arTs Università degli Studi di Trieste
Archivio della ricerca – postprint

Testosterone, cortisol, and status-striving personality features: A review and empirical evaluation of the Dual Hormone hypothesis

Nicholas M. Grebe[a,*], Marco Del Giudice[b], Melissa Emery Thompson[c], Nora Nickels[d], Davide Ponzi[e], Samuele Zilioli[f,g], Dario Maestripieri[d], Steven W. Gangestad[b]

[a] Department of Evolutionary Anthropology, Duke University, Durham, NC, USA
[b] Department of Psychology, University of New Mexico, Albuquerque, NM, USA
[c] Department of Anthropology, University of New Mexico, Albuquerque, NM, USA
[d] Department of Comparative Human Development, The University of Chicago, Chicago, IL, USA
[e] Department of Psychology, Oklahoma State University, Stillwater, OK, USA
[f] Department of Psychology, Wayne State University, Detroit, MI, USA
[g] Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI, USA

ARTICLE INFO

Keywords:
Testosterone
Cortisol
Dual Hormone hypothesis
Dominance
Aggression
Status-striving

ABSTRACT

Decades of research in behavioral endocrinology has implicated the gonadal hormone testosterone in the regulation of mating effort, often expressed in primates in the form of aggressive and/or status-striving behavior. Based on the idea that neuroendocrine axes influence each other, recent work among humans has proposed that links between testosterone and indices of status-striving are rendered conditional by the effects of glucocorticoids. The Dual Hormone hypothesis is one particular instance of this argument, predicting that cortisol blocks the effects of testosterone on dominance, aggression, and risk-taking in humans. Support for the Dual Hormone hypothesis is wide-ranging, but considerations of theoretical ambiguity, null findings, and low statistical power pose problems for interpreting the published literature. Here, we contribute to the development of the Dual Hormone hypothesis by (1) critically reviewing the extant literature—including *p*-curve analyses of published findings; and, (2) "opening the file drawer" and examining relationships between testosterone, cortisol, and status-striving personality features in seven previously published studies from our laboratories (total $N = 718$; median N per feature = 318) that examined unrelated predictions. Results from *p*-curve suggest that published studies have only 16% power to detect effects, while our own data show no robust interactions between testosterone and cortisol in predicting status-striving personality features. We discuss the implications of these results for the Dual Hormone hypothesis, limitations of our analyses, and the development of future research.

1. Introduction

Testosterone (T) is a gonadal hormone crucial for basic aspects of male reproductive physiology, such as spermatogenesis, the development of secondary sexual characteristics, and sexual arousal (Dixson, 1998). T is also broadly linked to behaviors that facilitate reproductive opportunities, such as courtship, in males of diverse animal taxa—for instance, singing in birds (e.g., Ball et al., 2003). Finally, T is also related to behavioral competition for access to resources such as food or mates. Intraspecific competitive behaviors are highly species-specific and context-dependent, but in highly social animals, including many primates, intraspecific competition is regulated by dominance and/or status hierarchies, which reduce the costs of direct physical aggression (Sapolsky, 2005).

Reported associations between T and competition, status-seeking, and aggression comprise one of the most robust literatures in behavioral endocrinology, with data from numerous contexts and animal taxa (with monogamous bird species, rodents, fishes, and primates being four frequently studied groups). In humans, the field has grown large enough to warrant a number of comprehensive reviews (some examples include Dabbs and Dabbs, 2000; Eisenegger et al., 2011; Montoya et al., 2012; Casto and Edwards, 2016; Zilioli and Bird, 2017). Quantitative meta-analyses have found generally modest but statistically significant associations between T and aggression (Archer et al., 1998; Archer et al., 2005), risk-taking (Kurath and Mata, 2018), and competition outcome (i.e., greater T increases after winning; Geniole et al., 2017)—though see Van der Meij et al. (2016) for a meta-analysis suggesting no relationship between T and leadership. Discussion

* Corresponding author at: Department of Evolutionary Anthropology, Box 90383, 130 Science Drive, Durham, NC 27708, USA.
E-mail address: nicholas.grebe@duke.edu (N.M. Grebe).

continues regarding the appropriate functional interpretation of men's aggression and status-striving behavior. While many scholars argue these traits represent men's mating effort (i.e., the Challenge Hypothesis; see Archer, 2006), some recent work instead suggests that mating effort may be better reflected by motivations to avoid social exclusion, rather than status-seeking (Neel et al., 2016).

Empirical findings also challenge the straightforward interpretation of T as a predictor of aggression and status-striving. Null associations between T and traits such as self-reported dominance (Josephs et al., 2006) or aggressive behavior (Popma et al., 2007), and even negative relationships with outcomes such as perceptions of leadership ability by others (Ronay and Carney, 2013) have led to researchers seeking explanations. Perhaps, given dozens of statistical tests of associations between T and status-striving in humans, and generally modest effect sizes, some non-significant correlations are to be expected based on sampling variability alone, even if there truly exists a positive link between T and status-seeking behavior. At the same time, there are several alternative reasons why established findings might fail to replicate: statistical power might be insufficient to detect an effect; studies might adopt measures with poor construct validity; the hypothesis itself could simply be incorrect. One particular kind of argument has recently gained popularity within psychological T research: associations between T and status-striving behavior are masked, or at least rendered conditional, by other moderating variables, especially other hormones. Other hormones are attractive candidates as moderators for both theoretical and practical reasons. Focusing the pleiotropic effects of hormones into more targeted outputs may be achieved within organisms via complex interactions between hormones (e.g., for a review of the scope of interactions between gonadal hormones, such as T, and the peptide hormone oxytocin, see Gimpl and Fahrenholz, 2001). And, as behavioral endocrinology studies have already collected biological samples for assay, examining additional hormones is possible without the need to carry out an additional study.

Interactions between T and glucocorticoids in particular have received substantial attention in the literature. Glucocorticoids—a class of steroid hormones that includes cortisol (C) in primates and corticosterone in rodents and birds—are secreted by the adrenal glands during stress and energetically demanding events. The basis for expecting interactions between T and glucocorticoids comes from both behavioral and physiological findings generated over the last 30 years. One seminal line of work has investigated effects of stress on reproductive behavior, likely reflecting interactions between T and glucocorticoids (though other hormones and biological mechanisms are also involved in the stress—reproduction link). Chronic elevation of glucocorticoids, whether via stressors (e.g., infections, handling in livestock, isolation in social species) or experimental administration, leads to a reduction in mating behavior, assessed via courtship effort (Moore and Miller, 1984), mounting latency and ejaculatory frequency (Retana-Marquez et al., 2003), or territoriality (Wingfield and Silverin, 1986; but see null effects on T). Notably, however, transient stressors may have null or even potentiating effects on mating behavior (for a review, see Tilbrook et al., 2000). And, species with temporal constraints on reproduction, such as some birds, can be buffered from mating deficits that typically result from elevated glucocorticoids (e.g., Astheimer et al., 2000).

There also exists suggestive evidence of more proximate, physiological interactions between T and glucocorticoids. Each hormone is the end product of their respective axes: T of the hypothalamic-pituitary-gonadal (HPG) axis, and C/corticosterone of the hypothalamic-pituitary-adrenal (HPA) axis. However, the upstream physiological systems producing these hormones communicate, and the HPA and HPG axes may mutually modulate one another (Viau, 2002). The level at which this mutual modulation occurs is a continuing debate within the extant literature. Some scholars have argued, for example, that T dampens the classic HPA 'stress response' upstream, at the level of the hypothalamus (Viau, 2002; Montoya et al., 2012),[1] though this may occur via androgen receptors outside of the hypothalamus (Handa and Weiser, 2014); others argue that inhibition occurs downstream only, at the level of the adrenal gland (Rubinow et al., 2005). There is more agreement on the idea that products of the HPA axis inhibit HPG function at all levels. Centrally, corticotropin-releasing hormone (CRH) inhibits secretion of gonadotropins, and glucocorticoids (such as cortisol) can decrease pituitary sensitivity to gonadotropins (Tilbrook et al., 2000). Peripherally, glucocorticoids may act directly upon the gonads to inhibit their endocrine functioning (Johnson et al., 1992). Once again, however, a contrasting view exists. Ketterson and Nolan Jr (1999) note that, in studies of birds, T positively associates with corticosterone. They propose that T induces a state of physiological stress. Through its metabolic effects, glucocorticoids may fuel activities that are also promoted by T (i.e., those that represent increased mating effort). Though Ketterson and Nolan Jr (1999) do not explicitly predict positive interactions between T and glucocorticoids, their proposal might lead one to expect positive interactions, at least for some energetically intensive activities. This proposal is also consistent with a large body of psychological studies in humans showing co-occurring T and C increases in response to challenges or stressors (see e.g. Bateup et al., 2002; Bobadilla et al., 2015; Shirtcliff et al., 2015). One recent paper suggests that administering T to men raises cortisol levels, though perhaps only men already high in dominance (Knight et al., 2017). In sum, while the precise mechanisms and directionality of interactions between T and glucocorticoids remain a matter of ongoing debate, a mounting body of evidence is consistent with the argument that the two hormonal systems can modulate one another, and that their interactions have consequences for downstream behavior or psychological traits.

1.1. The "Dual Hormone" hypothesis

Mehta and Josephs (2010) explicitly formulated a hormonal interaction hypothesis, claiming that C interacts with T to predict status-seeking behavior in humans. They also coined a label for this idea: the "Dual Hormone" hypothesis. Citing mutual inhibition between the HPA and HPG axes, Mehta and Josephs (2010) proposed that C "may block the influence of T on dominance" (p. 899). In two studies, they found that T and C interact to predict observers' ratings of dominance in men and women in a leadership task (Study 1) and likelihood of decisions to re-challenge one's opponent after competition among men (Study 2),[2] such that associations between T and these indices of status-striving behavior are strongest at low concentrations of C. This original prediction of the Dual Hormone hypothesis, that high C masks associations between T and dominance, has been replicated with other putative scales of dominance (e.g., from the International Personality Item Pool; see Pfattheicher, 2017) and extended to patterns of aggression (Popma et al., 2007), overbidding in an experimental economic game (van den van den Bos et al., 2013), destructive behavior (Pfattheicher et al., 2014), risk-taking (Mehta et al., 2015b), status among teammates (e.g., Edwards and Casto, 2013) and empathy (with an inverse association between empathy and T, modulated by C; Zilioli et al., 2015), among other domains (reviewed by Mehta and Prasad, 2015).

Yet, despite this level of support, the Dual Hormone hypothesis remains a relatively new idea that, for several reasons, demands a critical eye and requires further investigation. First, positive findings need to be considered within the larger literature of null and even reversed

[1] Though Montoya et al. (2012) cite Viau (2002) as support for the claim that T inhibits HPA activity at the level of the hypothalamus, Viau's argument is more specific: "Thus, while testosterone regulation of stress-induced ACTH release can be explained by the inhibition of AVP biosynthesis, sex steroid regulation of the neuroendocrine arm of the HPA axis cannot occur directly at level of the PVN" (p. 508)

[2] In Study 2, however, the T x C interaction was further moderated by whether men won or lost the competitive task. See Table 1 for a summary of three-way interactions within the literature on the Dual Hormone hypothesis.

findings. Multiple studies have now demonstrated a reversed T × C interaction, whereby the associations between T and behavior are strengthened at high levels of C for outcomes such as self-reported risk-taking (Barel et al., 2017) or aggressive behavior toward a competitor in a laboratory task (Denson et al., 2013). In addition to these interactions found in an unexpected direction, there are also a number of failures to replicate or find T × C interactions on aggressive behavior and attitudes (Scerbo and Kolko, 1994; Carré and Mehta, 2011[3]; Platje et al., 2015), psychopathy (Glenn et al., 2011), competitive behaviors (Salvador et al., 1999), risk-taking (Cueva et al., 2015), and past deviant behaviors (Mazur and Booth, 2014). Importantly, a new meta-analysis of Dual Hormone T × C effects finds a small average effect in the predicted direction, though this is qualified by evidence of publication bias and analytic flexibility (Dekkers et al., 2019).

