04	0.01
Δ coldhearted (DE)	.03	.12	0.25	0.00	.10	.11	0.97	0.01
Δ introverted (FG)	-.16	.12	-1.32	0.02	-.02	.11	-0.15	0.00
Δ submissive (HI)	-.15	.13	-1.15	0.01	-.02	.12	-0.15	0.00
Δ ingenuous (JK)	-.10	.13	-0.80	0.01	-.04	.11	-0.38	0.00
Δ nurturing (LM)	.02	.11	0.15	0.00	-.05	.10	-0.52	0.00
Δ extraverted (NO)	.09	.12	0.77	0.01	-.07	.11	-0.65	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample ; C=baseline cortisol (2nd sample); ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S53

Descriptive statistics and internal consistencies in Study 3a

	<i>M</i>	<i>SD</i>	Reliability
Self-assured pre	3.90	0.87	.96
Cooperative pre	4.19	0.68	.95
Self-displaying pre	3.53	0.87	.98
Self-assured post	4.09	0.88	.95

Cooperative post	4.18	0.71	.95
Self-displaying post	3.79	0.98	.98

Note: Reliability = Cronbach's α internal consistency.

Table S54

Bivariate correlations between T reactivity and observer-rated personality state changes in Study 3

	1 st T reactivity	2 nd T reactivity
1 st T reactivity	-	.57***
Δ Self-assured	.20*	.12
Δ Cooperative	-.01	.02
Δ Self-displaying	.11	.09

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; experimental group ($n=12x$); * $p<.05$, ** $p<.01$, *** $p<.001$.

Table S55

Interrater agreement on a facet level in Studies 3a and 3

	Male raters			Female raters		
	<i>pre</i>	<i>post</i>	<i>changes</i>	<i>pre</i>	<i>post</i>	<i>changes</i>
Study 3a						
Self-display	.83	.88	.61	.78	.82	.52
Self-assurance	.86	.87	.53	.84	.84	.29
cooperativeness	.73	.71	.36	.70	.75	.40
Study 3						
Self-display	.85	.88	.67	.86	.87	.65
Self-assurance	.89	.88	.55	.87	.87	.44
cooperativeness	.69	.72	.59	.72	.72	.42

Note: Interrater agreement = Cronbach's α ; Study 3a: $N=120$; Study 3b: $N=158$.

Table S56

Associations between testosterone reactivity and observer-rated pre-post social impression changes in Study 3a (Hypothesis 3)

	1 st T reactivity (β)	SE	<i>t</i>	Partial η^2	2 nd T reactivity (β)	SE	<i>t</i>	Partial η^2
Δ self-assurance	.19	.10	1.91* ^p	0.03	.11	.10	1.12	0.01
Δ cooperativeness	.07	.10	0.68	0.00	.04	.09	0.48	0.00
Δ self-display	.10	.10	1.01	0.01	.04	.10	0.40	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample ; Partial η^2 =partial eta-squared effect size; ^p = one-tailed p -value due to preregistered hypothesis; * $p<.05$, ** $p<.01$, *** $p<.001$.

Table S57

Associations between testosterone reactivity and observer-rated pre-post social impression changes in the experimental group in Study 3a (Hypothesis 3), including the two control variables age and relationship status

	1 st T reactivity (β)	SE	<i>t</i>	Partial η^2	2 nd T reactivity (β)	SE	<i>t</i>	Partial η^2
Δ self-assurance	.18	.10	1.83* ^p	0.03	.10	.10	1.06	0.01
Δ cooperativeness	.09	.10	0.90	0.01	.05	.09	0.52	0.00
Δ self-display	.09	.10	0.89	0.01	.04	.10	0.38	0.00

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample ; Partial η^2 =partial eta-squared effect size; ^p = one-tailed *p*-value due to preregistered hypothesis; **p*<.05, ***p*<.01, ****p*<.001.

Table S58

Associations between observer-rated pre-post social impression changes and the interaction between testosterone reactivity and baseline cortisol in the experimental group in Study 3a (Hypothesis 4)

	TxC for 1 st T reactivity				TxC for 2 nd T reactivity			
	β	SE	<i>t</i>	Partial η^2	β	SE	<i>t</i>	Partial η^2
Δ self-assurance	.18	.13	1.36	0.02	.21	.10	1.97	0.03
Δ cooperativeness	.04	.13	0.33	0.00	.06	.10	0.55	0.00
Δ self-display	.18	.13	1.36	0.02	.17	.10	1.64	0.02

Note: 1st/2nd T reactivity = residuals from 1st/2nd post-sample regressed on pre-sample; C = baseline cortisol (2nd sample); **p*<.05, ***p*<.01, ****p*<.001.