Evolution and Human Behavior 45 (2024) 106615

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Generosity as a status signal: Higher-testosterone men exhibit greater altruism in the dictator game

Julie Novakova [a,*], Petr Tureček [a,b], Kamila Machová [a], Kateřina Sýkorová [a], Vojtěch Zíka [a,c], Jaroslav Flegr [a]

[a] Laboratory of Evolutionary Biology, Department of Philosophy and History of Science, Faculty of Science, Charles University, Prague, Czech Republic
[b] Center for Theoretical Study, Prague, Czech Republic
[c] Center for Behavioral Experiments (CEBEX), Prague, Czech Republic

ARTICLE INFO

Keywords:
Androgen
Dictator game
Dominance
Dual hormone hypothesis
Evolution of altruism
Evolutionary psychology
Social decision making

ABSTRACT

Altruistic behavior can be modulated by many factors including hormonal levels, but their reported effects remain mixed. Understanding the proximate mechanisms of altruism such as these can help test predictions of ultimate, evolutionary explanations. We investigated the relationship of the endogenous salivary levels of testosterone and cortisol with Dictator Game (DG) offers as a proxy of altruism on a sample of general-population participants ($N = 158$, 84 F, 74 M). Bayesian data analysis and model comparison showed both testosterone and cortisol were negatively correlated with DG offers in women, while higher testosterone levels were associated with greater generosity in men. These results suggest that high testosterone may promote altruistic behavior in the service of status-seeking among men.

1. Introduction

1.1. Background

Altruism, meaning apparently selfless helping behavior toward others, has long constituted an evolutionary enigma. Hypotheses such as kin selection (Hamilton, 1964a, 1964b), reciprocity (Trivers, 1971) or costly signaling, dubbed 'handicap hypothesis' (Zahavi, 1995), have been proposed, and it seems that each could apply under some circumstances (West et al., 2011).

Proximate, mechanistic causes were also investigated, both of their own accord and tied to the ultimate evolutionary causes. Among the proximate mechanisms of altruistic behavior, hormonal levels had often been the focus in both humans and animals. But despite the growing attention toward this piece of the altruism puzzle, the specific hormonal influences and their connections to the ultimate causes remain largely unresolved.

Here we focus on the influence of testosterone (T) and cortisol (C) levels on altruism, especially in terms of the dual hormone hypothesis suggesting that basal T and C jointly influence behavioral traits associated with aggression and dominance (Mehta & Josephs, 2010). Conclusions insofar have been mixed, as summarized below. The aim of this study is to shed more light on this complex question and help integrate the proximate hormonal influences with ultimate evolutionary explanations, such as dominance displays in order to attract mates and cement one's social status.

1.2. Testosterone as an altruism modulator

A number of studies looked into the effect of sex hormones on prosocial behavior, with ambiguous results. Various experimental games are typically used as altruism proxies, among them the Dictator Game (DG; Forsythe et al., 1994), where one player is allocated a sum of money and may divide it between self and another player, who is passive in this case. People on average allocate around 28%, with women being usually more generous than men (Engel, 2011). In the earlier-devised Ultimatum Game (UG; Güth et al., 1982), the second player has agency and may accept or reject the first player's offer. In the latter case, neither gets anything. In the Public Goods Game (Marwell & Ames, 1981), players can contribute to a common pot where the contributions are multiplied and divided equally (not based on the initial contribution) between players. Out of these games, the DG is the closest proxy for "pure" altruism unaffected by other social considerations.

High-T men were found to offer non-significantly more in the UG

* Corresponding author at: Viničná 7, 128 00 Prague, Czech Republic.
E-mail address: julie.novakova@natur.cuni.cz (J. Novakova).

https://doi.org/10.1016/j.evolhumbehav.2024.106615
Received 23 November 2023; Received in revised form 2 August 2024; Accepted 6 August 2024
Available online 23 August 2024

than low-T men by Burnham (2007), and women with artificially raised T offered significantly more in the UG than those who'd received placebo (but at the same time, women who *believed* that they'd received testosterone offered less than those who believed they'd ingested placebo). Zethraeus et al. (2009) administered estrogen, testosterone or placebo to postmenopausal women, and found no effect of sex hormones on altruism, fairness and trust in several economic experiments, while Zak et al. (2009) found that men with artificially raised testosterone levels behaved *less* generously in the UG. Low-offer rejections in the UG, often interpreted in terms of fairness preferences but also as aggression, appeared to be increased by T (Burnham, 2007; Mehta & Beer, 2010), decreased by it (Kopsida et al., 2016), as well as independent of it (Cueva et al., 2017).

Such mixed evidence would suggest that the effects of T may be context-dependent and influenced by other variables. This conclusion is supported by Inoue et al. (2017), who found that high-T men with a higher social rank (as maintained in a university rugby team hierarchy) behaved more dominantly – offered less and demanded more – in the UG, while high-T men of a lower rank behaved submissively. The authors concluded that high T likely promotes strategic, rank-dependent social behavior rather than straightforward dominance. A follow-up study (Inoue et al., 2023) found that high-T men whose status has risen recently do in fact behave more dominantly. In addition, high T was also found to correlate with greater parochialism in experimental games (Diekhof et al., 2014; Reimers et al., 2019; Reimers & Diekhof, 2015).

Proxies of earlier androgen exposure can also be measured to elucidate the effects of T on behavior. The second to fourth digit (2D:4D) ratio is sexually dimorphic and thought to be dependent on prenatal T exposure (e.g. Manning et al., 1998, Hönekopp et al., 2007; experimentally in rats Talarovičová et al., 2009), hence potentially serving as a proxy of androgen levels in intrauterine development. Lower digit ratios are thought to indicate higher intrauterine androgen exposure. However, the ratio may simply reflect shifts along the allometric line – larger people have a lower 2D:4D ratio (Kratochvíl & Flegr, 2009). Or, as Knickmeyer et al. (2011) suggested, it might at least in early childhood indicate neonatal T levels rather than prenatal androgen exposure.

Again, some studies found that participants with lower digit ratios (potentially higher T exposure) were more generous, such as Millet and Dewitte (2009) in the DG, but exposure to aggression cues inverted the relationship, and partly Ronay and Galinsky (2011) in a rare face-to-face UG setting. Other studies, in contrast, found people with lower digit ratios more selfish, such as Brañas-Garza and Kovařík (2013) or Buser (2012) across several experimental games. However, the latter study didn't measure the ratio and only asked the participants whether their index or ring finger on each hand is longer or of equal length, which elicited criticism. Others found no effect (Brañas-Garza et al., 2019; Neyse et al., 2020; Parslow et al., 2019). The potential interplay of digit ratio and actual T levels is also unclear. Van Honk et al. (2012) investigated the influence of testosterone on behavior in the Public Goods Game, and found no main effect in either men or women, but there was a significant effect of interaction of testosterone with the 2D:4D ratio: Participants with low prenatal T exposure behaved more prosocially after T administration, while those with high prenatal exposure exhibited no significant change. The potential effect may not persist across ethnic groups: In Galizzi and Nieboer (2015), an inverted U-shaped relationship between DG giving and digit ratio was observed in European-origin participants, but not across the whole participant pool or in other single ethnic groups. To add to the complexity of the problem, Senci et al. (2020) observed a U-shaped relationship of digit ratio and DG giving in a control sample not primed by social norms, and no 2D:4D ratio effect in their experimental sample primed by prescriptive norms on generosity, concluding that biological foundation of altruistic behavior exists, but can be overridden by social norms.

The mixed conclusions and vastly varying experimental designs using only male or female participants, or both sexes, natural salivary or blood T, T administration or indirect cues of T exposure, different experimental games, situations or questions to assess participants' altruism, and often specific primings or cues make it nigh-impossible to reliably generalize the overall effects of T on altruistic behavior. Stanton (2017) reviewed the effects of T on decision-making in various contexts (economic, social, ethical, consumer behavior). The reported effects of T were mixed (mostly either promoting competition and risk-taking, or null). With caution, it can be concluded that the effects of T depend on context (such as the respective social rank, ingroup/outgroup status and sexual or aggression cues, of which little has been investigated thoroughly up to date).

1.3. Cortisol: stress response and altruistic decision-making

As a 'stress hormone', cortisol (C) can indicate impaired mental and/or physical health, and has been associated with impeded memory function, though the duration of the effect, influence of timing and relationship with acute/basal cortisol levels are not completely clear (Bermejo et al., 2022; Brunner et al., 2006; Het et al., 2005; Staufenbiel et al., 2013). Its overall relationship with cooperative and other behavior appears across a range of vertebrate species appears to be complex, in some cases inverse U-shaped (Raulo & Dantzer, 2018).