Still other findings provide a more ambiguous level of support for the Dual Hormone hypothesis, with T × C interactions only emerging with the addition of another moderator variable (e.g., winning vs. losing a competitive task—Mehta and Josephs, 2010; personality traits—Tackett et al., 2014; being in a 'social inclusion' experimental condition—Geniole et al., 2011). These more complex patterns go beyond the originally described hypothesis that "T and C jointly regulate behavior such that higher T should be positively related to dominance only when C is low" (Mehta and Josephs, 2010, p. 899).

In Table 1, we list a summary of published results from a literature search for T × C interactions predicting status-striving and aggression, grouped by level of support for the conventional (as defined by Mehta and Josephs, 2010) Dual Hormone hypothesis. To create this table, we started by including the studies listed in Mehta and Prasad's (2015) review of the Dual Hormone hypothesis (with the exception of Zilioli and Watson, 2012, which assesses a physiological, rather than behavioral, Dual Hormone outcome). We then identified additional pertinent effects by a) reviewing the publications listed by Mehta and Prasad and identifying any relevant null effects from these studies (e.g., for Tackett et al., 2014, we include both the significant T × C interaction on self-reports of externalizing behavior, as well as the null T × C interaction on parent reports of these same behaviors); b) searching Google Scholar using the terms "Dual Hormone", "testosterone cortisol behavior", "testosterone cortisol interaction", and "testosterone cortisol moderation"; and c) including any additional published papers familiar to us that were not identified by a) or b) but nevertheless assessed the interaction between T and C in predicting status-striving behavior. We concluded our search in May 2018. We extracted all relevant effects from the resulting papers that 1) measured T and C concentrations in humans and 2) examined how T × C interactions predicted indices of status-striving, aggression, dominance, or risk-taking. With these search strategies, we identified a total of 55 effects. Our table visually represents what we perceive as a lack of a clear trend: Though a number of findings support the Dual Hormone hypothesis, many null and negative results have appeared too. In many cases, a single study contains some supportive effects, but also null ones (e.g., Mehta et al., 2015a; Pfattheicher, 2017; Prasad et al., 2017; Tackett et al., 2014). Outcomes in Dual Hormone studies are also heterogeneous, assessing features ranging from anti-social punishment behavior (Pfattheicher et al., 2014) to social network centrality (Ponzi et al., 2016) to risk-taking in stock trading (Cueva et al., 2015). One outcome measure used across multiple studies (the Balloon Analog Risk Task) yields both significant and non-significant Dual Hormone effects (see Mehta et al., 2015b; Ronay et al., 2018).

Equivocal findings in the published literature relate to a second issue: a degree of imprecision or subjectivity in interpretation. By our reading, it is not completely clear how many results should be viewed in terms of their degree of support of the Dual Hormone hypothesis. As an illustrative example, consider Dual Hormone effects in the Ultimatum Game, a dyadic economic game in which a proposer makes an offer as to how to split a sum of money with a responder. The responder can accept the offer as proposed, or reject the offer so that both players receive nothing. Often, responders will reject 'unfair' (i.e., unequal) offers from proposers. One study reported that a negative T × C interaction (presented as high basal T, and short-term decreases in C) predicts the likelihood of responders rejecting unfair offers (though the authors found no T change × C change interaction; Prasad et al., 2017). A second paper reported that a similar attenuating interaction (increases in T, and decreases in C) predicted the *acceptance* of these same unfair offers (though the authors found no basal T × basal C interaction; Mehta et al., 2015a). In the case of these two opposing findings, it is unclear which particular combination of hormonal measurements (basal, short-term changes, or a mix) and outcome measurements (acceptance or rejection of unfair offers) offers the most appropriate test of the Dual Hormone hypothesis. Other questions of interpretation are more general. For instance, aggressive behavior is by no means the only (or even primary) tactic for status attainment in humans (see von Rueden et al., 2011; Eisenegger et al., 2011). Do Dual Hormone effects on aggressive behavior truly reflect status-striving? For which outcomes should a positive, potentiating T × C interaction be predicted, rather than the 'conventional' attenuating interaction (see e.g. Bobadilla et al., 2015)? In what instances should a moderator be expected for T × C (i.e., a three-way interaction—see Table 1 for examples)—and should the typical, 2-way interaction still emerge in such cases? Without clear definitions of what constitutes support, or lack thereof, for the Dual Hormone hypothesis, it is nearly impossible to evaluate the literature as a whole.

Finally, an analysis of the significant, predicted (i.e., attenuating) T × C interactions in the published literature suggests that most studies are underpowered to detect these effects. The *p*-curve (Simonsohn et al., 2014) is a distribution of *p*-values that are published, statistically significant (i.e., ranging from ~0 to 0.05), and in the predicted direction for a given research domain. The shape of the *p*-curve provides diagnostic information regarding the evidential value in a set of studies (versus the influence of *p*-hacking—subjective, defensible decisions made by scholars during the research process that artificially inflate the likelihood of obtaining statistically significant effects). But *p*-curve analysis also provides information about the estimated statistical power to detect a real effect in a set of studies. Given even modest statistical power (30%), a *p*-curve examining true effects will be markedly right-skewed, with 43% of *p*-values under .01 (this is due to the non-central distribution of test statistics; e.g. Hung et al., 1997). Fig. 1 shows a *p*-curve of the significant, independent,[4] attenuating T × C interaction *p*-values listed in Table 1. The *p*-curve analyses are equivocal as to the evidential value of the reported effects (i.e., the analyses cannot determine whether selective reporting is the sole explanation for the given effects; see Simonsohn et al., 2014 and supplementary online materials [SOM] for interpretation). But more firmly, *p*-curve estimates the power of these interaction tests to detect real non-zero effects, if they exist at the level *p*-curve offers a best-estimate of, to be just 16% [90% CI: 5%–44%]. Statistical power of only 16% presents problems of both an inflated false negative rate (i.e., given a real T × C interaction effect, underpowered studies will often fail to detect it) and an inflated estimate of effect size ("the winner's curse"; see Ioannidis, 2008).[5]

[3] In this article, Carré and Mehta cite a personal communication with an author of another study (Victoroff et al., 2011), who reported to them no T x C interaction on self-report measures of aggression.

[4] Two studies (Mehta et al., 2015b; Ponzi et al., 2016) contained multiple significant Dual Hormone effects. As *p*-curve should only be used to analyze statistically independent effects (Simonsohn et al., 2014), we selected the median *p*-value from studies with multiple significant effects.

[5] *P*-curve has also been discussed as a means to detect *p*-hacking, procedures through which researchers inflate Type I error rates by sifting through multiple ways of performing analyses. As the flat distribution of significant *p*-values in Fig. 1 can arise through sampling variability and publication bias alone, in absence of *p*-hacking, by no means do we imply that significant results have been *p*-hacked.

Table 1
Previous published findings on the Dual Hormone hypothesis.

Outcome examined	N[a]	Sex	Test statistic/*p*-value[b]	Reference
Effects consistent with the conventional "Dual Hormone Hypothesis" prediction (attenuating T × C interaction)				
Group performance in economic game	74	M & F	$r = -0.27$, $p = .03$	Akinola et al., 2016
Violence of convicted crime (prison inmates)	113	M	$t(112) = -2.11$, $p = .037$	Dabbs et al., 1991
Teammate–rated status	74	F	$F(1,70) = -6.62$, $p = .012$	Edwards and Casto, 2013
Performance in competitive reaction time task	115	F	$t(114) = -3.17$, $p = .002$	Henry et al., 2017
Aggressive behaviors (self−reported)	460	M & F	$b = -0.12$, $p = .011$	Grotzinger et al., 2018
Observer-rated dominance in leadership task	94	M & F	$F(1,82) = -4.96$; $p = .029$	Mehta and Josephs, 2010
Sellers' earning in negotiation (T & C reactivity)	64	M & F	$b = -0.74$, $p = .047$	Mehta et al., 2015a[c]
Behavioral risk-taking (balloon task)	160	M	$t(156) = -2.67$, $p = .008$	Mehta et al., 2015b
Trait risk-taking (self-report)	110	M & F	$t(106) = -2.36$, $p = .020$	Mehta et al., 2015b
Trait risk-taking (other-report)	92	M & F	$t(88) = -3.02$, $p = .003$	Mehta et al., 2015b
Antisocial punishment in economic game	72	M	$t(70) = -2.27$, $p = .026$	Pfattheicher et al., 2014 (Study 1)
Dominant behavior in dictator game	151	M	$\beta = -0.537$, $p = .006$	Pfattheicher, 2017
Social network centrality (popularity)	41	M	$\beta = -0.41$, $p = .03$	Ponzi et al., 2016
Social network centrality (betweenness)	41	M	$\beta = -0.42$, $p = .03$	Ponzi et al., 2016
Rejection of unfair offers (C reactivity)	39	M & F	$t(35) = -2.34$, $p = .025$	Prasad et al., 2017
Self-reported overt aggression	100	M	$t(91) = -2.37$, $p = .020$	Popma et al., 2007
Behavioral risk-taking (balloon task)	43	M	$b = -135.84$, SE = 49.5, $p < .05$	Ronay et al., 2018
Business executives' number of subordinates	74	M	$\beta = -0.35$, $p = .005$	Sherman et al., 2016
Externalizing behaviors (self−reported)	106	M & F	$\beta = -0.21$, $p = .032$	Tackett et al., 2014
Overbidding	23	M	$t(18) = -3.19$, $p = .005$	van den Bos et al., 2013
Self-reported trait empathy	315	M	$F(1, 311) = 5.77$, $p = .017$	Zilioli et al., 2015
Effects counter to the conventional "Dual Hormone Hypothesis" prediction (potentiating T × C interaction)				
Self-reported risk-taking	77	M & F	$t(75) = 2.41$, $p = .019$	Barel et al., 2017
Aggressive behavior toward competitor	53	F	$t(43) = 2.93$, $p = .005$	Denson et al., 2013
Cheating for financial gain (Study 1)	82	M & F	$\beta = 0.58$, $p = .03$	Lee et al., 2015
Cheating for financial gain (Study 2)	117	M & F	$\beta = 0.25$, $p = .04$	Lee et al., 2015
Self-reported psychopathy	187	M	$t(176) = 2.31$, $p = .022$	Welker et al., 2014
Null "Dual Hormone" findings (no significant T × C interaction)				
Aggressive response to threat	39	F	Not reported; $p > .478$	Buades-Rotger et al., 2016
Endorsement of sociopolitical aggression	41	M	Not reported; non-significant	Carré and Mehta, 2011; cf. Victoroff et al., 2011
Risk-taking in stock trading	137	M & F	$b = 1.256$, −0.289 (M, F); both $p > .05$	Cueva et al., 2015
Aggressive behavior in experimental game	63	M	Not reported; $p > .14$	Geniole et al., 2011
Aggressive behavior in experimental game	201	M & F	$t(99) = -1.15$, $t(92) = 0.47$ (M, F); $ps > 0.25$	Geniole et al., 2013
Clinician-rated psychopathy	178	M & F	$\beta = 0.06$, $p = .52$	Glenn et al., 2011
Rule-breaking behaviors (self−reported)	460	M & F	$b = -0.05$, $p > .05$	Grotzinger et al., 2018
Self-reported deviant behavior	4462	M	Not reported; non − significant	Mazur and Booth, 2014
Emergent status in triadic interactions	45	M	$t(25) = 1.96$, $t(9) = 0.46$ (Study 1, 2); $ps > 0.05$	Mazur et al., 2015
Buyers' price negotiated (T & C reactivity)	64	M & F	Not reported; $p > .60$	Mehta et al., 2015a
Sellers' earning in negotiation (basal T & C)	64	M & F	Not reported; $p > .40$	Mehta et al., 2015a[c]
Dominant decisions in economic game	98	M & F	Not reported; $p > .20$	Mehta et al., 2017
Self-reported aggressive behavior	259	M & F	$\beta = 0.20$, $p = .55$	Platje et al., 2015
Social network centrality (gregariousness)	41	M	$\beta = -0.08$, $p = .69$	Ponzi et al., 2016
Self-reported covert aggression	100	M	Not reported; non − significant	Popma et al., 2007
Self-reported psychopathic traits	128	M	$t(123) = -0.197$, $p = .844$	Pfattheicher, 2016
Trait dominance	151	M	Not reported; $p = .67$	Pfattheicher, 2017
Rejection of unfair offers (basal C)	39	M & F	$\beta = 0.25$, $p = .15$	Prasad et al., 2017
Behavioral risk-taking (balloon task)	97	F	$b = 21.04$, $p > .05$	Ronay et al., 2018
Dominant/offensive behavior in judo match	27	M	Not reported; non − significant	Salvador et al., 1999
Teacher/staff rated aggression in children	40	M & F	Not reported; non − significant	Scerbo and Kolko, 1994
Preferences for financial risk−taking	247	M	$t(239) = 0.74$, $p = .463$	Smith and Apicella, 2017
Externalizing behaviors (parent−reported)	106	M & F	$\beta = -0.11$, $p = .246$	Tackett et al., 2014
Self-reported dominance	36	M	Not reported; non − significant	van der Westhuizen and Solms, 2015
Self-reported psychopathy	187	F	$t(176) = -0.64$, $p = .53$	Welker et al., 2014

