# Testosterone administration decreases sensitivity to angry facial expressions in healthy males: A computational modeling approachTestosterone and threat perception

Yu Nan, Pranjal Mehta, Jiajun Liao, Yueyuan Zheng, Chengyang Han, Yin Wu


PII: S0306-4530(23)00926-5

DOI: https://doi.org/10.1016/j.psyneuen.2023.106948

Reference: PNEC106948

To appear in: *Psychoneuroendocrinology*

Received date: 9 August 2023
Revised date: 10 November 2023
Accepted date: 23 December 2023

Please cite this article as: Yu Nan, Pranjal Mehta, Jiajun Liao, Yueyuan Zheng, Chengyang Han and Yin Wu, Testosterone administration decreases sensitivity to angry facial expressions in healthy males: A computational modeling approachTestosterone and threat perception, *Psychoneuroendocrinology,* (2023) doi:https://doi.org/10.1016/j.psyneuen.2023.106948

Running title: Testosterone and threat perception

# Testosterone administration decreases sensitivity to angry facial expressions in healthy males: A computational modeling approach

Yu Nan [a, b], Pranjal Mehta [c], Jiajun Liao [d], Yueyuan Zheng [e], Chengyang Han [f], Yin Wu [a, g *]

[a] Department of Applied Social Sciences, Hong Kong Polytechnic University, Hong Kong SAR, China

[b] School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

[c] Department of Experimental Psychology, University College London, London, United Kingdom

[d] School of Psychology, South China Normal University, Guangzhou, China

[e] Department of Psychology, University of Hong Kong, Hong Kong SAR, China

[f] Department of Psychology, Hangzhou Normal University, Hangzhou, China

[g] Research Institute for Sports Science and Technology, Hong Kong Polytechnic University, Hong Kong SAR, China

[*] Address correspondence: Dr. Yin Wu, Department of Applied Social Sciences, Hong Kong Polytechnic University. E-mail: y.wu@polyu.edu.hk

## Highlights

• Exogenous testosterone attenuates sensitivity to facial threat.

• Exogenous testosterone group showed reduced sensitivity to angry facial expressions.

• Exogenous testosterone has a causal effect on men's perception of facial threat.

• The effects of exogenous testosterone on threat cues was primarily driven by angry facial expressions.



## Abstract

Previous research indicates that higher testosterone levels are related to increased aggressive and dominant behaviors, particularly in males. One possible mechanism for these hormone-behavior associations could involve threat perception. However, the causal influence of testosterone on men's recognition of threatening facial expressions remains unknown. Here, we tested the causal effect of exogenous testosterone on men's sensitivity to facial threat by combining a psychophysical task with computational modeling. We administered a single dose (150 mg) of testosterone or placebo gel to healthy young men (n = 120) in a double-blind, placebo-controlled, between-participant design. Participants were presented with morphed emotional faces mixing anger/fear and neutral expressions and made judgments about the emotional expression. Across typical regression analysis, signal detection analysis, and drift

diffusion modeling, our results consistently showed that individuals who received testosterone (versus placebo) exhibited a lower perceived sensitivity to angry facial expressions. But we observed no significant effects of testosterone administration on fearful facial expressions. The findings indicate that testosterone attenuates sensitivity to facial threat, especially angry facial expressions, which could lead to a misestimation of others' dominance and an increase in one's own aggressive and dominant behaviors.

**Keywords**

Testosterone; Social threat; Angry facial expressions; Fearful facial expressions; Emotion

## 1. Introduction

Testosterone is a major androgen hormone that has significant effects on individual growth and development, and it also affects many aspects of social behavior. The challenge hypothesis provides a theoretical framework for understanding the social effects of testosterone. This hypothesis was originally developed to describe intra- and interspecies differences in testosterone secretion among male birds (Wingfield et al., 1990). According to this hypothesis, testosterone levels should increase acutely during intermale contests such as competition for mates or status, and these testosterone increases should promote aggressive and dominant behaviors towards male competitors as well mating behaviors towards females.