Human studies also don't show a clear linear pattern. Starcke et al. (2011) compared control and stress-exposed participants' responses to hypothetical realistic moral dilemmas with an altruistic and egotistical reaction option, and while there was no significant difference between both groups, participants whose C was more elevated after the stress task exhibited more egotistical behavior.

Vinkers et al. (2013) studied the time-dependent effect of stress – immediately and with a 75-min latency – on decision-making in the DG and UG, and found the stressed participants to be less generous (time-independently) and more prone to altruistic punishment immediately after stress exposure, but less prone to it than controls after the latency period. Measured salivary C levels did not correlate with altruism in the games at any point. A study by Schulreich et al. (2022) corroborated these results, showing a negative effect of cortisol increase on charitable donations in the control group as well as the experimental group who were subjected to the Trier Social Stress Test.

Contrary to those findings, Singer et al. (2017) found that young men exposed to acute stress via the Trier Social Stress Test behaved more altruistically in experimental situations. Their C levels correlated positively with altruism. Sparrow et al. (2019) also found young adults with higher stress reactivity to be more generous, in this case in the DG. Interestingly, older adults showed a (non-significant) negative association of cortisol and generosity (while in general, older adults were still more altruistic than young adults, as shown in previous studies (Bailey et al., 2013, Matsumoto et al., 2016, Nakavachara, 2018, Ogawa et al., 2018, Pornpattananangkul et al., 2019). Finally, Zhang et al. (2019) found no effect of stress task-induced salivary cortisol on DG or UG behavior in women, but found men to behave more altruistically in the DG with more elevated C.

Schulreich et al. (2022) speculated that the apparent discrepancy between studies finding an increase or decrease of altruism, respectively, in relation to cortisol levels or their changes might come down to competing effects of acute stress: impairing mentalizing (comprehension of one's and others' intentions and behavior), while increasing empathy. The salience of various factors (such as the perceived need of the donation recipient) might be decisive in which of these opposite effects becomes more pronounced.

All these experiments, though, concerned cortisol levels after exposure to some form of acute stress, while basal C levels could be more relevant in the long term. These were studied by Pfattheicher and Keller (2014), who found that basal C correlated negatively with costly punishment of free-riders in the Public Goods Game (PGG). Costly punishment is generally associated with altruistic concerns. However, PGG contributions did not correlate with C levels. The study had male

participants only.

Self-reported chronic stress, which would be associated with increased basal C, was found to have no effect of donations in an anonymous Dictator Game (Ceccato, Kettner, Kudielka, Schwieren, & Voss, 2018). The role of basal cortisol – more interesting from the perspective of Triversian reciprocity, as chronic stress negatively affects memory and expected lifespan – in altruism has received little attention as opposed to cortisol changes under acute stress, and it remains to be investigated in more detail. The present study aims to be a part of that.

We are not aware of any C administration studies using the framework of experimental games up to this point, making clearly distinguishing its effects less simple. However, since high C, especially in the long term, is linked to stress and potentially impaired health and chances of long-term survival, we should expect it to correlate negatively with altruism, viewed through the lens of the reciprocal altruism theory that requires high chances of future interactions and good memory of previous ones.

1.4. Testosterone-cortisol relationship and the dual hormone hypothesis

The dual hormone hypothesis (DHH), posited by Mehta and Josephs (2010), suggests that basal T and C jointly influence behavioral traits associated with aggression and dominance, making high-T, low-C individuals more dominant and status-seeking, while high-T, high-C individuals would exhibit more status loss avoidance. Montoya et al. (2012) reviewed the literature on the role of T and C in reactive (impulsive) and proactive (instrumental) aggression, including social dilemmas and games, and found the findings from experimental games-utilizing studies with endogenous T and C alone or exogenous T suggestive of the above-described interplay of T and C. T seems to promote aggression and social status-seeking, while C is linked to social withdrawal, and its high levels can nullify or override the effects of testosterone.

Denson et al. (2013) investigated the effect of T*C in women on reactive (retaliative, impulsive) aggression, and observed women with both high T and C reacted more aggressively when insulted, while low levels of both hormones predicted more submissive behavior. More in line with the DHH predictions, Pfattheicher et al. (2014) found that high-T, low-C corresponded to destructive (antisocial), but not altruistic punishment in the PGG. In other words, high-T men with low C levels significantly more punished other players who'd contributed just as much as they or more in the PGG, but they did not punish free-riders any more. Neither of the hormones correlated with costly punishment. The study had male participants only. Behavior that could be interpreted as antisocial was similarly found in high-T, low-C individuals with chronic dominance in the Pfattheicher (2017) study, where male participants played a "reversed DG", where they could increase their earnings by decreasing another (dummy) player's earnings.

Zilioli et al. (2015) studied the effects of T and C on empathy, and found that high-T individuals with low basal C had lower empathy scores (in accordance with the folk perception of T as an 'anti-empathy' hormone), while the effect was reversed in individuals with high C levels. Sex differences were observed by Sýkora (2018), who found support for the dual hormone hypothesis in risk-taking in women, but not men, and not in competitiveness in either sex.

Prasad et al. (2019) found support for DHH when investigating the interplay of testosterone and acute stress cortisol response on DG offers; high-T individuals with low C response exhibited most dominant behavior, keeping most of the endowment for themselves. Men and women in the study showed similar patterns. In an earlier study using the UG and the interplay of acute stress-induced cortisol change with testosterone, though, sex differences were observed in the relationship of acute stress and "retaliation" (rejecting an unfair UG offer): women in the high-stress condition were less likely to retaliate than in the low-stress condition, while men were more likely to do so (Prasad et al., 2017).

Both basal and acute cortisol levels may therefore play a key role in modulating the dominance-seeking effects of testosterone. Moreover, the dual hormone hypothesis might explain the inconsistency in reported effects of T and C in studies investigating only one of these hormones. The DHH-framework studies outlined above also show the importance of differentiating between various behaviors of a single category (impulsive/reactive v. instrumental/proactive aggression in Denson et al. (2013) and Pfattheicher et al. (2014)) and highlight the role of sex.

A review and series of studies by Grebe et al. (2019) suggest that the majority of published studies are underpowered to detect a T-C relationship with behavioral traits. Their seven independent studies showed no T*C effects, and only a weak effect of T on status-seeking. They point out, however, that social environment context should be investigated with regard to the dual hormone hypothesis (where single-sex or mixed-sex settings could play a role, and opponent status cues may influence the realized behavior (Knight et al., 2022)). Dekkers et al. (2019) performed a meta-analysis of 30 papers investigating the DHH, yielding tentative support to the hypothesis (at least where social status is concerned). The effects seemed to be stronger in men. However, further studies are needed to elucidate the subject, which we're hoping to contribute to.

In the present study, we investigated the relationship of basal T and C (measured from salivary samples) with altruism in the DG. The majority of previous studies suggest that high-T, low-C individuals exhibit more dominant and status-seeking behavior, but the T*C connection with altruism remains far less clear, therefore we aimed to fill this gap.

2. Methods

2.1. Participants and experimental procedure

The experiments were conducted at the Charles University in Prague in June and July 2018 (for male participants) and 2019 (for female participants). Participants were invited through Facebook. A maximum of 18 enlisted participants could attend one session, and in total, 158 people participated (all central-European, 74 men, 84 women, mean age = 29, median of age = 28, SD = 9.3). The sessions took place in late afternoon only, always with same-sex heterosexual participants and always at the same time and the same season, in order to minimize the noise of diurnal and seasonal fluctuation of hormonal levels.

On site, the participants were greeted at the reception and led to the computer lab (each to their own station divided by partitions), where they read and signed their informed consent and experiment guidelines. Their saliva was collected into salivettes by passive salivation (rolling a swab in their mouth for three minutes). They were aware that C and T levels will be measured in the samples. After the salivettes were recovered by the experimenter, they could begin the online questionnaire including the Dictator Game and general demographic questions (sex, age, education, socioeconomic standing). Other variables (self-reported number of sexual and romantic partners, short personality survey, health, temporal discounting and risk-taking, Ultimatum Game behavior) were measured for a related study (Novakova et al., 2021); optional participant photographs (to be used in a later study) were also taken for a future study after the questionnaire was completed.

The DG was played for 400 tokens translatable into 40 CZK (roughly 2 USD). The participants were first presented with the rules and truthfully told that they would be randomly matched with another person in the room and that their earnings at the end of the experiment would reflect the game outcome. Before playing, the participants answered a control question to assess whether they understood the rules of the game correctly, and a question about having any previous knowledge of either DG or UG (from lectures, literature etc.). Finally, the participants were individually informed about the goal of the experiment and paid their reward gained in the experiment, with the opportunity to gamble in a die roll for a greater reward.