(*continued on next page*)

Table 1 (*continued*)

Outcome examined	N[a]	Sex	Test statistic/*p*-value[b]	Reference
Reading the Mind in the Eyes Test	315	M & F	Not reported; $p = .996$	Zilioli et al., 2015
Moderated "Dual Hormone" findings (significant 3-way, but not 2-way, T × C interactions)				
Friendly behavior in role–playing activity	80	M & F	$B = 31.15$, $p = .04$	Lozza et al., 2017 (moderated by experimental condition)
Decision to re-challenge an opponent	57	M	$\chi^2 = 6.30$, $p < .05$	Mehta and Josephs, 2010 (moderated by outcome)
Externalizing behaviors (parent–reported)	106	M & F	$\beta = -0.25$, $p < .001$	Tackett et al., 2014 (moderated by personality traits)

[a] These sample sizes represent the effective sample size used in statistical analyses; frequently, this number was smaller than the total number of participants, due to exclusion of outliers, insufficient sample volume for multiple assays, etc.

[b] For ease of interpretation, all test statistics/parameters in the direction predicted by the Dual Hormone hypothesis are assigned a negative value in the table, even if coded to have a positive sign in the original paper.

[c] This paper also includes a second study and Dual Hormone finding. However, this other effect the authors highlight as supportive of the Dual Hormone Hypothesis—the accepting of unfair offers in the Ultimatum Game—is at odds with characterizations of attenuating T × C interactions as potentiating status-striving behaviors (indeed, Prasad et al. [2017] explicitly interpret the **rejection** of unfair offers as an aggressive response that supports the Dual Hormone hypothesis). For this reason, we omit this effect from our table.

1.2. The present studies

Given the state of the field, additional work is needed. Studies with larger samples are especially desirable, as they possess reasonable power to detect real effects, and positive effects are less likely to be chance findings. In the current paper, we report findings from seven studies, with total *N* of 718. These studies were not designed with the intent of testing the Dual Hormone hypothesis. Nonetheless, on these samples, some of which were relatively large, we happened to have measures related to status-striving and risk-taking (median *N* for measures used = 318), as well as measures of salivary T and C (in many instances, multiple baseline measures, such that T and C could be aggregated). Hence, we see our contribution as an opening of the file-drawer, which may be one route by which researchers can gain insight into the nature and strength of effects. Recently, questions about replicability within psychology have led some labs to make public all of their data that pertain to effects of interest, whether statistically significant or not (e.g., Lane et al., 2016; Mann and Spellman, 2016). Like meta-analysis, opening the file drawer aims to clarify the nature of an association by considering in aggregate the evidence provided from the sum of available evidence.

2. Methods

2.1. Participants

Participants were recruited in seven separate studies, which were conducted to test unrelated predictions.

2.1.1. Sample 1

A total 98 men were recruited for a study examining the association between fluctuating asymmetry and oxidative stress biomarkers. We report on a subsample of 97 for which we obtained both T and C levels (for further details, see Gangestad et al., 2010a).

2.1.2. Sample 2

A total of 150 participants (75 couples) were recruited for a study examining oxytocin levels and responses in romantically involved partners. Once again, the subset here ($N = 142$; 70 women) includes those for which we had measures of both T and C (for further details, see Grebe et al., 2017).

2.1.3. Sample 3

A total of 70 couples were recruited for a study examining changes in romantically involved women's sexual interests, and male partners' reactions to them, across the cycle. The subset on which we obtained assays of salivary T and C totals 90 participants (46 women). In this sample, all women were normally ovulating (for further details, see Gangestad et al., 2005, 2010b, 2014; Garver-Apgar et al., 2006).

2.1.4. Sample 4

A total of 152 adults (77 men) were recruited for a study on personality traits, sociosexual relationships, and hormones (measured at one time point). Out of the total sample, we include a subset of 122 participants for whom one T and one C datapoint were available (only one saliva sample per subject was collected in this study; for further details, see Maestripieri et al., 2013; Maestripieri et al., 2014; Del Giudice et al., 2014).

2.1.5. Sample 5

A total of 109 men were recruited for a study on personality traits, responses to stressful and sexual stimuli, and hormones. In this study, three T and three C samples were collected on two separate days, but here we only report on the two baseline samples. The subset with available T concentrations comprises 107 men, while 108 had available C concentrations (for further details, see Ponzi et al., 2016; Zilioli et al., 2016).

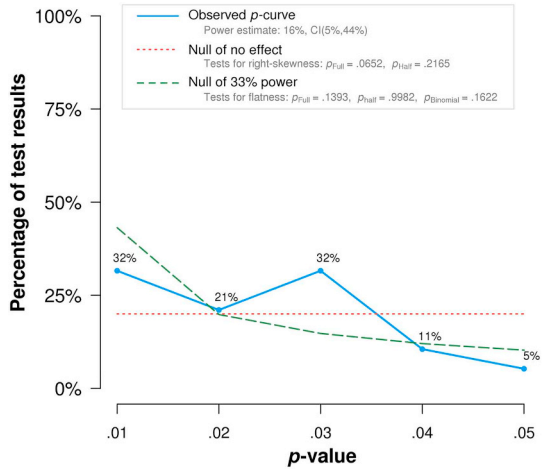

Fig. 1. *P*-curve of 19 significant, independent published results supporting the Dual Hormone hypothesis (see Table 1). Figure generated by p-curve.com, application version 4.06.

2.1.6. Sample 6

A total of 126 participants (65 men) were recruited for a study on personality traits, chronotypes, impulsivity, and responses to stressful stimuli. In this study, three T and three C samples were collected on a single day; we only report on the two baseline samples for each hormone. 120 participants had T and C measurements available (for further details, see Marvel-Coen et al., 2018).

2.1.7. Sample 7

A total of 41 male rugby players were recruited for a study of dominance, prestige, and social networks. Thirty-nine participants provided one saliva sample before and one saliva sample after team practice. One participant provided his saliva sample on a non-practice day, while one participant did not provide saliva. T and C were assayed on the saliva samples provided before team practice and the saliva sample provided on a non-practice day (final $N = 40$). For further details, see Ponzi et al. (2016).

2.2. Psychological measures

Across the seven samples, we obtained a variety of self- and partner-reported personality measures that relate to status-striving or risk-taking—the predominant constructs represented in the Dual Hormone literature. We note that self-reports may not be ideal for capturing overt behaviors—the outcomes some scholars argue to be most closely predicted by hormonal fluctuations (e.g., Mazur and Booth, 1998). However, a) five of the 21 previously published, statistically significant Dual Hormone effects similarly rely on self-report measures, suggesting these measures possess some validity as tests of the Dual Hormone hypothesis; and b) we obtained partner-reports for some of our measures (in Sample 3; see below), which provide validating information for self-reports and may reflect observable behavior, even independent of self-reports (see, e.g., Connelly and Ones, 2010, who argue that observer reports are "strong predictors of behavior" [p. 1092]); c) some "self-reports" are not personality scales but rather are, in effect, self-reported behavioral surrogates—reports of whether participants had engaged in particular behaviors in the past two days; Samples 2 and 3, below).

We divided our outcome measures into two separate categories—"core" and "secondary"—based on the degree to which they reflect the original formulation of the Dual Hormone hypothesis. Below, we justify our choice of measurements and their categorization.

2.2.1. Social potency

The Social Potency Scale, one measure on the Multidimensional Personality Questionnaire (MPQ; Tellegen, 1982), consists of 26 items and is generally considered a measure of social dominance, validated with peer-reports (Tellegen, 1982). E.g., "People consider me forceful;" "I am quite good at getting others to see my way." This measure was administered to Samples 1–3; mean $\alpha = 0.86$ (SD = 0.02). The romantic partners of participants in Sample 3 completed the same measure re-worded to concern a partner. E.g., "People consider him (her) forceful;" $\alpha = 0.89$. Self-reports correlated substantially with partner-reports, $r(68) = 0.48$ and 0.43, $p < .001$, for men and women, respectively. Due to its strong associations with published Dual Hormone outcomes such as psychopathic personality (e.g., Benning et al., 2003) and descriptions of social potency as a measure of traits such as "fondness for leadership roles" (Caspi et al., 1997), social potency was categorized as a "core" outcome in analyses.

2.2.2. Non-submissiveness

This measure consists of 16 items and was created to tap individual differences in unwillingness to tolerate, without a counter-response, actions by others aimed to diminish one's status or social standing. E.g., "When other men [women] "cross the line" with me, I am not afraid to enter into a conflict with them;" "Most people probably respect me for my willingness to stand up for myself." This measure was administered to Samples 1–3, mean $\alpha = 0.81$ (SD = 0.06). The romantic partners of participants in Sample 3 completed the same measure re-worded to concern a partner; $\alpha = 0.91$. Self-reports correlated substantially with partner-reports, $r(68) = 0.55$ and 0.39, $p \leq .001$, for men and women, respectively. Non-submissiveness covaries substantially with Social Potency as well: for self-reports, $r = 0.55$ and 0.54 for women in Samples 2 and 3 ($N = 73$, 71) and 0.38, 0.54, and 0.51 for men in Samples 1–3 ($N = 73$, 97, 70; all $p < .001$). Non-submissiveness was also categorized as a "core" outcome.