In agreement with the challenge hypothesis, previous research has found a marked increase in mating and aggressive behavior following exogenous testosterone administration

across several bird species (i.e., the song sparrow, the house sparrow, and the European starling) (Peters, 2002). In addition, studies in several species of fish (Hirschenhauser et al., 2004), two types of lizards (Greenberg & Crews, 1990), and primates (such as the ring-tailed lemur, rhesus monkey, and chimpanzee) (Muller & Wrangham, 2004) found similar associations between hormonal and behavioral changes, providing evidence that the challenge hypothesis applies more broadly to other vertebrate species. In humans, competitive interactions lead to increases in men's salivary testosterone concentrations (Trumble et al., 2014), which trigger more aggressive and dominant behaviors towards rivals (Carré et al., 2009) as well as affiliative social interactions with potential mates (Van der Meij et al., 2012). Overall, these findings corroborate the challenge hypothesis in showing testosterone's links with mating, aggressive, and dominant behaviors particularly in males.

Despite growing support for testosterone's role in behaviors such as aggression and dominance, the mechanisms for these hormone-behavior associations remain unclear. One possible mechanism could involve threat perception of conspecifics, which serve to guide one's own social behavior. Indeed, accurate assessments of potential social threat could regulate fight-or-flight decisions that are critical for survival and reproduction (Geniole et al., 2020).

Facial expressions play a crucial role in human social and emotional behavior, including aggressive and dominant behaviors. In social interactions, individuals interpret others' affective states, motivations, intentions, and personalities through the emotional information conveyed by the facial expressions of others (Parkinson, 2005). Six basic emotions, namely happiness, anger, disgust, fear, sadness and surprise, are generally recognized in various cultural contexts (Elfenbein & Ambady, 2002). During social interaction, these expressions have the following

functions: first, they can provide emotional and intentional information of the expressor; second, they can elicit a response from the perceiver; and finally, they provide a reward for the desired social behavior (Keltner, 2003). Collectively, perceptions of facial expressions could serve as signals of others' emotional states that guide future behavior of perceivers (Keltner et al., 2003).

Of these expressions mentioned above, angry facial expressions are considered threatening social signals (Pichon et al., 2009). Anger represents imminent verbal or physical aggression from the expressor, with a high degree of unpleasantness and arousal (Adams Jr et al., 2003). As a threat signal, an angry facial expression could trigger fight-or-flight responses in the recipient, and testosterone may bias perceptions of this threat signal (Carré et al., 2017). In line with this theorizing, one study found that testosterone administration reduced women's accuracy in recognizing angry faces (Van Honk & Schutter, 2007). The authors interpreted their findings as suggesting that accurate perception of anger serves as a social corrective signal that guides one's own behavior, and testosterone disrupts perception of this signal. More specifically, the authors proposed that testosterone - by reducing accuracy in recognizing angry facial expressions - may lead individuals to misestimate the social threat posed by conspecifics, which in turn may increase one's own aggressive and dominant behaviors.

In line with this proposed mechanism, subsequent research found that female participants who received testosterone administration had a reduced tendency to avoid social threats, particularly towards angry facial expressions rather than neutral or happy expressions (Enter et al., 2014). Additional research found that individuals exhibit a *threat premium effect*: a behavioral tendency to cede more resources to others with high-threat facial expressions

compared to others with low-threat facial expressions (Geniole et al., 2017). Importantly, testosterone treatment reduced this threat premium effect, consistent with the view that testosterone reduces sensitivity to threatening facial expressions that guides one's own behavior (Geniole et al., 2019).

Overall, these findings suggest that testosterone may increase aggressive and dominant behaviors by reducing sensitivity to social threat cues. However, other studies reported results that cannot be reconciled easily with this proposed mechanism, suggesting instead that testosterone may enhance – rather than inhibit – sensitivity to social threat. For example, exogenous testosterone administration resulted in enhanced physiological and neural responses to angry faces (Goetz et al., 2014), and higher testosterone levels were thought to be associated with increased emotion recognition performance (Vongas & Al Hajj, 2017). Other research found non-significant associations between testosterone levels and recognition of angry facial expressions (Derntl et al., 2009; Rukavina et al., 2018). Given these inconsistencies across studies, testosterone's role in facial threat processing requires further investigation.