2.2. Saliva treatment and analysis

All saliva samples were stored in salivettes in a freezer with −20 °C until radioimmunoassayed for T and C levels at the Institute of Endocrinology in Prague. For the full methodology see Flegr et al. (2008). In case of insufficient sample volume, only T assays were done, since T effects were expected to play a more substantial role in altruism, based on literature review. Insufficient material for analysis of either hormone was found in four women and ten men. Only T was analyzed two women and eleven men, and only C (where the amount of saliva for T analysis was found insufficient at a later stage) in two women (no men).

2.3. Statistical analysis

We have conducted a Bayesian data analysis and model comparison to assess the relationship between hormonal levels and altruism.

Sex (as an index variable with two levels), Testosterone, Cortisol, and the interaction between Testosterone and Cortisol (as a T/C ratio) were considered as potential predictors of the amount given by the participants in the DG. Logarithms of hormonal levels and the logarithm of T/C ratio were used to avoid the measurement's arbitrariness and other issues connected to raw T/C ratios raised by Del Giudice & Gangestad (2022). The logarithmic scale emphasizes the importance of the relative increase (i.e. 10% increase in the T level) over the absolute increase (i.e. 1 particle/mol). Base 1.1 was used for all logarithms to create an intuitive predictor scale with 1 unit representing a 10% increase. Logarithms were standardised before the analysis to allow for easy assignment of priors to model parameters. Still, despite being made on logarithmic scale, some predictions are visualised on the original scale to allow for a straightforward interpretation. In addition, because some studies that aim to evaluate the dual hormone hypothesis work with a more conventional interaction term, the T*C product, we include results of alternative analysis that works with the product instead of the ratio in Supplement S11.

To avoid the omission of non-trivial relationships, we fitted a broad set of models on the data. Each continuous predictor could enter the analysis in 6 levels of increasing complexity: 0: No association between the predictor and the altruism, 1: Linear relationship, 2–5: Spline model of a given order (order 2 utilised linear, order 3 quadratic, 4–5 cubic B-splines with nodes aligned with appropriate quantiles; see Supplement S1 for spline visualisation and further introduction of the method). Splines were constructed using the bs( ) function from the "splines" package (RCoreTeam, 2019). We used splines instead of more common polynomials, because spline-defined functions cover more versatile set of curves in the acceptable minimum-maximum boundaries for the same number of parameters as compared to polynomials, and because all parameters corresponding to spline weights are advised to have the same prior distribution, unlike slopes multiplying higher order polynomials (see also McElreath, 2020 for further discussion on Bayesian spline models).

A similar approach was recently brought forward as specification curve analyses by Simonsohn et al. (2020) who, however, work in the frequentist framework and explicit hypothesis testing. Our approach combines multi-model inference common in the Bayesian framework with model comparison and spline curve specification.

Sex could have a null relationship with the altruism (i.e. there was a common intercept parameter for men and women), or influence the altruism baseline (i.e. there was a separate intercept by sex), or interact with continuous variables (i.e. each slope or spline weight had a separate value for men and women). Thus 647 models were created ($6 \times 6 \times 6 \times 3-1$; because the by-sex intercept and slope models without hormonal predictors represent the same structure). Models' out-of-sample predictive accuracy was compared using WAIC (Widely Applicable Information Criterion) and the resulting weights.

$$w_i = e^{-0.5\Delta_i} / \Sigma_j e^{-0.5\Delta_j}$$

where Δ_i represents the difference between ith model's WAIC and the lowest WAIC in the set (McElreath, 2020).

It is well known that the dependent variable (donation in the DG) has a challenging distribution. There is an overrepresentation of people donating exactly half of the entrusted money (200 tokens = 20 CZK), and the distribution is constrained by the budget (0–400 here). We have selected a mixture distribution reflecting this fact, a middle-inflated beta-binomial distribution. There are two multiple logistic regressions with the same predictors and model terms in each model. Parameter values in these regressions are, however, independent. First regression, linked to a probability term p in a simple binomial distribution, decides the "middle inflation"; whether or not the participant donates the exact half of the money regardless of other regression terms. The second regression uses the standard beta-binomial distribution with $N = 400$, $\alpha = \mu \, x \, \theta$, $\beta = (1 - \mu) \, x \, \theta$ as a likelihood function. (The expected proportion of the budget donated (μ) and the overdispersion (θ) around this mean. These parameters are sufficient to model discrete values between 0 and 400.)

The regression terms are linked to the log-odds of p and μ. Parameter θ is assumed to be constant and independent of sex.

We used weakly regularised unbiased priors in all models. Prior distributions of regression parameters, both intercepts and slopes, were characterised by normal distributions N ($mean = 0$, $sd = 2$). The posterior distributions were extracted using the Hamiltonian Monte Carlo sampler Stan (StanDevelopmentTeam, 2021) called by the rethinking stan() infrastructure (McElreath, 2020). All compared Stan models were built by a bespoke compiler. All data and computer code used in the analysis are available at https://osf.io/pwztm/.

3. Results

3.1. Descriptive statistics

3.1.1. Dictator game

All 158 responses were valid: the participants finished the experiment and correctly answered the control question for assessing whether they understood the game. The mean amount allocated to another player in the DG was 154.4 out of 400 tokens (~38.6%), median 200 (50%), sd 76.2. Women gave on average 165 tokens, men 142.4 tokens. Previous knowledge of the DG or UG had no statistically significant effect on participants' behavior.

3.1.2. Testosterone

For 16 of the 158 participants, testosterone levels could not be reliably measured due to the small volume of extracted saliva. In the 142 remaining ones, the mean testosterone level was 0.566 nmol/l for men (sd = 0.341), and 0.667 nmol/l for women (sd = 0.222). Two men had very low testosterone levels (together with one woman they were marked as potential outliers, i.e. influential points, on the logarithmic scale), therefore we decided to re-run the analysis without the three data points. The results did not change (see Supplement S10). The two men were, however, not enough to drag the whole male mean down; the whole distribution of testosterone values in men resembles the distribution of T values in women, just shifted by approx. 0.1 (see raw data distributions in Supplement S9). We discuss the observed difference further in Discussion.

3.1.3. Cortisol

Cortisol levels were available for fewer participants ($N = 131$), due to the insufficient sample volume for both T and C assays in some cases. The mean C level was 2.250 nmol/l for men (sd = 1.550), and 2.900 nmol/l for women (sd = 1.250).

3.1.4. Testosterone-cortisol ratio

For participants with both T and C levels measured, T/C ratios were computed. The mean ratio level was 0.33420 for men (sd = 0.383), and

0.28625 for women (sd = 0.231). (Also see the comparison of raw hormonal levels, interaction terms, and their logarithms in Supplement S9.)

3.2. Models of T/C association with altruism

Biological sex and hormonal levels together influenced the expected donation in the DG. Models that contained separate intercepts and slopes for each sex offered a much better out-of-sample predictive accuracy (total weight = 0.79) than models with common slopes (total weight = 0.21) and models independent of sex (total weight = 0.00). Out of the 100 best models, 77 were models with the interaction between sex and hormonal levels, 23 had at least by-sex intercepts. 100 worst models were exclusively models without the effect of sex. This can be interpreted as participant sex playing a key role in predicting altruistic behavior.

The best model (slo_051) combined the complicated relationship between log(C) and DG donation (fifth-order spline structure) with a straightforward linear effect of log(T/C) ratio (Fig. 1, see Supplement S3 for the posterior distribution of model parameters). Although non-trivial, C-related changes in the mean donation seem to be represented faithfully by this curve. In the first 39 models, the relationship between