2.2.3. Winning intrasexual competitive behaviors

Participants in Samples 2 and 3 were asked on two different occasions within a month to report how often, in the past two days, they had engaged in specific behaviors that "got the best of" a same-sex person (5 items, 3 items for the Samples 2 and 3, respectively). Items included, "I came away feeling I got the upper hand after a conflict with another man [woman];" "I humiliated another man [woman]." Responses were aggregated across the two reports; $\alpha = 0.90$ and 0.73 for Study 2 and 3, respectively. These reports, which reflect an individual's motivation to maintain status over same-sex competitors, were categorized as "core" outcomes.

2.2.4. Extraversion

The 60-item NEO Five-Factor Inventory (McCrae and Costa, 2004) was administered to participants in Samples 1 and 2. The Extraversion Scale consists of 12 items. E.g., "I like to have a lot of people around me;" "I like to be where the action is." $\alpha = 0.78$, 0.85. Extraversion as a personality dimension includes facets that relate to dominance, assertiveness, and sensation-seeking, which are higher in men (consistent with these traits being modulated by T). However, extraversion also contains a sociability element that is higher in women (Weisberg et al., 2011); thus, extraversion is not an ideal index of status-striving. Here we include it as a "secondary" outcome.

2.2.5. Agreeableness

The Agreeableness Scale of the Short-Form of the NEO Personality Inventory consists of 12 items. E.g., "Some people think I'm selfish and egotistical" (reverse-coded); "If someone starts a fight, I'm ready to fight back" (reversed-coded). Reversed agreeableness was included as a "core" outcome, due to its sizable positive relation to empathy and negative relation to aggression (Tremblay and Ewart, 2005; Graziano et al., 2007). It was administered to Samples 1 and 2 ($\alpha = 0.76$, 0.68).

2.2.6. Eysenck Impulsiveness Questionnaire (EIQ) Venturesomeness and Impulsivity

Eysenck and Eysenck (1978) developed a measure of two facets of impulsivity, Venturesomeness (17 items) and Impulsiveness (26 items). Eysenck and colleagues (see Eysenck and Eysenck, 1978; Eysenck et al., 1984) define both facets in terms of a propensity to take risks, but with Impulsiveness characterized by a lack of awareness of such risks, and Venturesomeness characterized by an acute recognition of risk. As Venturesomeness directly taps sensation-seeking, a close relative to risk-taking (Lejuez et al., 2002; Whiteside and Lynam, 2001), we categorized it as a "core" outcome. Impulsiveness was categorized as a "secondary" outcome, due to its smaller correlation with sensation-seeking, and small sex differences (the large majority of status-striving traits show significant sex differences; see Del Giudice et al., 2014). The measures were administered to Samples 4–6. Mean α across both dimensions and all three samples was 0.70 (SD = 0.11).

2.2.7. Zimbardo Time Perspective Inventory (ZTPI)

The ZTPI (Zimbardo and Boyd, 1999; D'Alessio et al., 2003) consists of 38 items, which factor analysis shows tap three dimensions (D'Alessio et al., 2003): Hedonistic Present, Fatalistic Present, and Future Time Perspective (16, 9, and 13 items loading on each factor, respectively). All three dimensions strongly relate to impulsivity, and only moderately to sensation-seeking; thus, they were categorized as "secondary" outcomes (see Del Giudice et al., 2014). The measures were administered to Samples 4–6. Mean α across dimensions and samples was 0.73 (SD = 0.12).

2.2.8. Scale of Intrasexual Competitiveness (SIC)

The SIC was developed by Buunk and Fisher (2009) to assess the degree to which people view interactions with same-sex individuals in a competitive way. The scale consists of 12 items, e.g., "I always want to beat other men [women] in competition;" "When I'm at a party, I enjoy it when women [men] pay more attention to me than other men [women]." The items of the SIC involve striving for status and attention from the other sex, and the scale predicts competitive behaviors in various contexts (Buunk and Fisher, 2009). For this reason, this measure was categorized as a "core" outcome. The SIC was administered to Samples 4 and 5 (α = 0.87, 0.78).

2.2.9. Dominance and prestige

Cheng et al. (2010) developed a scale assessing two broadly different means of obtaining status in human societies. Within the 22-item scale, two readily interpretable factors emerged: one that contains items pertaining to dominance (10 items; e.g., "I try to control others rather than permit them to control me"; "I dislike giving orders" [reverse-scored]), and a second pertaining to prestige (12 items; e.g. "My unique talents and abilities are recognized by others"; "Others seek my advice on a variety of matters"). This scale was administered to Sample 7; α = 0.78 and 0.67 for dominance and prestige, respectively. These measures, both intended to capture status-striving, were categorized as "core" outcomes.

2.3. Hormone assays

For all samples, participants provided a sufficient volume of saliva for hormonal assays through passive drool on at least one occasion: for Samples 1 and 2, two samples collected approximately 1 week apart; for Sample 3, two consecutive days' samples, where the two collections were typically separated by 1 to 3 weeks; for Sample 4, one sample taken at the beginning of laboratory procedures; for Sample 5, two baseline samples provided at the beginning of a pair of experimental sessions; for Sample 6, a single sample given in the laboratory; for Sample 7, two saliva samples were collected (for more details, see Participants section). For Samples 1 and 2, the first sample was collected at the beginning of a lab session and the second sample was collected shortly upon morning awakening. For Sample 3, all samples were collected early evening. For Sample 4, all samples were collected between noon and 4 PM; Samples 5 and 6 were collected throughout the day. Sample 7 was provided between 6 PM and 9 PM.

T and C were assayed from all available samples (Samples 1, 2 and 4: the Hominoid Reproductive Ecology Lab, Department of Anthropology, University of New Mexico; Sample 3: the Clinical Ligand Assay Service Satellite, University of Michigan; Samples 5–6: Institute of Mind and Biology, University of Chicago; Sample 7: Behavioral Neuroendocrinology Lab, Simon Fraser University). Prior to assays, frozen saliva samples were thawed, mixed by vortexing, then centrifuged for 15 min to break up and precipitate mucins. For all samples, T was assayed using an enzyme immunoassay kit from Salimetrics; C was measured with an in-house immunoassay using reagents and protocols provided by the University of California at Davis Clinical Endocrinology Laboratory (Samples 1 and 2), or a Salimetrics kit (Samples 3–7). The minimum sensitivity for T was 1 pg/mL; minimum sensitivity for C was 16 pg/mL for the in-house assay, and 70 pg/mL for the Salimetrics assay. Across studies, intra-assay CVs ranged from 3.0% - 9.4% for T (mean: 5.6%), and 4.0% - 8.9% for C (mean: 6.5%). Inter-assay CVs ranged from 5.9% to 10.7% for T (mean: 8.5%) and 5.5% to 12.8% (mean: 8.4%) for C.

Within each sample, all T and C values were corrected for time since awakening by regressing hormone values on a three-term (linear, quadratic, and cubic) polynomial function of passage of time, where all terms including and under the highest order term that was significant were retained. Residual hormone values were saved, and the mean uncorrected value was added to place values on the same scale as the original. For analyses, we averaged all time-corrected baseline T and C values (that is, prior to any laboratory procedures or manipulations) available for an individual to calculate mean T and C.

2.4. Procedures

All studies were largely questionnaire-based. Participants completed psychological and behavioral measures as part of larger questionnaire packets. A substantial proportion of other questionnaires in Samples 2 and 3 concerned participants' behaviors, thoughts, and attitudes directed toward their partners and from their partners toward them. Samples 1 and 2 included questionnaires about general health and fitness. Samples 4–6 included questionnaires regarding personality measures (including autistic-like traits), romantic relationships, sexual behavior, general health, family composition, childhood experiences, and puberty timing. Additional self-reported measures and body and strength measurements were collected in Sample 7. See the published papers cited above for more details.

2.5. Statistical analyses

Within-sample, a specific psychological measure was regressed on T levels, C levels, and the interaction between T and C levels. All levels were log-transformed, in light of expected effects of hormone concentrations that are proportional (vs. absolute) (Jones, 1996). T and C levels were zero-centered within sample. For samples that included both men and women, sex-specific analyses were conducted, given large sex differences in mean and variance of T levels. Each effect – the main effect of T, the main effect of C, and the T × C interaction – was expressed as the partial *r* between the predictor and a given outcome, controlling for the other two effects. To assess the robustness of sex-specific effects of a certain kind, we computed *z*-statistics on overall effects for a given outcome, averaged across all sex-specific samples.

In addition to computing the statistical significance of partial correlations, we performed a Bayesian analysis of interaction effects using *Bayes factors* (see Rouder et al., 2018; Wagenmakers et al., 2017). In contrast with frequentist *p* values, Bayes factors can be used to quantify the strength of the evidence supporting the null hypothesis versus the

alternative. In the present analysis, values larger than 1 indicate evidence in favor of the null hypothesis (in this case, no T × C interaction); values smaller than 1 indicate evidence in favor of the alternative hypothesis (nonzero T × C interaction). Bayesian analyses were performed with JASP 0.8.6 (JASP Team, 2018) using summary statistics from regression models. For the alternative hypothesis, we used a JZS prior with scale factor $r = 0.2$, to account for the fact that interaction effects in observational studies tend to be comparatively small even in the presence of strong interactions (McClelland and Judd, 1993). Bayesian analyses are limited to individual studies, which have lower power than frequentist analyses on overall mean effects. See the SOM for details.

2.5.1. Open data

Data from Samples 1–7 are publicly available at https://osf.io/5aeyt/.

3. Results

3.1. Interactions between T and C levels

Partial correlations and their significance levels are presented in Table 2. We first consider the effects of primary interest: Interactions between T and C. For no outcome did we observe a robust effect for men, women, or overall. The Dual Hormone hypothesis expects interaction effects on status-striving and risk-taking outcomes to be negative. For males, the mean effect size for measures of core traits (with Agreeableness reversed) was slightly positive, 0.06. The mean effect for measures of secondary traits (with future orientation reversed) was slightly negative, −0.01. For females, these values were −0.03 and 0.07, respectively. Overall, effects averaged close to zero, 0.02 in both males and females. For individual traits, no interaction effects were statistically significant at $p < .05$. Just two effects had $p < .10$, and both ran in the positive, not negative, direction: for males, Dominance (0.28); for females, Impulsivity (0.18).

Bayes factors for all the interaction effects are reported in the SOM. With the current choice of priors, the large majority of Bayes factors (29 out of 34 for core traits, 23 out of 26 for secondary traits) indicated weak evidence in favor of the null hypothesis. Specifically, values supporting the null ranged from 1.04 to 1.98 for core traits and from 1.23 to 1.96 for secondary traits; the range was 1.21 to 1.77 in females and 1.04 to 1.98 in males. Robustness checks showed that these results do not depend critically on our choice of scale factor for the prior, and would be qualitatively similar with larger scale factors (up to about $r = 0.5$; see the SOM). There were no instances of moderate or strong support for the null (Bayes factors > 3), which is unsurprising given the small size of individual samples. Smaller samples provide less information about the value of model parameters; as a result, even findings consistent with absence of the hypothesized effects may provide limited support for the null. In two instances, Bayes factors indicated moderate evidence in favor of the alternative hypothesis of a nonzero interaction (0.24 for Winning behaviors in Sample 3 females, and 0.28 for Extraversion in Sample 2 males). However, in both cases the direction of the interaction was positive, rather than negative as predicted by the Dual Hormone hypothesis.