Several limitations of prior studies undermine their generalizability and may contribute to inconsistent results. First, many of these studies were correlational (Derntl et al., 2009). Confounding factors may produce inconsistent results in these studies, and such studies do not enable causal inference. Second, many of the studies relied on indirect measures to study sensitivity to emotional facial expressions (e.g., heart rate, gaze duration) (Terburg et al., 2012), and these approaches do not directly examine perceived facial threat. Third, most studies to date recruited relatively small samples (fewer than 20 participants in several cases), which could have produced false negative or false positive results (Button et al., 2013). Fourth, most

of the testosterone manipulation studies used female participants (Van Honk et al., 2005). Violent and aggressive behavior are more prevalent in men compared to women, the challenge hypothesis focuses on testosterone's role in male social behavior, and gender differences have been observed in testosterone's associations with behaviors such as aggression (Van Honk & JLG Schutter, 2007). Thus, it is important to examine how testosterone administration affects men's sensitivity to threatening stimuli. Fifth, some prior work used both male and female facial expressions as experimental stimuli(Van Honk & JLG Schutter, 2007), but the experimental stimuli of same-sex conspecifics to study social threat are more in line with the predictions of the challenge hypothesis. Finally, many of the studies examined testosterone's role in emotional processing with traditional statistical models alone (e.g., regression). Testosterone's influence on facial threat processing can be further understood by integrating computational modeling approaches (Konovalov et al., 2018), which can provide greater insight into the underlying mental process.

The present study addressed these limitations in a novel design by combining pharmacological hormone administration, computational modeling, and psychophysical approaches to examine how testosterone affects men's sensitivity to threat stimuli. We employed an emotion recognition task in which participants had to identify presented facial expressions of each trial as quickly and accurately as possible (Graham et al., 2007). We adopted morphed faces with an artificial combination of angry and neutral expressions, with 7 morph levels. We adopted three complementary analysis strategies: mixed-effect logistic regression analysis, signal detection analysis, and drift diffusion modeling. The two parameters (sensitivity and criterion) of signal detection theory (SDT) provide insight into the

psychological process of emotion identification beyond traditional regression analysis (Tsoi et al., 2008). Hierarchical Bayesian estimation of the drift-diffusion model (HDDM) (Wiecki et al., 2013) was also employed to assess the latent decision-making dynamics during emotion processing, including speed of evidence accumulation and decision threshold. HDDM could help us better understand the mental process of emotion recognition by modeling trial-by-trial reaction times (RTs) using several free parameters. Thus, in the present study, beyond typical regression analysis, we examined the process by which individuals perceived threatening facial expressions through a combination of model-based (SDT) and model-free (HDDM) approaches. Our primary research question focuses on anger, because prior work on testosterone has also focused on angry facial expressions as a threat cue. However, since fear and anger share some similarities in terms of potency and arousal, we included fear-related stimuli in the study design for control purposes. We hypothesized that single-dose testosterone administration would impact sensitivity in the perception of angry facial expressions. Given the inconsistencies in prior work and a lack of experimental research in men, we were agnostic about the direction of the effect.



## 2. Methods

### 2.1 Participants

One hundred and twenty healthy males (mean age = 19.98 years, *SD* = 1.55, age range = 18–25) were recruited through advertisement in a Chinese university. We screened participants through an online questionnaire to ensure that all of them were right-handed, had

normal vision and were not taking psychotropic medications or suffering from any psychiatric/neurological disorder. Twenty-four hours before the experiment, participants were asked not to take any alcohol, caffeine or smoke. The study employed a double-blind, placebo-controlled, between-participant design, in which each participant randomly received either a single dose of Androgel or placebo gel. Two participants (both in the placebo condition), whose data contained more than 30% of invalid trials (Brennan & Baskin-Sommers, 2020), were not included in the final analysis. We collected the informed consent signed by all participants and paid them a participation fee of 150 Chinese Yuan (~$21). This study was approved by local research ethics committee and was conducted in accordance with the Declaration of Helsinki.