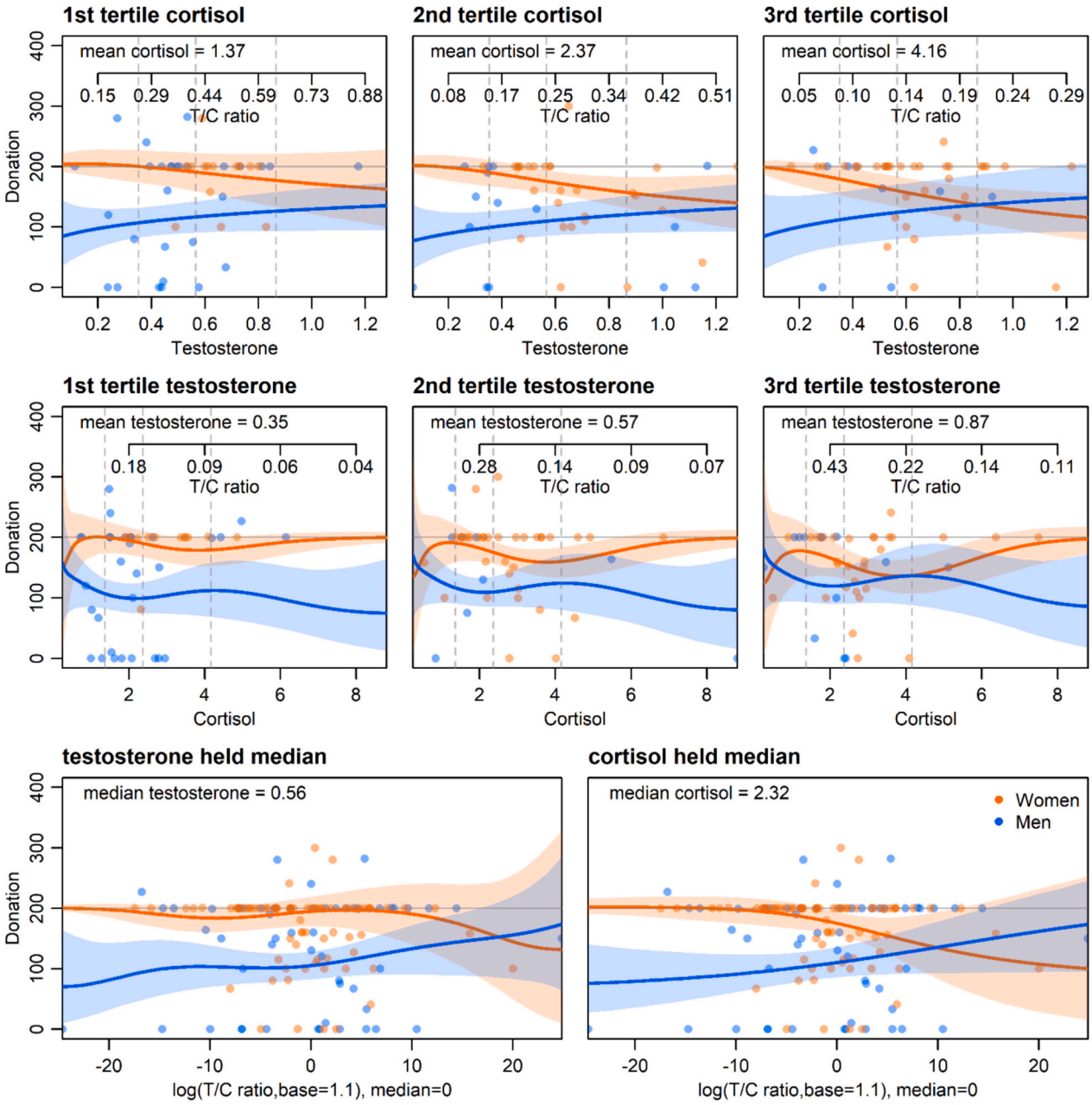

Fig. 1. Predictions of the best model. The dashed vertical lines in the two upper rows indicate positions of the respective tertile means in complementary panels. The raw data in these rows are limited to tertiles in panel titles. In the predictions from the perspective of T/C ratio, all raw data points are included (women shown in orange, men in blue). The counterfactual value in the top left corner was held constant to produce predictions of change in donation with the x-axis in each panel (e.g. if the T was held median, only C values varied to produce desired T/C ratios). The additional T/C ratio axes by tertiles in top two rows are added for easier interpretation of the interaction term, which changes with both T and C. They allow to read all panels as having T/C ratio on the x-axis (the mean value for each respective tertile is used to calculate corresponding T/C values for each x-axis tick). The solid lines represent mean predictions, and the semi-transparent corridors span 89% compatibility intervals based on 10,000 samples from the posterior distribution. To see the inverse visualisation with hormonal levels logarithms and the ratio on the original scale, see Fig. S4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

C and donation is either the above-mentioned fifth-order spline (25 models), or none (14 models). The best simpler model can be found on position 40 (slo_220), and even the best model using just cortisol to predict altruism suggests the more complicated relationship (slo_050, 29th best model). Although the null C models fall second (total weight 0.28), it is improbable that the relationship is non-existent and we identified a false pattern. It just means that no simple curve can faithfully represent the relationship over the whole recorded cortisol range. The null model minimises the risk of overfitting, but at the expense of overall low accuracy. In overall, the models with by-sex slopes and fifth-order B-spline structure for log(C) predictor have an accumulated weight of 0.47. Given the predictive accuracy and the comparison to other models, the best model can be interpreted as realistically capturing the complex hormonal influences.

The cortisol-altruism relationship can be described as U-shaped, with U-reversal at the very beginning of the detected C range (i.e. incomplete M-shaped in women and incomplete W-shaped in men; see Fig. 1, and also Figs. S17, and S18, with the results of models that fall second and third in the comparison). Both men and women with very low C levels are predicted to donate similar amounts. Then, the donations rise sharply in women and decrease in men to decrease, respectively rise again to a similar amount around the mean of the third tertile. The lines reach plateaus of high donations in women and low donations in men for extremely high C values. Due to the limited sample size, however, the compatibility intervals around the extreme ends are very wide, and one shall be cautious in their interpretation. The concave profile of the curve characterising women donations up to the third tertile and the convex profile of donations by men do seem well-grounded.

With increasing T (either relative to C or absolute, depending on the model, because, to some extent, these variables are interchangeable), men's donations increase, while women's donations decrease (Fig. 2). Parcellation of the model into two logistic regressions brings little new information. If the donation is expected to rise according to the beta-binomial model's central tendency μ, so does the probability of donating the exact half of the budget p (see Supplementary Figs. S9-S18). That is because the fair share represents a very generous offer. This conclusion was not hard-wired in our model; the parameters of the regressions were independent. The fact that the general patterns agree is a good sanity check. It suggests that despite a complicated mixture

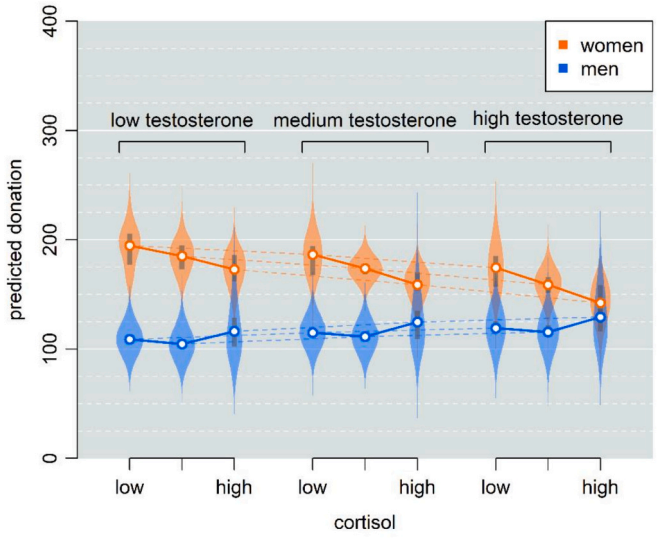

Fig. 2. Aggregated prediction based on posterior distributions of the 10 best models. The density plots depict the complete distribution of mean donation prediction, including sample variation within each model. The shaded rectangles span 89% of the posterior distribution means. White points indicate grand means of the posterior distribution means. Low, medium, and high levels correspond to mean values per 1st, 2nd, and 3rd tertile of data distributions.

distribution, decisions whether or not to split the budget equally and how much to donate otherwise are not governed by fundamentally different underlying processes in participants' minds.

T performs better as an isolated predictor than C (models with only T have a cumulative weight of 0.08, C-only models just 0.03, models using only T/C ratio have a cumulative weight of 0.07). The best model using a single hormonal predictor employs two T splines (model slo_200, overall 5th best model; see Supplementary Fig. S19 for the visualisation: the function is growing, respectively declining over the whole T range, the first spline only moderates the increase, respectively decrease of the donation).

If testosterone is given enough space to interact with cortisol in men, an interesting pattern emerges. The positive effect of T on DG offers is most pronounced if C is sufficiently low. For higher C values in men, the relationship between T and generosity gets slightly flattened or, possibly even reversed (see, for instance, the testosterone-perspective plots in the alternative models with log(T*C) product in Supplementary Fig. S32. The best log(T*C) model WAIC = 934.53 was even slightly better than the best log(T/C) model WAIC = 936.48). Caution, however, is advisable when interpreting this interaction, because of the relatively low number of men with high C and T values (notice wide 89% CIs of posterior mean – shaded rectangles – in high-C men in Fig. 2).