3.2. Main effects

The mean association of T with core measures of status-striving was positive but modest (partial $r = 0.07$ and 0.09 for men and women, respectively). In overall analyses, T predicted intrasexual competitive winning and partner-reported non-submissiveness (only at $p < .10$), both in a positive direction. For secondary measures, mean associations with T were near zero (partial $r = 0.04$ and −0.01 for men and women). In men, T was positively associated with extraversion (a secondary trait in this context). The mean association of outcomes with C was near-zero. In overall analyses, C was not associated with any of our measures. See Table 2.

4. Discussion

Across seven independent samples of adults, totaling 718 participants (median sample size per analysis = 318), we tested whether T, C, or the interaction of the two hormones predicted a variety of self- and partner-reported personality measures relating to status-striving or risk-taking. Our aggregate analysis of all seven studies did not yield a robust T × C interaction effect, with the overall effect going slightly in the opposite direction of that predicted by the Dual Hormone hypothesis. We followed up our aggregate analysis by performing Bayesian analyses of interaction effects in individual samples. Most of these analyses indicated weak evidence in favor of the null hypothesis. We acknowledge upfront, however, that several of these findings on individual traits were only able to provide weak evidence due to small sample sizes.

T levels modestly but non-significantly related to core measures of status-striving (mean partial $r = 0.07$); secondary measures were even more weakly related (mean partial $r = 0.03$). C levels did not reliably predict psychological traits and behavioral variables.

4.1. Limitations and possible interpretations

The proper interpretation of null findings is often far from straightforward. Here we review the limitations of the present study, and address possible interpretations of results. Failure to detect an effect can stem from a number of sources. In our view, failure to detect an overall, substantively meaningful mean negative interaction effect that truly exists is unlikely. The mean interaction effect aggregated over multiple measures and across subsamples (mathematically equivalent to a meta-analytic estimate of the average effect) actually runs in a positive direction, meaning the average relationship between T and our outcomes became more positive as C increased (though this effect fell well short of statistical significance). At the level of individual traits, our power to detect a true interaction effect varies, and some analyses are based on small samples. Thus, we emphasize our aggregate results as the most informative findings pertaining to the Dual Hormone hypothesis, while fully acknowledging that our results can say much less about results regarding individual traits. As our power analysis below demonstrates, the aggregate sample contains excellent power to detect a range of plausible effect sizes.

Another possibility is a problem with our operationalization of the Dual Hormone hypothesis. In particular, our results may arise from the predominant use of self-report measures. This possibility is consistent with one perspective advanced for main effects of T, in which T impacts implicit status-related motivations that are not necessarily connected to explicit self-reports (see e.g. Schultheiss et al., 1999; Schultheiss and Rohde, 2002; Terburg and van Honk, 2013). Such effects on motivation and behavior may be detectable by others (e.g., partners, acquaintances) or via experimental manipulation (e.g. the presentation of angry faces; Terburg et al., 2012) while remaining opaque to the affected person. The first point in response is that, as Table 1 demonstrates, varying operationalizations of 'status-striving' or 'dominance' have already been employed in the Dual Hormone literature, including self-report measures very similar to the variables we assess (e.g., self-reports of externalizing behaviors [Tackett et al., 2014; Grotzinger et al., 2018] are strongly predicted by reversed agreeableness (Miller et al., 2008), one measure used in our analyses). Sixteen of the 55 previously published effects (29%) listed in Table 1 rely on self-report measures, including five of the 20 statistically significant effects (a near-identical 25%). Based on these frequencies, in conjunction with theoretical arguments regarding the substantial links between personality traits and overt behaviors (e.g. Back and Vazire, 2015), we do not believe that our analyses are systematically less representative of the Dual Hormone hypothesis than many other previously published tests. Second, to

Table 2
Effects of testosterone (log), cortisol (log), and testosterone (log) × cortisol (log) interactions: partial *r*. SIC: Scale of Intersexual Competitiveness; EIQ: Eyesenck Impulsiveness Questionnaire; ZTPI: Zimbardo Time Perspective Inventory.

Sample		S1	S2	S2	S3	S3	S4	S4	S5	S6	S6	S7	Weighted means			
Sex		M	M	F	M	F	M	F	M	M	F	M	M	F	All	*Trait N*
Core traits																
Social potency	T	0.15	0.18	−0.15	−0.08	0.24							0.10	0.03	0.07	
	C	−0.04	−0.07	0.10	0.12	0.09							−0.01	−0.01	−0.01	
	T × C	0.07	0.20	−0.08	0.00	0.06							0.10	−0.01	0.05	325
Non-submissiveness	T	0.05	0.08	0.09	0.00	0.22							0.05	0.15	0.09	
	C	−0.03	0.00	−0.05	0.03	0.22							0.00	0.07	0.02	
	T × C	0.05	0.08	0.18	0.04	0.05							0.06	0.13	0.09	323
Social potency (partner-report)	T				0.08	0.24							0.08	0.24	0.16	
	C				−0.03	0.00							−0.03	0.00	−0.02	
	T × C				0.08	−0.17							0.18	−0.17	−0.04	88
Non-submissiveness (partner-report)	T				0.23	0.14							0.23	0.14	*0.19*	
	C				−0.18	0.17							−0.18	0.17	0.00	
	T × C				0.14	−0.17							0.14	−0.17	−0.01	88
Win Intrasex Competition	T		0.05	−0.10	*0.30*	**0.30**							*0.16*	0.08	**0.12**	
	C		−0.09	−0.19	−0.05	0.10							0.03	−0.06	−0.02	
	T × C		0.15	−0.17	−0.05	0.37							0.06	0.07	0.07	233
SIC	T						0.12	0.09	−0.03				0.03	0.09	0.05	
	C						−0.06	−0.02	0.06				−0.01	−0.02	0.00	
	T × C						−0.13	−0.15	−0.06				−0.09	−0.15	−0.11	196
Dominance	T											0.11	0.11	–	0.11	
	C											−0.25	−0.25	–	−0.25	
	T × C											*0.28*	*0.28*	–	*0.28*	40
Prestige	T											−0.13	−0.13	–	−0.13	
	C											0.09	0.09	–	0.09	
	T × C											−0.04	−0.04	–	−0.04	40
Agreeableness (reversed)	T	0.00	−0.13	0.07			−0.09	−0.09	−0.08				−0.07	0.00	−0.05	
	C	0.06	0.16	−0.16			0.17	−0.05	0.05				*0.10*	−0.11	0.04	
	T × C	0.11	−0.05	0.00			−0.02	0.03	−0.03				0.01	0.01	0.01	436
EIQ Venturesome	T						0.12	−0.04	0.09	0.12	0.05		0.11	−0.01	0.07	
	C						0.05	−0.08	−0.11	−0.05	−0.06		−0.05	−0.01	−0.03	
	T × C						−0.08	0.11	0.00	−0.09	−0.04		−0.05	0.03	−0.02	318
MEAN	T												0.07	0.09	0.07	
	C												−0.03	0.01	−0.02	
	T × C												0.06	−0.03	0.02	
Secondary traits																
Extraversion	T	0.14	**0.36**	−0.13			0.13	−0.12	0.02				**0.16**	−0.13	0.07	
	C	0.16	−0.10	0.00			0.12	0.02	0.08				0.07	0.01	0.05	
	T × C	−0.16	**0.28**	−0.06			0.16	0.11	−0.10				0.03	0.01	0.02	436
EIQ impulsivity	T						−0.05	−0.04	−0.13	**0**.04	−0.10		−0.05	−0.07	−0.06	
	C						0.21	0.12	0.02	0.08	−0.03		0.09	0.04	0.07	
	T × C						−0.02	0.09	−0.02	−0.11	*0.25*		−0.05	*0.18*	0.03	318
ZTPI present-hedonistic	T						0.02	−0.18	0.04	*0.23*	0.05		0.09	−0.06	0.04	
	C						0.09	**0.35**	−0.02	0.02	0.00		0.02	*0.16*	*0.08*	
	T × C						−0.04	0.16	−0.01	0.07	0.11		0.01	0.14	0.05	318
ZTPI present-fatalistic	T						0.03	**0**.07	−0.08	0.08	0.15		0.00	0.11	0.04	
	C						0.05	*0.29*	0.01	0.03	0.04		0.03	0.16	0.07	
	T × C						−0.06	−0.13	−0.06	0.11	0.08		−0.01	−0.02	−0.01	318
ZTPI future (reversed)	T						−0.07	0.10	0.02	0.08	0.08		0.01	0.08	0.04	
	C						0.01	0.17	−0.05	0.00	−0.02		−0.02	0.07	0.01	
	T × C						−0.06	0.05	−0.06	0.09	0.06		−0.01	0.05	0.01	318
MEAN	T												0.04	−0.01	0.03	
	C												0.04	0.09	0.06	
	T × C												−0.01	0.07	0.02	
Median *N*		97	71	70	44	45	46	45	107	62	58	40				

Notes. All effects expressed as partial *r*, with the other two effects controlled. All effects significant at $p < .05$ are in bold. All effects with $p < .10$ are in italics. *p*-Values not corrected for multiple comparisons.

empirically address differences between self-report and behavioral effect sizes, we compared the right skew of the *p*-curves (as an index for evidential value; see Simonsohn et al., 2014) for the self-report effects to those not relying on self-reports. The effects available to us do not suggest a significant difference in right skew ($p = .193$, in the direction of more right-skew for the behavioral effects; see S2 and S3 in SOM for separate *p*-curves of self-report and behavioral outcomes). We fully acknowledge that, with only five eligible self-report effects, and 15 eligible behavioral effects, this analysis has limited power to detect differences. We hence do not argue that it shows evidence for no or minimal differences across outcome type; we merely note that the analysis does not offer clear evidence *for* them. Third, self-reports of some of our measures (Social Potency, Non-Submissiveness) were paired with reports by partners, which showed substantial "self-other agreement" (self-partner correlations averaged 0.52 and 0.41 for men and women, respectively); clearly, these correlations are not possible if individuals do not behave in ways both tapped by the self-report measure and observable to others. The mean overall interaction effect sizes for partner reports specifically were close to zero as well (−0.04 and −0.01 for Social Potency and Non-Submissiveness, respectively—and more positive [hence, in the unpredicted direction] for men, for whom T may be measured more validly (Shirtcliff et al., 2002;

Crewther et al., 2018), than for females [mean partial $r = 0.16$ and -0.17, respectively). Effects for self-reported instances of specific behaviors (winning competitions) in the past two days ran slightly in a positive direction on average (0.07). All in all, then, we see little evidence that results vary as a function of outcome type, which is not to deny the possibility that they do. We note that this conclusion is supported by a new meta-analysis of the Dual Hormone hypothesis (Dekkers et al., 2019), which does not find a significant difference in effect sizes comparing self-report to behavioral tasks.

Our results relatedly leave open the possibility that some effects of T (and C) occur predominantly (or even exclusively) at an implicit level, and require specific experimental procedures to be reliably detected. This perspective is consistent with findings showing associations between hormone concentrations and behaviour when controlling for self-reports of personality (e.g., Geniole et al., 2013; Akinola et al., 2016). However, we note that studies supporting the idea that T affects automatic processes outside of conscious awareness (e.g., gaze aversion, early threat processing) are typically based on small samples, and the relevant effects often emerge from 2-way or 3-way interaction tests (e.g., Terburg et al., 2012; van Honk et al., 2005; van Peer et al., 2017; Welling et al., 2016; Wirth and Schultheiss, 2007). Some of the same methodological issues that may have led to spurious and/or inflated findings in the Dual Hormone literature also apply to the literature on implicit motivation. While we do not intend to dismiss the existing findings on implicit motivation, we argue that those findings should be submitted to critical scrutiny before they can be accepted with confidence.