The overall best model with T/C ratio suggests that the interaction may be non-trivially present in the effect of the choice to split the budget in half in women (see parameter estimates in Supplementary Fig. S6). Women with higher log(T/C) are expected to show lower probability of splitting the budget in half, while the continuous effect on the donation otherwise does not show conclusive difference from 0. Further studies should aim to replicate our effort to faithfully reproduce the mixture distribution present in DG offers to support or revise such subtleties in how hormonal levels are translated to realized signs of generosity.

Outside of extreme values, the set of models agree well that between means of the first and the third tertile of T levels (see the dashed vertical lines in Fig. 1 that indicate by-tertile means), the expected donation declines with increasing T in women and rises with increasing T in men. The effect of C between the respective tertiles is similar in women; they donate conclusively less with higher C, unlike men.

The aggregated donation predictions for tertile-mean combinations based on the set of 10 best models (Fig. 2) is in accord with this finding (see Fig. S5 for their overview). As T increases, men donate more and women less. Women's donations tend to decrease with increasing C. In contrast, men's donations do not respond strongly to this hormone's changes, perhaps except for the moderating effect of C on the effect of T (which is most pronounced when C is low as described above). These effects are preserved even if we use the 100 best models instead of 10 (see Supplementary Fig. S20, see Fig. S21 if you prefer plot alignment with alternating T levels).

These reported effects of hormones are robust, even though the model omitting hormones' effects scored relatively high (see Supplementary Fig. S4, look for model int_000, 10th best model overall, the relative weights of the best model slo_051 and this one are 0.68 and 0.32).

If we disrupt the relationship between hormones and donation by randomly shuffling the vectors holding T and C values, we arrive at a completely different picture (see Supplement S6). Models using separate intercepts and slopes for each sex no longer cluster at the beginning of the models' weight comparison; they become interchangeable with models predicting donations using common slopes. Furthermore, models with a low number of splines are much favoured over the complex models, which is not the case in the actual-data comparison in Fig. S4 (this difference between the empirical and the scrambled dataset is too large to be attributable to the sampling variation, see Supplement S7).

The difference in donation between men and women, regardless of the hormonal levels, is well established. The null model (no_000) is at the end of the overall comparison. The weights in a simulation, where

hormonal levels and sex labels are randomly permuted (Supplement S6), suggest the reversal of the predicted out-of-sample accuracy than reported in Fig. S4 (see Fig. S23).

Still, there is a lot of unexplained variation in altruism even after sex and hormonal balance are accounted for. The θ characterising the overdispersion of donation around the predicted budget allocation is 1.93 [89% CI 1.42–2.52] according to the best model, which suggests relatively evenly spread beta distribution (higher θ correspond to the tighter distribution of probability density). The WAIC of the best model is 936.49, and the WAIC of the worst model is 957.05, while the estimated standard error of the best model's WAIC is 66.50. That signals that even the best model does not perform much better than the worst model in the set. It is the huge overdispersion that causes it (see Supplement S8), indicating there is still a lot of variation to explain. Put simply, the best model likely captures the real complexity of T and C influences on DG donations, but explains only a small variance in the DG donation distribution. Future studies should also investigate the mutual relationships of other physiological or neurological correlates. Nevertheless, our results suggest that the interaction between hormonal levels and sex is an integral part of any model aiming to predict donation in the DG precisely.

4. Discussion

Consistently with previous studies (see the Engel, 2011 meta-study), women in our sample were more generous in the DG than men. We observed that DG allocations consistently declined with increasing T in women and rose, albeit less steeply, with higher T in men. The relationship between altruism and C was more complicated, roughly N-shaped in women and И-shaped in men, again across all tertiles of T levels; but it can be approximated that women contributed less with rising C levels, while men exhibited a weaker opposite pattern (see Fig. 2) – this is what happens between the mean value of the first and third tertile.

According to the dual hormone hypothesis, behavioral patterns commonly associated with high T tend to be more pronounced in individuals with low basal C. The results we observed do not neatly fit this interpretation due to the nontrivial shape of the relationship of C with altruism; high-T and low-C, but to some extent also high-C men as opposed to median-C men behaved more generously in the DG.

While T, especially in men, is often thought to promote aggression and selfishness (more or less suggested also by the studies of Zak et al., 2009, Diekhof et al., 2014, Inoue et al., 2017, Reimers et al., 2019), the actual reality appears more complicated. In fact, high-T men in our sample donated more than low-T men (especially if they had low C levels), more in line with Burnham (2007). The dual hormone hypothesis suggests that high T (combined with low C) should be linked to social status-seeking and aggression, and high C to social withdrawal (modulating the effect of T on status-seeking).

In our study, greater altruism in high-T men could be interpreted in terms of status-seeking and status display, which corresponds to Durkee et al. (2020) findings that benefit-generation predicts status allocation cross-culturally, whilst cost-infliction only predicts status weakly or not at all. Interestingly, this interpretation is at odds with Inoue et al. (2017, 2023), who found high-T men to offer less and demand more in the UG, and Pfattheicher (2017), where high-T, low-C dominant men took more in a reversed DG. While the UG and reverse-DG can be compared to the DG only with caution, these findings and their interpretation invite the question: Is altruism a signal of dominance, or submission?

We suspect that the answer would be very much context-dependent, and that the question itself, asked without adding "under conditions…", may be much too simplified to elicit any truly useful answers. It would be useful to view both the hormonal influences on altruism and its ultimate benefits more in the view of life histories (and thus varying costs and benefits), both within- and across studies (Roney, 2016).

If we interpret our results in light of broadly generalized mate choice and seeking strategies between the sexes – where men are more likely to showcase the abundance of resources and women to look for these signals (e.g. Todd et al., 2007; Waynforth & Dunbar, 1995; Wiederman, 1993), it makes sense that the proximate mechanisms of hormonal influences on altruistic behavior should differ by sex, shaped by different ultimate causes, here status display (relevant for mate attraction and intra-sexual competition) for men.

A single study, of course, cannot bring any strong conclusions about the evolutionary origins of human altruism. It can, however, represent one important piece of the whole puzzle of factors contributing to the evolution of altruistic behavior. The data-driven approach that we employed revealed an interesting pattern that invites both further theoretical development and empirical replication.

There were several potential limitations to the study. In our sample, puzzlingly, women had higher mean T levels than men. A plausible explanation is that salivary T correlates closely with serum T in men, but not necessarily in women (Shirtcliff et al., 2002), making this a possible limitation. Higher salivary T for women may reflect relatively lower serum testosterone, since the linear regression equations for serum and salivary T differ (Salimetrics, 2019). Fiers et al. (2014) raise concerns about salivary T comparability to serum T levels and its difference between the sexes, where women had relatively raised salivary T compared to serum T – a fact which may explain the difference observed in our study.

In general, observed salivary T-behavior connections in men should be treated as reliable, in women with some caution. Although salivary testing of T in women was considered reliable based on multiple results (e.g. Baxendale et al., 1982; Dabbs Jr, 1990), later studies showed that it could reflect serum levels poorly (Fiers et al., 2014; Flyckt et al., 2009), indeed to such an extent that some authors warned before using it to estimate women's T levels (Davison, 2009). High circadian variability with episodic fluctuations exceeding the typical morning maxima was detected by Al-Dujaili and Sharp (2012). Age may have played a role as well, since T levels drop with age in both sexes to a similar relative degree, but the absolute decrease is much greater in men (Keevil et al., 2017). Inter-laboratory differences may also exist, although results from multiple labs tended to be fairly in agreement in Dabbs Jr et al. (1995). Lastly, the method of collection may have influenced the measured T levels, especially in women; Prasad et al. (2017) found that when using salivettes, there "was inflation of testosterone values at the lower end of the distribution (ostensibly in female participants)". While the salivette cotton swab technique and passive drool collection exhibited close correlation between measured hormonal levels in their study, it cannot be excluded that the collection technique may have inflated the measured T levels in women in our study. Moreover, while cotton swabs and passive drool lead to very similar mean measured cortisol levels, some studies indicate that the within-participant correlation may be low (Kozaki et al., 2009), as some molecules contained in cotton may have cross-linking with assay antibodies, and the cotton might absorb some of the cortisol. However, as Kozaki et al. (2009) conclude, all hormonal collection methods may introduce their own noise and potentially bias.

Salivary testosterone corresponds to bioavailable testosterone more than serum T levels do, which is why we chose this collection method, but it is very dependent on the experimental conditions, as outlined above. It's possible than women adhered more to the pre-experimental instructions to drink more water before coming to the experimental session and to salivate on the swab in their mouth for the whole three minutes' duration of saliva collection. That could also explain why women's testosterone results measured on average higher. In women, there were also fewer samples where the swab contained too little material for both T and C assays, which would support this potential explanation.

We must stress that this result, while puzzling, doesn't constitute a major problem for the aims of the current study, since the Bayesian analyses were done separately for men and women. One important takeaway for future studies is that separate sessions for different participant

groups should not be held, because the effect of the session (despite the carefully maintained identical conditions of the session room, time of day, time of year, instructions, researchers on-site and material equipment) may potentially override the effect of treatment.

Given the more debatable comparability of salivary and serum T levels in women and possibly raised salivary T (as compared to serum T) in women relatively to men (Davison, 2009; Fiers et al., 2014; Flyckt et al., 2009), we suspect that more studies arriving to similarly peculiar results may have been conducted, but possibly unpublished due to publication bias and concerns over data reliability. If that is so, their publication is desirable in order to achieve greater reliability and replicability in future studies.

Another potential limitation could be that only a single measurement of C and T levels was done for each participant, making it potentially prone to fluctuations. However, the experiments were conducted always at the same time of day and season of the year, and all the participants received the same instructions beforehand (that they should only come if healthy, go to sleep at their usual time the previous day, avoid overt stress or physical exertion before the experiment, drink enough water and avoid consuming food and flavored drinks at least an hour before the start of the experiment). Any out-of-the-usual hormonal fluctuation in our sample should therefore only introduce noise, but no systematic error.

Finally, while the stakes were only approx. 2 USD, a recent meta-analysis of stake size in DG and UG found no effect of stakes on the offer size in UG and a statistically significant, but practically very small effect in DG (Larney et al., 2019), although it cannot be excluded that some effect might exist within the context of hormonal levels.

To summarize, our findings provide tentative support for the dual-hormone theory, where high-T men (especially if also low-C) may express their dominance through altruism as a status signal. This finding is not that surprising given different mate choice strategies across sexes, especially pertaining to resource signaling (Todd et al., 2007; Waynforth & Dunbar, 1995; Wiederman, 1993).

Further studies and meta-analyses are sorely needed to ascertain the hormonal effects on altruism (including predictions made by different ultimate explanations of altruism) and the role of the dual hormone hypothesis. The current evidence remains mixed, but differences in study design (e.g. measuring salivary, blood or hair cortisol or testosterone; working with endogenous hormone levels or using hormonal administration; putting the participants in either the role of proposer or responder in a given game, or both; using experimental games or other measures of altruism; anonymous or face-to-face setting in the game; hypothetical or real stakes, and their value; mentioning specifically testosterone before the experiment, etc.) may be responsible for some of the observed differences, highlighting the need for carefully constructed meta-analyses and for far more replication studies. For studying relatively easily measurable proxies of androgen exposure, such as the digit ratio, large-scale studies such as Neyse et al. (2020) are preferable. For measuring hormonal levels directly, this is of course not attainable due to the cost of obtaining and analyzing the samples, and future meta-studies and replications can at least partly overcome the resulting small-sample issues.

The findings ascertained here will hopefully contribute to untangling the current enigma of hormonal influences of altruistic behavior, which appears to be complex and often situationally-dependent. Our results show that both high T and C were related to lower generosity in women, while the relationship of T ran in the opposite direction in men, where it was further modulated by the levels of C in a complex pattern. This outcome tentatively suggests that altruism could be interpreted in terms of status-seeking and display, in line with earlier findings of Durkee et al. (2020) that benefit-generation rather than cost-affliction leads to higher status allocation. However, further research is needed to test these conclusions more thoroughly and in a greater variety of contexts and methods, such as mixed-sex groups or higher stakes.

CRediT authorship contribution statement

Julie Novakova: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Petr Tureček:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Kamila Machová:** Writing – original draft, Project administration, Methodology. **Kateřina Sýkorová:** Writing – original draft, Project administration, Methodology. **Vojtěch Zíka:** Writing – original draft, Project administration, Methodology. **Jaroslav Flegr:** Writing – original draft, Validation, Supervision, Resources, Methodology, Investigation, Conceptualization.

Financial support

This work was funded by the project GA UK 1502018, supported by the Charles University's grant agency.

Ethics Approval

The study was approved by the IRB of the Faculty of Science, Charles University.

Research Transparency & Reproducibility

All data and computer code used in the analysis are publicly available at https://osf.io/pwztm/.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2024.106615.

References

Al-Dujaili, E. A. S., & Sharp, M. A. (2012). Female salivary testosterone: Measurement, challenges and applications. *Steroids-From Physiology to Clinical Medicine*, 129–167. https://doi.org/10.5772/53648

Bailey, P. E., Ruffman, T., & Rendell, P. G. (2013). Age-related differences in social economic decision making: The ultimatum game. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 68*(3), 356–363. https://doi.org/10.1093/geronb/gbs073

Baxendale, P. M., Jacobs, H. S., & James, V. H. T. (1982). Salivary testosterone: Relationship to unbound plasma testosterone in normal and hyperandrogenic women. *Clinical Endocrinology, 16*(6), 595–603. https://doi.org/10.1111/j.1365-2265.1982.tb03176.x

Bermejo, J. L., Valldecabres, R., Villarrasa-Sapiña, I., Monfort-Torres, G., Marco-Ahulló, A., & Do Couto, B. R. (2022). Increased cortisol levels caused by acute resistance physical exercise impair memory and learning ability. *PeerJ, 10*, Article e13000. https://doi.org/10.7717/peerj.13000

Brañas-Garza, P., Espín, A. M., García-Muñoz, T., & Kovářík, J. (2019). Digit ratio (2D:4D) and prosocial behaviour in economic games: No direct correlation with generosity, bargaining or trust-related behaviours. *Biology Letters, 15*(8), 20190185. https://doi.org/10.1098/rsbl.2019.0185

Brañas-Garza, P., & Kovařík, J. (2013). Digit ratios and social preferences: A comment on Buser (2012). ESI Working Paper 13–31. Retrieved from http://digitalcommons.chapman.edu/esi_working_papers/29.

Brunner, R., Schaefer, D., Hess, K., Parzer, P., Resch, F., & Schwab, S. (2006). Effect of high-dose cortisol on memory functions. *Annals of the New York Academy of Sciences, 1071*(1), 434–437. https://doi.org/10.1196/annals.1364.037

Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society B: Biological Sciences, 274*(1623), 2327–2330. https://doi.org/10.1098/rspb.2007.0546

Buser, T. (2012). Digit ratios, the menstrual cycle and social preferences. *Games and Economic Behavior, 76*(2), 457–470. https://doi.org/10.1016/j.geb.2012.07.006

Ceccato, S., Kettner, S. E., Kudielka, B. M., Schwieren, C., & Voss, A. (2018). Social preferences under chronic stress. *PloS one, 13*(7), Article e0199528. https://doi.org/10.1371/journal.pone.0199528

Cueva, C., Roberts, R. E., Spencer, T. J., Rani, N., Tempest, M., Tobler, P. N., … Rustichini, A. (2017). Testosterone administration does not affect men's rejections of low ultimatum game offers or aggressive mood. *Hormones and Behavior, 87*, 1–7. https://doi.org/10.1016/j.yhbeh.2016.09.012

Dabbs, J. M., Jr. (1990). Salivary testosterone measurements: Reliability across hours, days, and weeks. *Physiology & Behavior, 48*(1), 83–86. https://doi.org/10.1016/0031-9384(90)90265-6

Dabbs, J. M., Jr., Campbell, B. C., Gladue, B. A., Midgley, A. R., Navarro, M. A., Read, G. F., … Worthman, C. M. (1995). Reliability of salivary testosterone measurements: A multicenter evaluation. *Clinical Chemistry, 41*(11), 1581–1584. https://doi.org/10.1093/clinchem/41.11.1581

Davison, S. (2009). Salivary testing opens a Pandora's box of issues surrounding accurate measurement of testosterone in women. *Menopause, 16*(4), 630–631. https://doi.org/10.1210/js.2016-1029

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? *Neuroscience & Biobehavioral Reviews, 96*, 250–271. https://doi.org/10.1016/j.neubiorev.2018.12.004

Del Giudice, M., & Gangestad, S. W. (2022). Hormone ratios suffer from striking lack of robustness to measurement error. *Psychoneuroendocrinology, 142*, 105802. https://doi.org/10.1016/j.psyneuen.2022.105802

Denson, T. F., Mehta, P. H., & Tan, D. H. (2013). Endogenous testosterone and cortisol jointly influence reactive aggression in women. *Psychoneuroendocrinology, 38*(3), 416–424. https://doi.org/10.1016/j.psyneuen.2012.07.003

Diekhof, E. K., Wittmer, S., & Reimers, L. (2014). Does competition really bring out the worst? Testosterone, social distance and inter-male competition shape parochial altruism in human males. *PLoS One, 9*(7), Article e98977. https://doi.org/10.1371/journal.pone.0098977