Another limitation of the present study is that we assayed T and C levels on either one day or two different days in our studies, which could result in levels that are not broadly representative of mean hormone levels produced by the participant. However, many previous studies examining associations with personality measures assessed hormone levels on just one day (e.g., Mehta and Josephs, 2010; Mehta et al., 2015a, 2015b; Ponzi et al., 2016; Zilioli et al., 2015). One recent study found a 'conventional' negative T × C interaction on reports of aggressive behavior using hair samples, representing a longer time interval of hormone assessment (Grotzinger et al., 2018). This preliminary evidence raises the possibility that more sensitive and/or aggregated measurements of steroid hormones will more reliably yield interactions, but further research is needed to test this speculation.

Finally, one might point to challenges to the validity of ELISA procedures as a potential limitation of our study (see Welker et al., 2016). Some researchers (e.g., Welker et al., 2016; Schultheiss et al., 2018) argue that a superior method for measuring salivary T concentrations is mass spectrometry (two recent papers in our review have adopted this approach: Grotzinger et al., 2018; Ronay et al., 2018). We welcome the use of more precise methods of hormone measurement but, unlike some scholars who believe the field needs to jettison ELISAs and "go back to square one" (Mazur and Clifton, 2018, p. 8) with mass spectrometry, we take a somewhat different stance. Evidence from several sources reports strong and significant correspondence between salivary T assessed via ELISA and mass spectrometry ($r = 0.808$, $p < .001$, $N = 40$ males, Yasuda et al., 2008), between salivary T and serum T determined via ELISA (total T: $r = 0.96$, $p < .001$, $N = 28$ males and females, Salimetrics LLC, State College, PA; total T: $r = 0.916$, $N = 45$ males, Crewther et al., 2018), and among results from the three major manufacturers of salivary T ELISA kits ($0.774 < r < 0.921$, $N = 50$ males and females, Andersson et al., 2017). Studies have reported poorer performance of salivary assays for detecting salivary T in women due to poor resolution of low concentrations (Shirtcliff et al., 2002; Crewther et al., 2018), but this issue is not specific to ELISAs. Limiting our results to men, correlations with core outcomes specifically ran somewhat in the unpredicted direction (results for partner-reports in particular also ran in the wrong direction). Mass spectrometry undeniably possesses some advantages over ELISA that warrants its expanded use, but at the same time, the cost-effectiveness and performance of ELISAs makes them viable methods for continued use in behavioral endocrinology studies.

4.2. Concluding thoughts and recommendations

The strongest interpretation of the present results—i.e., that the hypothesis of T × C interactions predicting human status-striving should be rejected—is clearly premature, in our view. In addition to the aforementioned limitations of our dataset, at a physiological level, an empirical rationale for interactions between T and C is reasonable, and these interactions plausibly affect downstream behavior. The Dual Hormone literature has pointed to evidence of inhibitory physiological interactions (e.g., Viau, 2002; Tilbrook et al., 2000), which may well manifest in C blocking the effects of T on dominance behavior, in line with the traditional Dual Hormone prediction. At the same time, there are also theoretical and physiological reasons to expect *positive* interactions. For instance, Schoech et al. (1999) show that, in dark-eyed juncos, males treated with T exhibit baseline C and larger C responses to stressors than do controls. This comports with a physiological explanation the authors propose: increases in mating effort (e.g., expanding home ranges, singing more frequently) facilitated by T exact substantial energetic costs. The well-known effects of C on the mobilization of energy stores, they propose, is one important way males 'pay for' the costs of T-mediated behaviors. Behavioral endocrinology work in chimpanzees (e.g., Muller and Wrangham, 2004) and red-fronted lemurs (Ostner et al., 2008) pose a comparable argument regarding C and dominance in primates: successful investments by males in mating effort and status striving must occur within a physiological environment of enhanced stress. Lastly, Table 1 provides examples of positive T × C interactions on several indices of status-striving or dominance in humans. Thus, the inconsistent pattern of interactions in the literature, in conjunction with the null effects we report, may indicate the presence of *contingent* T × C interactions that influence status-seeking behavior or traits, rather than an absence of meaningful interactions.

Perhaps the clearest implication of our results, then, is the need for further theoretical and conceptual refinement. We strongly advocate for the development of precise predictions regarding which behaviors and traits should be subject to negative T × C interactions, which are expected to be subject to positive interactions, and perhaps which are not expected to yield significant interaction effects. Several studies have also reported moderated Dual Hormone effects (e.g. 3-way T × C × personality feature interactions; see Tackett et al., 2014 and Pfattheicher, 2017; T × C × experimental condition interactions; see Mehta and Josephs, 2010, Geniole et al., 2011, and Henry et al., 2017); we similarly encourage the development of precise predictions regarding the cases in which these 3-way interactions should emerge. We hope that our results aid in the development of these predictions, both by highlighting some effects that do not appear to be robust, and by encouraging other labs to open their file drawers, which may contain replication attempts of Dual Hormone predictions with outcome measures unavailable to us—for instance, regarding overt behaviors. As we note above, *p*-curves of the published literature do not provide evidence of significantly stronger effects for behavioral outcomes, but the addition of file-drawered effects may modify this preliminary conclusion, and thus advance debates regarding the use of implicit versus explicit measures in Dual Hormone studies.

Going forward, we encourage pre-registration of Dual Hormone predictions before they are tested empirically, in light of recent discussion regarding analytical flexibility in psychological science (e.g., the 'garden of forking paths'; Gelman and Loken, 2014). High-powered, pre-registered replication studies will serve as one of the most valuable tools to definitively address the robustness and boundary conditions of the Dual Hormone hypothesis. This call is echoed by the authors of a new meta-analysis on the Dual Hormone Hypothesis (Dekkers et al., 2019).

While the median sample size in our analyses (318) is larger than

many Dual Hormone studies, simulations suggest that even larger samples may be necessary to achieve acceptable statistical power, particularly to detect interaction effects (see, e.g., Gelman, 2018). If T × C interactions account for an additional 2% of variance explained in a linear regression, for instance, then sample sizes of ~400 are needed to achieve 80% statistical power. Raising the effect size to 4% still requires ~200 participants for 80% power.[6] Future research may also benefit from study designs that measure T and C at multiple time points, either via collecting multiple 'baseline' measurements, or capturing short-term hormonal changes. Single, instantaneous hormone concentrations—by far the most common measurement type in the Dual Hormone literature—may be affected by a number of extraneous factors (e.g., recent aerobic exercise or sexual activity), which contributes noise and increases the chances of both Type I and Type II errors (see Loken and Gelman, 2017). Another potential way forward may stem from experimental manipulation of T. Recent work administering exogenous T to human participants suggests effects on aggression (e.g., Carré et al., 2017; Wagels et al., 2018) and threat processing (e.g., van Peer et al., 2017); perhaps the effects of these manipulations can be compared across a gradient of naturally occurring C concentrations (e.g., corresponding to time since waking). In sum, we believe that a triangulation effort, involving numerous techniques and approaches, will be necessary to clarify the nature of Dual Hormone effects on human behavior.

Competing interests

We have no competing interests to declare.

References

Akinola, M., Page-Gould, E., Mehta, P.H., Lu, J.G., 2016. Collective hormonal profiles predict group performance. Proc. Natl. Acad. Sci. 113, 9774–9779.

Andersson, C.R., Bergquist, J., Theodorsson, E., Ström, J.O., 2017. Comparisons between commercial salivary testosterone enzyme-linked immunosorbent assay kits. Scand. J. Clin. Lab. Invest. 77, 582–586.

Archer, J., 2006. Testosterone and human aggression: an evaluation of the challenge hypothesis. Neurosci. Biobehav. Rev. 30 (3), 319–345.

Archer, J., Birring, S.S., Wu, F.C., 1998. The association between testosterone and aggression among young men: empirical findings and a meta-analysis. Aggress. Behav. Off. J. Int. Soc. Res. Aggress. 24 (6), 411–420.

Archer, J., Graham-Kevan, N., Davies, M., 2005. Testosterone and aggression: a reanalysis of book, Starzyk, and Quinsey's (2001) study. Aggress. Violent Behav. 10 (2), 241–261.

Astheimer, L.B., Buttemer, W.A., Wingfield, J.C., 2000. Corticosterone treatment has no effect on reproductive hormones or aggressive behavior in free-living male tree sparrows, Spizella arborea. Horm. Behav. 371, 31–39.

Back, M.D., Vazire, S., 2015. The social consequences of personality: six suggestions for future research. Eur. J. Personal. 29, 296–307.

Ball, G.F., Castelino, C.B., Maney, D.L., Appeltants, D., Balthazart, J., 2003. The activation of birdsong by testosterone. Ann. N. Y. Acad. Sci. 100, 211–231.

Barel, E., Shahrabani, S., Tzischinsky, O., 2017. Sex hormone/cortisol ratios differentially modulate risk-taking in men and women. Evol. Psychol. 151 (1474704917697333).

Bateup, H.S., Booth, A., Shirtcliff, E.A., Granger, D.A., 2002. Testosterone, cortisol, and women's competition. Evol. Hum. Behav. 23 (3), 181–192.

Benning, S.D., Patrick, C.J., Hicks, B.M., Blonigen, D.M., Krueger, R.F., 2003. Factor structure of the psychopathic personality inventory: validity and implications for clinical assessment. Psychol. Assess. 15 (3), 340.

Bobadilla, L., Asberg, K., Johnson, M., Shirtcliff, E.A., 2015. Experiences in the military may impact dual-axis neuroendocrine processes in veterans. Dev. Psychobiol. 576, 719–730.

Buades-Rotger, M., Engelke, C., Beyer, F., Keevil, B.G., Brabant, G., Krämer, U.M., 2016. Endogenous testosterone is associated with lower amygdala reactivity to angry faces and reduced aggressive behavior in healthy young women. Sci. Rep. 6, 38538.

Buunk, A.P., Fisher, M., 2009. Individual differences in intrasexual competition. J. Evol. Psychol. 7, 37–48.

Carré, J.M., Mehta, P.H., 2011. Importance of considering testosterone–cortisol interactions in predicting human aggression and dominance. Aggress. Behav. 376, 489–491.

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

Caspi, A., Begg, D., Dickson, N., Harrington, H., Langley, J., Moffitt, T.E., Silva, P.A., 1997. Personality differences predict health-risk behaviors in young adulthood: evidence from a longitudinal study. J. Pers. Soc. Psychol. 73 (5), 1052.

Casto, K.V., Edwards, D.A., 2016. Testosterone, cortisol, and human competition. Horm. Behav. 82, 21–37.

Cheng, J.T., Tracy, J.L., Henrich, J., 2010. Pride, personality, and the evolutionary foundations of human social status. Evol. Hum. Behav. 31, 334–347.

Connelly, B.S., Ones, D.S., 2010. An other perspective on personality: meta-analytic integration of observers' accuracy and predictive validity. Psychol. Bull. 136, 1092–1122.

Crewther, B.T., Obmiński, Z., Orysiak, J., Al-Dujaili, E.A., 2018. The utility of salivary testosterone and cortisol concentration measures for assessing the stress responses of junior athletes during a sporting competition. J. Clin. Lab. Anal. 32, e22197.