Durkee, P. K., Lukaszewski, A. W., & Buss, D. M. (2020). Psychological foundations of human status allocation. *Proceedings of the National Academy of Sciences, 117*(35), 21235–21241. https://doi.org/10.1073/pnas.2006148117

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*, 583–610. https://doi.org/10.1007/s10683-011-9283-7

Fiers, T., Delanghe, J., T'Sjoen, G., Van Caenegem, E., Wierckx, K., & Kaufman, J. M. (2014). A critical evaluation of salivary testosterone as a method for the assessment of serum testosterone. *Steroids, 86*, 5–9. https://doi.org/10.1016/j.steroids.2014.04.013

Flegr, J., Lindová, J., Pivoňková, V., & Havlíček, J. (2008). Brief communication: Latent toxoplasmosis and salivary testosterone concentration—Important confounding factors in second to fourth digit ratio studies. *American Journal of Physical Anthropology, 137*(4), 479–484. https://doi.org/10.1002/ajpa.20888

Flyckt, R. L., Liu, J., Frasure, H., Wekselman, K., Buch, A., & Kingsberg, S. A. (2009). Comparison of salivary versus serum testosterone levels in postmenopausal women receiving transdermal testosterone supplementation versus placebo. *Menopause, 16*(4), 680–688. https://doi.org/10.1097/gme.0b013e318199d5c4

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*(3), 347–369. https://doi.org/10.1006/game.1994.1021

Galizzi, M. M., & Nieboer, J. (2015). Digit ratio (2D: 4D) and altruism: Evidence from a large, multi-ethnic sample. *Frontiers in Behavioral Neuroscience, 9*, 41. https://doi.org/10.3389/fnbeh.2015.00041

Grebe, N. M., Del Giudice, M., Thompson, M. E., Nickels, N., Ponzi, D., Zilioli, S., … Gangestad, S. W. (2019). Testosterone, cortisol, and status-striving personality features: A review and empirical evaluation of the dual hormone hypothesis. *Hormones and Behavior, 109*, 25–37. https://doi.org/10.1016/j.yhbeh.2019.01.006

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Hamilton, W. D. (1964a). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology, 7*(1), 1–16. https://doi.org/10.1016/0022-5193(64)90038-4

Hamilton, W. D. (1964b). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology, 7*(1), 17–52. https://doi.org/10.1016/0022-5193(64)90039-6

Het, S., Ramlow, G., & Wolf, O. T. (2005). A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology, 30*(8), 771–784. https://doi.org/10.1016/j.psyneuen.2005.03.005

Hönekopp, J., Bartholdt, L., Beier, L., & Liebert, A. (2007). Second to fourth digit length ratio (2D: 4D) and adult sex hormone levels: New data and a meta-analytic review. *Psychoneuroendocrinology, 32*(4), 313–321. https://doi.org/10.1016/j.psyneuen.2007.01.007

Inoue, Y., Burriss, R. P., Hasegawa, T., & Kiyonari, T. (2023). Testosterone promotes dominance behaviors in the ultimatum game after players' status increases. *Scientific Reports, 13*(1), 18029. https://doi.org/10.1038/s41598-023-45247-4

Inoue, Y., Takahashi, T., Burriss, R. P., Arai, S., Hasegawa, T., Yamagishi, T., & Kiyonari, T. (2017). Testosterone promotes either dominance or submissiveness in the ultimatum game depending on players' social rank. *Scientific Reports, 7*(1), 1–9. https://doi.org/10.1038/s41598-017-05603-7

Keevil, B. G., Clifton, S., Tanton, C., Macdowall, W., Copas, A. J., Lee, D., … Mercer, C. H. (2017). Distribution of salivary testosterone in men and women in a british general population-based sample: The third national survey of sexual attitudes and lifestyles (Natsal-3). *Journal of the Endocrine Society, 1*(1), 14–25. https://doi.org/10.1210/js.2016-1029

Knickmeyer, R. C., Woolson, S., Hamer, R. M., Konneker, T., & Gilmore, J. H. (2011). 2D: 4D ratios in the first 2 years of life: Stability and relation to testosterone exposure and sensitivity. *Hormones and Behavior, 60*(3), 256–263. https://doi.org/10.1016/j.yhbeh.2011.05.009

Knight, E. L., Morales, P. J., Christian, C. B., Prasad, S., Harbaugh, W. T., Mehta, P. H., & Mayr, U. (2022). The causal effect of testosterone on men's competitive behavior is moderated by basal cortisol and cues to an opponent's status: Evidence for a context-dependent dual-hormone hypothesis. *Journal of Personality and Social Psychology, 123*(4), 693. https://doi.org/10.1037/pspa0000305

Kopsida, E., Berrebi, J., Petrovic, P., & Ingvar, M. (2016). Testosterone administration related differences in brain activation during the ultimatum game. *Frontiers in Neuroscience, 10*, 66. https://doi.org/10.3389/fnins.2016.00066

Kozaki, T., Hashiguchi, N., Kaji, Y., Yasukouchi, A., & Tochihara, Y. (2009). Effects of saliva collection using cotton swab on cortisol enzyme immunoassay. *European Journal of Applied Physiology, 107*(6), 743–746. https://doi.org/10.1007/s00421-009-1178-3

Kratochvíl, L., & Flegr, J. (2009). Differences in the 2nd to 4th digit length ratio in humans reflect shifts along the common allometric line. *Biology Letters, 5*(5), 643–646. https://doi.org/10.1098/rsbl.2009.0346

Larney, A., Rotella, A., & Barclay, P. (2019). Stake size effects in ultimatum game and dictator game offers: A meta-analysis. *Organizational Behavior and Human Decision Processes, 151*, 61–72. https://doi.org/10.1016/j.obhdp.2019.01.002

Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: A predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Human Reproduction (Oxford, England), 13*(11), 3000–3004. https://doi.org/10.1093/humrep/13.11.3000

Marwell, G., & Ames, R. E. (1981). Economists free ride, does anyone else? Experiments on the provision of public goods, IV. *Journal of Public Economics, 15*(3), 295–310. https://doi.org/10.1016/0047-2727(81)90013-X

Matsumoto, Y., Yamagishi, T., Li, Y., & Kiyonari, T. (2016). Prosocial behavior increases with age across five economic games. *PLoS One, 11*(7). https://doi.org/10.1371/journal.pone.0158671

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Oxon: CRC Press. https://doi.org/10.1201/9781315372495

Mehta, P. H., & Beer, J. (2010). Neural mechanisms of the testosterone–aggression relation: The role of orbitofrontal cortex. *Journal of Cognitive Neuroscience, 22*(10), 2357–2368. https://doi.org/10.1162/jocn.2009.21389

Mehta, P. H., & Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior, 58*(5), 898–906. https://doi.org/10.1016/j.yhbeh.2010.08.020

Millet, K., & Dewitte, S. (2009). The presence of aggression cues inverts the relation between digit ratio (2D: 4D) and prosocial behaviour in a dictator game. *British Journal of Psychology, 100*(1), 151–162. https://doi.org/10.1348/000712608X324359

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. *Motivation and Emotion, 36*(1), 65–73. https://doi.org/10.1007/s11031-011-9264-3

Nakavachara, V. (2018). The economics of altruism: The old, the rich, the female. *Journal of Human Behavior in the Social Environment, 28*(1), 28–43. https://doi.org/10.1080/10911359.2017.1393367

Neyse, L., Vieider, F. M., Ring, P., Probst, C., Kaernbach, C., van Eimeren, T., & Schmidt, U. (2020). Risk attitudes and digit ratio (2D: 4D): Evidence from prospect theory. *Journal of Risk and Uncertainty, 60*(1), 29–51. https://doi.org/10.1007/s11166-020-09321-w

Novakova, J., Machová, K., & Flegr, J. (2021). Looking like a million dollars: Does attractiveness priming increase altruistic behavior in experimental games? *Frontiers in Psychology, 12*, Article 658466. https://doi.org/10.3389/fpsyg.2021.658466

Ogawa, K., Kawamura, T., & Matsushita, K. (2018). *Higher the cognitive ability of, and/or the younger the dictator, the more selfish: Experimental evidence*. Available at SSRN 3267146. https://doi.org/10.2139/ssrn.3267146.

Parslow, E., Ranehill, E., Zethraeus, N., Blomberg, L., von Schoultz, B., Hirschberg, A. L., … Dreber, A. (2019). The digit ratio (2D: 4D) and economic preferences: No robust associations in a sample of 330 women. *Journal of the Economic Science Association, 5*(2), 149–169. https://doi.org/10.1007/s40881-019-00076-y

Pfattheicher, S. (2017). Illuminating the dual-hormone hypothesis: About chronic dominance and the interaction of cortisol and testosterone. *Aggressive Behavior, 43*(1), 85–92. https://doi.org/10.1002/ab.21665

Pfattheicher, S., & Keller, J. (2014). Towards a biopsychological understanding of costly punishment: The role of basal cortisol. *PLoS One, 9*(1), Article e85691. https://doi.org/10.1371/journal.pone.0085691

Pfattheicher, S., Landhäußer, A., & Keller, J. (2014). Individual differences in antisocial punishment in public goods situations: The interplay of cortisol with testosterone and dominance. *Journal of Behavioral Decision Making, 27*(4), 340–348. https://doi.org/10.1002/bdm.1811

Pornpattananangkul, N., Chowdhury, A., Feng, L., & Yu, R. (2019). Social discounting in the elderly: Senior citizens are good Samaritans to strangers. *The Journals of Gerontology: Series B, 74*(1), 52–58. https://doi.org/10.1093/geronb/gbx040

Prasad, S., Knight, E. L., & Mehta, P. H. (2019). Basal testosterone's relationship with dictator game decision-making depends on cortisol reactivity to acute stress: A dual-hormone perspective on dominant behavior during resource allocation. *Psychoneuroendocrinology, 101*, 150–159. https://doi.org/10.1016/j.psyneuen.2018.11.012

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. *Hormones and Behavior, 92*, 128–140. https://doi.org/10.1016/j.yhbeh.2016.10.020
Raulo, A., & Dantzer, B. (2018). Associations between glucocorticoids and sociality across a continuum of vertebrate social behavior. *Ecology and Evolution, 8*(15), 7697–7716. https://doi.org/10.1002/ece3.4059
RCoreTeam. (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org.
Reimers, L., & Diekhof, E. K. (2015). Testosterone is associated with cooperation during intergroup competition by enhancing parochial altruism. *Frontiers in Neuroscience, 9*, 183. https://doi.org/10.3389/fnins.2015.00183
Reimers, L., Kappo, E., Stadler, L., Yaqubi, M., & Diekhof, E. K. (2019). Endogenous testosterone correlates with parochial altruism in relation to costly punishment in different social settings. *PeerJ, 7*, Article e7537. https://doi.org/10.7717/peerj.7537
Ronay, R., & Galinsky, A. D. (2011). Lex talionis: Testosterone and the law of retaliation. *Journal of Experimental Social Psychology, 47*(3), 702–705. https://doi.org/10.2139/ssrn.1872648
Roney, J. R. (2016). Theoretical frameworks for human behavioral endocrinology. *Hormones and Behavior, 84*, 97–110. https://doi.org/10.1016/j.yhbeh.2016.06.004
Salimetrics. (2019). Expanded range salivary testosterone enzyme immunoassay kit. Retrieved from: https://salimetrics.com/wp-content/uploads/2018/03/testosterone-saliva-elisa-kit.pdf.
Schulreich, S., Tusche, A., Kanske, P., & Schwabe, L. (2022). Altruism under stress: Cortisol negatively predicts charitable giving and neural value representations depending on mentalizing capacity. *Journal of Neuroscience, 42*(16), 3445–3460. https://doi.org/10.1523/jneurosci.1870-21.2022
Senci, C. M., Breccia, F., & Freidin, E. (2020). Attention to prescriptive norms increases dictator game generosity in women but not men: Using the 2D: 4D digit ratio to test the role of biology. *Journal of Psychology, 7*, 21. https://doi.org/10.12974/2313-1047.2020.07.2
Shirtcliff, E. A., Granger, D. A., & Likos, A. (2002). Gender differences in the validity of testosterone measured in saliva by immunoassay. *Hormones and Behavior, 42*(1), 62–69. https://doi.org/10.1006/hbeh.2002.1798
Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z
Singer, N., Sommer, M., Döhnel, K., Zänkert, S., Wüst, S., & Kudielka, B. M. (2017). Acute psychosocial stress and everyday moral decision-making in young healthy men: The impact of cortisol. *Hormones and Behavior, 93*, 72–81. https://doi.org/10.1016/j.yhbeh.2017.05.002
Sparrow, E. P., Armstrong, B. A., Fiocco, A. J., & Spaniol, J. (2019). Acute stress and altruism in younger and older adults. *Psychoneuroendocrinology, 100*, 10–17. https://doi.org/10.1016/j.psyneuen.2018.09.025
StanDevelopmentTeam. (2021). Stan modeling language users guide and reference manual. Retrieved from https://mc-stan.org.
Stanton, S. J. (2017). The role of testosterone and estrogen in consumer behavior and social & economic decision making: A review. *Hormones and Behavior, 92*, 155–163. https://doi.org/10.1016/j.yhbeh.2016.11.006
Starcke, K., Polzer, C., Wolf, O. T., & Brand, M. (2011). Does stress alter everyday moral decision-making? *Psychoneuroendocrinology, 36*(2), 210–219. https://doi.org/10.1016/j.psyneuen.2010.07.010
Staufenbiel, S. M., Penninx, B. W., Spijker, A. T., Elzinga, B. M., & van Rossum, E. F. (2013). Hair cortisol, stress exposure, and mental health in humans: A systematic review. *Psychoneuroendocrinology, 38*(8), 1220–1235. https://doi.org/10.1016/j.psyneuen.2012.11.015
Sýkora, Z. (2018). *Hormones and competitive behavior*. Master's thesis. Faculty of Social Sciences, Charles University. Retrieved from https://dspace.cuni.cz/bitstream/handle/20.500.11956/99536/120298438.pdf.
Talarovičová, A., Kršková, L., & Blažeková, J. (2009). Testosterone enhancement during pregnancy influences the 2D: 4D ratio and open field motor activity of rat siblings in adulthood. *Hormones and Behavior, 55*(1), 235–239. https://doi.org/10.1016/j.yhbeh.2008.10.010
Todd, P. M., Penke, L., Fasolo, B., & Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proceedings of the National Academy of Sciences, 104*(38), 15011–15016. https://doi.org/10.1073/pnas.0705290104
Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*(1), 35–57. https://doi.org/10.1086/406755
Van Honk, J., Montoya, E. R., Bos, P. A., Van Vugt, M., & Terburg, D. (2012). New evidence on testosterone and cooperation. *Nature, 485*(7399), E4–E5. https://doi.org/10.1038/nature11136
Vinkers, C. H., Zorn, J. V., Cornelisse, S., Koot, S., Houtepen, L. C., Olivier, B., … Joëls, M. (2013). Time-dependent changes in altruistic punishment following stress. *Psychoneuroendocrinology, 38*(9), 1467–1475. https://doi.org/10.1016/j.psyneuen.2012.12.012
Waynforth, D., & Dunbar, R. I. (1995). Conditional mate choice strategies in humans: Evidence from'Lonely Hearts' advertisements. *Behaviour*, 755–779. https://doi.org/10.1163/156853995X00135
West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior, 32*(4), 231–262. https://doi.org/10.1016/j.evolhumbehav.2010.08.001
Wiederman, M. W. (1993). Evolved gender differences in mate preferences: Evidence from personal advertisements. *Ethology and Sociobiology, 14*(5), 331–351. https://doi.org/10.1016/0162-3095(93)90003-Z
Zahavi, A. (1995). Altruism as a handicap: The limitations of kin selection and reciprocity. *Journal of Avian Biology, 26*(1), 1–3. https://doi.org/10.2307/3677205
Zak, P. J., Kurzban, R., Ahmadi, S., Swerdloff, R. S., Park, J., Efremidze, L., … Matzner, W. (2009). Testosterone administration decreases generosity in the ultimatum game. *PLoS One, 4*(12), Article e8330. https://doi.org/10.1371/journal.pone.0008330
Zethraeus, N., Kocoska-Maras, L., Ellingsen, T., Von Schoultz, B. O., Hirschberg, A. L., & Johannesson, M. (2009). A randomized trial of the effect of estrogen and testosterone on economic behavior. *Proceedings of the National Academy of Sciences, 106*(16), 6535–6538. https://doi.org/10.1073/pnas.0812757106
Zhang, Q., Ma, J., & Nater, U. (2019). How cortisol reactivity influences prosocial decision-making: The moderating role of sex and empathic concern. *Frontiers in Human Neuroscience, 13*, 415. https://doi.org/10.3389/fnhum.2019.00415
Zilioli, S., Ponzi, D., Henry, A., & Maestripieri, D. (2015). Testosterone, cortisol and empathy: Evidence for the dual-hormone hypothesis. *Adaptive Human Behavior and Physiology, 1*(4), 421–433. https://doi.org/10.1007/S40750-014-0017-X