Cueva, C., Roberts, R.E., Spencer, T., Rani, N., Tempest, M., Tobler, P.N., Herbert, J., Rustichini, A., 2015. Cortisol and testosterone increase financial risk taking and may destabilize markets. Sci. Rep. 5, 11206.

Dabbs, J.M., Dabbs, M.G., 2000. Heroes, Rogues, and Lovers: Testosterone and Behavior. McGraw-Hill, New York.

Dabbs, J.M., Jurkovic, G.J., Frady, R.L., 1991. Salivary testosterone and cortisol among late adolescent male offenders. J. Abnorm. Child Psychol. 19 (4), 469–478.

D'Alessio, M., Guarino, A., De Pascalis, V., Zimbardo, P.G., 2003. Testing Zimbardo's Stanford time perspective inventory STPI-short form. Time Soc. 12, 333–347.

Dekkers, T.J., van Rentergem, J.A.A., Meijer, B., Popma, A., Wagemaker, E., Huizenga, H.M., 2019. A meta-analytical evaluation of the dual-hormone hypothesis: does cortisol moderate the relationship between testosterone and status, dominance, risk taking, aggression, and psychopathy? Neurosci. Biobehav. Rev. 96, 250–271.

Del Giudice, M., Klimczuk, A.C., Traficonte, D.M., Maestripieri, D., 2014. Autistic-like and schizotypal traits in a life history perspective: diametrical associations with impulsivity, sensation seeking, and sociosexual behavior. Evol. Hum. Behav. 35, 415–424.

Denson, T.F., Mehta, P.H., Tan, D.H., 2013. Endogenous testosterone and cortisol jointly influence reactive aggression in women. Psychoneuroendocrinology 38, 416–424.

Dixson, A., 1998. Primate Sexuality. John Wiley Sons, Ltd.

Edwards, D.A., Casto, K.V., 2013. Women's intercollegiate athletic competition: cortisol, testosterone, and the dual-hormone hypothesis as it relates to status among teammates. Horm. Behav. 64 (1), 153–160.

Eisenegger, C., Haushofer, J., Fehr, E., 2011. The role of testosterone in social interaction. Trends Cogn. Sci. 15, 263–271.

Eysenck, S.B.G., Eysenck, H.J., 1978. Impulsiveness and venturesomeness: their position in a dimensional system of personality description. Psychol. Rep. 43, 1247–1255.

Eysenck, S.B., Easting, G., Pearson, P.R., 1984. Age norms for impulsiveness, venturesomeness and empathy in children. Personal. Individ. Differ. 5, 315–321.

Gangestad, S.W., Thornhill, R., Garver-Apgar, C.E., 2005. Adaptations to ovulation: implications for sexual and social behavior. Curr. Dir. Psychol. Sci. 14, 312–316.

Gangestad, S.W., Merriman, L.A., Thompson, M.E., 2010a. Men's oxidative stress, fluctuating asymmetry and physical attractiveness. Anim. Behav. 80, 1005–1013.

Gangestad, S.W., Thornhill, R., Garver-Apgar, C.E., 2010b. Men's facial masculinity predicts changes in their female partners' sexual interests across the ovulatory cycle, whereas men's intelligence does not. Evol. Hum. Behav. 31, 412–424.

Gangestad, S.W., Garver-Apgar, C.E., Cousins, A.J., Thornhill, R., 2014. Intersexual conflict across women's ovulatory cycle. Evol. Hum. Behav. 35, 302–308.

Garver-Apgar, C.E., Gangestad, S.W., Thornhill, R., Miller, R.D., Olp, J.J., 2006. Major histocompatibility complex alleles, sexual responsivity, and unfaithfulness in romantic couples. Psychol. Sci. 17, 830–835.

Gelman, A., 2018, March 15. You Need 16 Times the Sample Size to Estimate an Interaction Than to Estimate a Main Effect [Blog Post]. Available from: https://andrewgelman.com/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/.

Gelman, A., Loken, E., 2014. Data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. Am. Sci. 10, 460.

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. Biol. Psychol. 87, 137–145.

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. Horm. Behav. 64, 790–798.

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. Horm. Behav. 92, 37–50.

Gimpl, G., Fahrenholz, F., 2001. The oxytocin receptor system: structure, function, and regulation. Physiol. Rev. 81, 629–683.

Glenn, A.L., Raine, A., Schug, R.A., Gao, Y., Granger, D.A., 2011. Increased testosterone-to-cortisol ratio in psychopathy. J. Abnorm. Psychol. 120 (2), 389.

Graziano, W.G., Habashi, M.M., Sheese, B.E., Tobin, R.M., 2007. Agreeableness, empathy, and helping: a person × situation perspective. J. Pers. Soc. Psychol. 93, 583.

Grebe, N.M., Kristoffersen, A.A., Grøntvedt, T.V., Thompson, M.E., Kennair, L.E.O., Gangestad, S.W., 2017. Oxytocin and vulnerable romantic relationships. Horm. Behav. 90, 64–74.

Grotzinger, A.D., Mann, F.D., Patterson, M.W., Tackett, J.L., Tucker-Drob, E.M., Harden, K.P., 2018. Hair and salivary testosterone, hair cortisol, and externalizing behaviors

[6] Analysis performed in G*Power 3.1.

in adolescents. Psychol. Sci. 29, 688–699.
Handa, R.J., Weiser, M.J., 2014. Gonadal steroid hormones and the hypothalamo–pituitary–adrenal axis. Front. Neuroendocrinol. 35, 197–220.
Henry, A., Sattizahn, J.R., Norman, G.J., Beilock, S.L., Maestripieri, D., 2017. Performance during competition and competition outcome in relation to testosterone and cortisol among women. Horm. Behav. 92, 82–92.
Hung, H.M., O'Neill, R.T., Bauer, P., Köhne, K., 1997. The behavior of the P-value when the alternative hypothesis is true. Biometrics 53, 11–22.
Ioannidis, J.P., 2008. Why most discovered true associations are inflated. Epidemiology 19, 640–648.
JASP Team, 2018. JASP (Version 0.8.6) [Computer Software]. https://jasp-stats.org.
Johnson, E.O., Kamilaris, T.C., Chrousos, G.P., Gold, P.W., 1992. Mechanisms of stress: a dynamic overview of hormonal and behavioral homeostasis. Neurosci. Biobehav. Rev. 16, 115–130.
Jones, K.A., 1996. Summation of basic endocrine data. In: Gass, G.A., Kaplan, H.M. (Eds.), Handbook of Endocrinology, second edition. vol. 1. CRC Press, Boca Raton FL, pp. 2–42.
Josephs, R.A., Sellers, J.G., Newman, M.L., Mehta, P.H., 2006. The mismatch effect: when testosterone and status are at odds. J. Pers. Soc. Psychol. 90, 999–1013.
Ketterson, E.D., Nolan Jr., V., 1999. Adaptation, exaptation, and constraint: a hormonal perspective. Am. Nat. 154, S4–S25.
Knight, E.L., Christian, C.B., Morales, P.J., Harbaugh, W.T., Mayr, U., Mehta, P.H., 2017. Exogenous testosterone enhances cortisol and affective responses to social-evaluative stress in dominant men. Psychoneuroendocrinology 85, 151–157.
Kurath, J., Mata, R., 2018. Individual differences in risk taking and endogeneous levels of testosterone, estradiol, and cortisol: a systematic literature search and three independent meta-analyses. Neurosci. Biobehav. Rev.
Lane, A., Luminet, O., Nave, G., Mikolajczak, M., 2016. Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the file drawer of one laboratory. J. Neuroendocrinol. 28.
Lee, J.J., Gino, F., Jin, E.S., Rice, L.K., Josephs, R.A., 2015. Hormones and ethics: understanding the biological basis of unethical conduct. J. Exp. Psychol. Gen. 144, 891.
Lejuez, C.W., Read, J.P., Kahler, C.W., Richards, J.B., Ramsey, S.E., Stuart, G.L., Brown, R.A., 2002. Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). J. Exp. Psychol. Appl. 8, 75.
Loken, E., Gelman, A., 2017. Measurement error and the replication crisis. Science 355, 584–585.
Lozza, N., Spoerri, C., Ehlert, U., Hubmann, P., Kesselring, M., Farahmand, F., Sollberger, S., La Marca, R., 2017. Predicting social behavior: basal and dynamic joint effects of testosterone and cortisol. Adap. Hum. Behav. Physiol. 3 (3), 255–274.
Maestripieri, D., Klimczuk, A.C., Seneczko, M., Traficonte, D.M., Wilson, M.C., 2013. Relationship status and relationship instability, but not dominance, predict individual differences in baseline cortisol levels. PLoS One 8 (12), e84003.
Maestripieri, D., Klimczuk, A., Traficonte, D., Wilson, C., 2014. A greater decline in female facial attractiveness during middle age reflects women's loss of reproductive value. Front. Psychol. 5, 179.
Mann, T., Spellman, B.A., 2016, September. Opening our own file drawers: issues in creating a more complete and useful psychological science. In: Symposium conducted at the meeting of the Society of Experimental Social Psychology, Santa Monica, CA.
Marvel-Coen, J., Nickels, N., Maestripieri, D., 2018. The relationship between morningness-eveningness, psychosocial variables, and cortisol reactivity to stress from a life history perspective. Evol. Behav. Sci. 12, 71–86.
Mazur, A., Booth, A., 1998. Testosterone and dominance in men. Behav. Brain Sci. 21 (3), 353–363.
Mazur, A., Booth, A., 2014. Testosterone is related to deviance in male army veterans, but relationships are not moderated by cortisol. Biol. Psychol. 96, 72–76.
Mazur, A., Clifton, S., 2018. Enzyme immunoassay may be inadequate for measuring salivary testosterone in older men. Aging Male 1–9.
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.
McClelland, G.H., Judd, C.M., 1993. Statistical difficulties of detecting interactions and moderator effects. Psychol. Bull. 114, 376–390.
McCrae, R.R., Costa, P.T., 2004. A contemplated revision of the NEO Five-Factor Inventory. Personal. Individ. Differ. 36, 587–596.
Mehta, P.H., Josephs, R.A., 2010. Testosterone and cortisol jointly regulate dominance: evidence for a dual-hormone hypothesis. Horm. Behav. 58, 898–906.
Mehta, P.H., Prasad, S., 2015. The dual-hormone hypothesis: a brief review and future research agenda. Curr. Opin. Behav. Sci. 3, 163–168.
Mehta, P.H., Mor, S., Yap, A.J., Prasad, S., 2015a. Dual-hormone changes are related to bargaining performance. Psychol. Sci. 26, 866–876.
Mehta, P.H., Welker, K.M., Zilioli, S., Carré, J.M., 2015b. Testosterone and cortisol jointly modulate risk-taking. Psychoneuroendocrinology 56, 88–99.
Mehta, P.H., DesJardins, N.M.L., van Vugt, M., Josephs, R.A., 2017. Hormonal underpinnings of status conflict: testosterone and cortisol are related to decisions and satisfaction in the hawk-dove game. Horm. Behav. 92, 141–154.
Miller, J.D., Lynam, D.R., Jones, S., 2008. Externalizing behavior through the lens of the five-factor model: a focus on agreeableness and conscientiousness. J. Pers. Assess. 90, 158–164.
Montoya, E.R., Terburg, D., Bos, P.A., Van Honk, J., 2012. Testosterone, cortisol, and serotonin as key regulators of social aggression: a review and theoretical perspective. Motiv. Emot. 36, 65–73.
Moore, F.L., Miller, L.J., 1984. Stress-induced inhibition of sexual behavior: corticosterone inhibits courtship behaviors of a male amphibian Taricha granulosa. Horm. Behav. 18, 400–410.
Muller, M.N., Wrangham, R.W., 2004. Dominance, cortisol and stress in wild chimpanzees Pan troglodytes schweinfurthii. Behav. Ecol. Sociobiol. 55, 332–340.
Neel, R., Kenrick, D.T., White, A.E., Neuberg, S.L., 2016. Individual differences in fundamental social motives. J. Pers. Soc. Psychol. 110 (6), 887.
Ostner, J., Kappeler, P., Heistermann, M., 2008. Androgen and glucocorticoid levels reflect seasonally occurring social challenges in male red-fronted lemurs Eulemur fulvus rufus. Behav. Ecol. Sociobiol. 62, 627–638.
Pfattheicher, S., 2016. Testosterone, cortisol and the Dark Triad: narcissism but not Machiavellianism or psychopathy is positively related to basal testosterone and cortisol. Personal. Individ. Differ. 97, 115–119.
Pfattheicher, S., 2017. Illuminating the dual-hormone hypothesis: about chronic dominance and the interaction of cortisol and testosterone. Aggress. Behav. 43, 85–92.
Pfattheicher, S., Landhäußer, A., Keller, J., 2014. Individual differences in antisocial punishment in public goods situations: the interplay of cortisol with testosterone and dominance. J. Behav. Decis. Mak. 27, 340–348.
Platje, E., Popma, A., Vermeiren, R.R., Doreleijers, T.A., Meeus, W.H., van Lier, P.A., Koot, H.M., Branje, S.J.T., Jansen, L., 2015. Testosterone and cortisol in relation to aggression in a non-clinical sample of boys and girls. Aggress. Behav. 41, 478–487.
Ponzi, D., Zilioli, S., Mehta, P.H., Maslov, A., Watson, N.V., 2016. Social network centrality and hormones: the interaction of testosterone and cortisol. Psychoneuroendocrinology 68, 6–13.
Popma, A., Vermeiren, R., Geluk, C.A., Rinne, T., van den Brink, W., Knol, D.L., Jansen, L.M.C., van Engeland, H., Doreleijers, T.A., 2007. Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. Biol. Psychiatry 61, 405–411.
Prasad, S., Narayanan, J., Lim, V.K., Koh, G.C., Koh, D.S., Mehta, P.H., 2017. Preliminary evidence that acute stress moderates basal testosterone's association with retaliatory behavior. Horm. Behav. 92, 128–140.
Retana-Marquez, S., Bonilla-Jaime, H., Vazquez-Palacios, G., Martinez-Garcia, R., Velazquez-Moctezuma, J., 2003. Changes in masculine sexual behavior, corticosterone and testosterone in response to acute and chronic stress in male rats. Horm. Behav. 44, 327–337.
Ronay, R., Carney, D.R., 2013. Testosterone's negative relationship with empathic accuracy and perceived leadership ability. Soc. Psychol. Personal. Sci. 4, 92–99.
Ronay, R., van der Meij, L., Oostrom, J.K., Pollet, T.V., 2018. No evidence for a relationship between hair testosterone concentrations and 2D: 4D ratio or risk taking. Front. Behav. Neurosci. 12, 30.
Rouder, J.N., Haaf, J.M., Vandekerckhove, J., 2018. Bayesian inference for psychology, part IV: parameter estimation and Bayes factors. Psychon. Bull. Rev. https://doi.org/10.3758/s13423-017-1420-7.
Rubinow, D.R., Roca, C.A., Schmidt, P.J., Danaceau, M.A., Putnam, K., Cizza, G., Chrousos, G., Nieman, L., 2005. Testosterone suppression of CRH-stimulated cortisol in men. Neuropsychopharmacology 30, 1906–1912.
Salvador, A., Suay, F., Martinez-Sanchis, S., Simon, V.M., Brain, P.F., 1999. Correlating testosterone and fighting in male participants in judo contests. Physiol. Behav. 68, 205–209.
Sapolsky, R.M., 2005. The influence of social hierarchy on primate health. Science 308, 648–652.
Scerbo, A.S., Kolko, D.J., 1994. Salivary testosterone and cortisol in disruptive children: relationship to aggressive, hyperactive, and internalizing behaviors. J. Am. Acad. Child Adolesc. Psychiatry 33, 1174–1184.
Schoech, S.J., Ketterson, E.D., Nolan Jr., V., 1999. Exogenous testosterone and the adrenocortical response in dark-eyed juncos. Auk 116, 64–72.
Schultheiss, O.C., Rohde, W., 2002. Implicit power motivation predicts men's testosterone changes and implicit learning in a contest situation. Horm. Behav. 41, 195–202.
Schultheiss, O.C., Campbell, K.L., McClelland, D.C., 1999. Implicit power motivation moderates men's testosterone responses to imagined and real dominance success. Horm. Behav. 36, 234–241.
Schultheiss, O.C., Dlugash, G., Mehta, P.H., 2018. Hormone measurement in social neuroendocrinology: a comparison of immunoassay and mass spectroscopy methods. In: Schultheiss, O.C., Mehta, P.H. (Eds.), Routledge International Handbook of Social Neuroendocrinology. Routledge, Abingdon, UK.
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. J. Pers. Soc. Psychol. 110, 921.
Shirtcliff, E.A., Granger, D.A., Likos, A., 2002. Gender differences in the validity of testosterone measured in saliva by immunoassay. Horm. Behav. 42, 62–69.
Shirtcliff, E.A., Dismukes, A.R., Marceau, K., Ruttle, P.L., Simmons, J.G., Han, G., 2015. A dual-axis approach to understanding neuroendocrine development. Dev. Psychobiol. 57 (6), 643–653.
Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. P-curve: a key to the file-drawer. J. Exp. Psychol. Gen. 1432, 534.
Smith, K.M., Apicella, C.L., 2017. Winners, losers, and posers: the effect of power poses on testosterone and risk-taking following competition. Horm. Behav. 92, 172–181.
Tackett, J.L., Herzhoff, K., Harden, K.P., Page-Gould, E., Josephs, R.A., 2014. Personality × hormone interactions in adolescent externalizing psychopathology. Personal. Disord. 53, 235.
Tellegen, A., 1982. Brief Manual for the Multidimensional Personality Questionnaire. Unpublished manuscript. University of Minnesota, Minneapolis (1031–1010).
Terburg, D., van Honk, J., 2013. Approach–avoidance versus dominance–submissiveness: a multilevel neural framework on how testosterone promotes social status. Emot. Rev. 5, 296–302.
Terburg, D., Aarts, H., van Honk, J., 2012. Testosterone affects gaze aversion from angry faces outside of conscious awareness. Psychol. Sci. 23, 459–463.
Tilbrook, A.J., Turner, A.I., Clarke, I.J., 2000. Effects of stress on reproduction in non-rodent mammals: the role of glucocorticoids and sex differences. Rev. Reprod. 5, 105–113.

Tremblay, P.F., Ewart, L.A., 2005. The Buss and Perry Aggression Questionnaire and its relations to values, the Big Five, provoking hypothetical situations, alcohol consumption patterns, and alcohol expectancies. Personal. Individ. Differ. 38, 337–346.
van den Bos, W., Golka, P., Effelsberg, D., McClure, S., 2013. Pyrrhic victories: the need for social status drives costly competitive behavior. Front. Neurosci. 7, 189.
van der Westhuizen, D., Solms, M., 2015. Social dominance and the affective neuroscience personality scales. Conscious. Cogn. 33, 90–111.
Van der Meij, L., Schaveling, J., van Vugt, M., 2016. Basal testosterone, leadership and dominance: a field study and meta-analysis. Psychoneuroendocrinology 72, 72–79.
van Honk, J., Peper, J.S., Schutter, D.J., 2005. Testosterone reduces unconscious fear but not consciously experienced anxiety: implications for the disorders of fear and anxiety. Biol. Psychiatry 58, 218–225.
van Peer, J.M., Enter, D., van Steenbergen, H., Spinhoven, P., Roelofs, K., 2017. Exogenous testosterone affects early threat processing in socially anxious and healthy women. Biol. Psychol. 129, 82–89.
Viau, V., 2002. Functional cross-talk between the hypothalamic-pituitary-gonadal and-adrenal axes. J. Neuroendocrinol. 14, 506–513.
Victoroff, J., Quota, S., Adelman, J.R., Celinska, B., Stern, N., Wilcox, R., Sapolsky, R.M., 2011. Support for religio-political aggression among teenaged boys in Gaza: Part II: neuroendocrinological Findings. Aggress. Behav. 37 (2), 121–132.
von Rueden, C., Gurven, M., Kaplan, H., 2011. Why do men seek status? Fitness payoffs to dominance and prestige. Proc. R. Soc. B 278, 2223–2232.
Wagels, L., Votinov, M., Kellermann, T., Eisert, A., Beyer, C., Habel, U., 2018. Exogenous testosterone enhances the reactivity to social provocation in males. Front. Behav. Sci. 12, 37.
Wagenmakers, E.J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q.F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J.N., Morey, R.D., 2017. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. Psychon. Bull. Rev. https://doi.org/10.3758/s13423-017-1343-3.
Weisberg, Y.J., DeYoung, C.G., Hirsh, J.B., 2011. Gender differences in personality across the ten aspects of the Big Five. Front. Psychol. 2, 178.
Welker, K.M., Lozoya, E., Campbell, J.A., Neumann, C.S., Carré, J.M., 2014. Testosterone, cortisol, and psychopathic traits in men and women. Physiol. Behav. 129, 230–236.
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.
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.
Whiteside, S.P., Lynam, D.R., 2001. The five factor model and impulsivity: using a structural model of personality to understand impulsivity. Personal. Individ. Differ. 30, 669–689.
Wingfield, J.C., Silverin, B., 1986. Effects of corticosterone on territorial behavior of free-living male song sparrows Melospiza melodia. Horm. Behav. 20, 405–417.
Wirth, M.M., Schultheiss, O.C., 2007. Basal testosterone moderates responses to anger faces in humans. Physiol. Behav. 90, 496–505.
Yasuda, M., Honma, S., Furuya, K., Yoshii, T., Kamiyama, Y., Ide, H., Muto, S., Horie, S., 2008. Diagnostic significance of salivary testosterone measurement revisited: using liquid chromatography/mass spectrometry and enzyme-linked immunosorbent assay. J. Men's Health 5, 56–63.
Zilioli, S., Watson, N.V., 2012. The hidden dimensions of the competition effect: basal cortisol and basal testosterone jointly predict changes in salivary testosterone after social victory in men. Psychoneuroendocrinology 37, 1855–1865.
Zilioli, S., Bird, B.M., 2017. Functional significance of men's testosterone reactivity to social stimuli. Front. Neuroendocrinol. 47, 1–18.
Zilioli, S., Ponzi, D., Henry, A., Maestripieri, D., 2015. Testosterone, cortisol and empathy: evidence for the dual-hormone hypothesis. Adapt. Hum. Behav. Physiol. 1, 421–433.
Zilioli, S., Ponzi, D., Henry, A., Kubicki, K., Nickels, N., Wilson, M.C., Maestripieri, D., 2016. Interest in babies negatively predicts testosterone responses to sexual visual stimuli among heterosexual young men. Psychol. Sci. 27 (1), 114–118.
Zimbardo, P.G., Boyd, J.N., 1999. Putting time in perspective: a valid, reliable individual-differences metric. J. Pers. Soc. Psychol. 77, 1271–1288.