### 2.2 Testosterone Administration

The experiment started at 13:00 in the afternoon and ended around 17:00. A male research assistant, who was blind to both the purpose of the study and experimental condition, applied the gel to the shoulders and upper arms of participants. Participants in the testosterone group received 150 mg of testosterone (Androgel®), while those in the placebo group received a colorless hydroalcoholic gel (the placebo gel had identical an appearance to the testosterone gel). We administered the emotional recognition task 3 hours post-dosing. The dose and timing of the task were selected based on previous pharmacokinetic data, which showed that testosterone concentrations peaked 3 hours after receiving 150 mg of testosterone (Eisenegger et al., 2013) but see (Bird et al., 2016). Participants took a rest during the 3-hour waiting period. During the waiting period, participants remained in the testing rooms and were given unrelated newspapers and magazines to ensure they were not exposed to content pertinent to the current

study. In addition to the emotion recognition task, participants also completed two additional tasks unrelated to the present study.

### 2.3 Stimuli

To prepare the stimuli of the experiment, we recruited 42 male university students as actors and photographed the stimuli for the task. We asked participants to take emotional photos (including angry, fearful, and neutral expressions) in the same environment, location, and lighting conditions. The final 126 images of facial expressions were standardized using PhotoShop, including the same parameters such as brightness and contrast of the pictures. Then the background of all the pictures was changed to black. We also concealed the extra parts of the actor's hair and clothes, which could affect the participants' decision-making. We only kept an oval face as the experimental material (see Fig. 1A). After normalization, we recruited another 48 male university students to identify and rate the 126 images. Finally, faces of 6 actors were selected for the stimulus set used in the following task, and the average recognition accuracy of the three emotional faces of the six actors was over 0.8 (anger: mean correct rate = 0.81, *SD* = 0.09; fear: mean correct rate = 0.81, *SD* = 0.08; neutral: mean correct rate = 0.83, *SD* = 0.04).

Stimuli were generated by blending the two images using face-morphing software (Abrosoft, 2018, FantaMorph for Windows, Version 5.5.0) to create two types of emotion blends: anger-neutral blends and fear-neutral blends. The morphed pictures were equally distributed between the neutral and the angry/fearful expressions and resulted in 7 morph levels (20% anger to 80% anger, 20% fear to 80% fear, in 10% increments)(Wudarczyk et al., 2016).

### 2.4 Emotion-Recognition Task

The study used a modified emotion recognition task (Graham et al., 2007), which was programmed using Presentation software (Neurobehavioral System Inc.). The final 84 images (2 blends of 6 actors, 7 morph levels) were presented in a randomized order, and each image was presented twice (168 trials in total). The task consisted of two blocks; all facial expression images were presented for the first time in the first block and a second time in the second block. In each block, participants viewed both emotional blends: the anger-neutral blend followed by the fear-neutral blend. At the end of each emotional blend (after viewing all morph levels for either anger-neutral or fear-neutral), participants were provided with a 30-second rest period. After completing both emotional blends in the first block, a longer two-minute rest period was allocated to minimize visual fatigue effects before proceeding to the second block.

For each trial (see Fig. 1B), a fixation point was first presented on the screen with a random display time between 500 ms and 1000 ms. Following that, emotional pictures were presented, and participants were asked to use the “F” and “J” keys on the keyboard within 3 seconds to identify the expression on each face as quickly and as accurately as possible. While “F” represented the expression on the bottom left of the screen, “J” represented the expression on the bottom right of the screen (the keys were counter-balanced within participants). To better understand the task, the participants were asked to practice before the formal experiment, and the emotional face materials used in the practice trials were different from the formal experiment. No result feedback was given after selection in each trial.

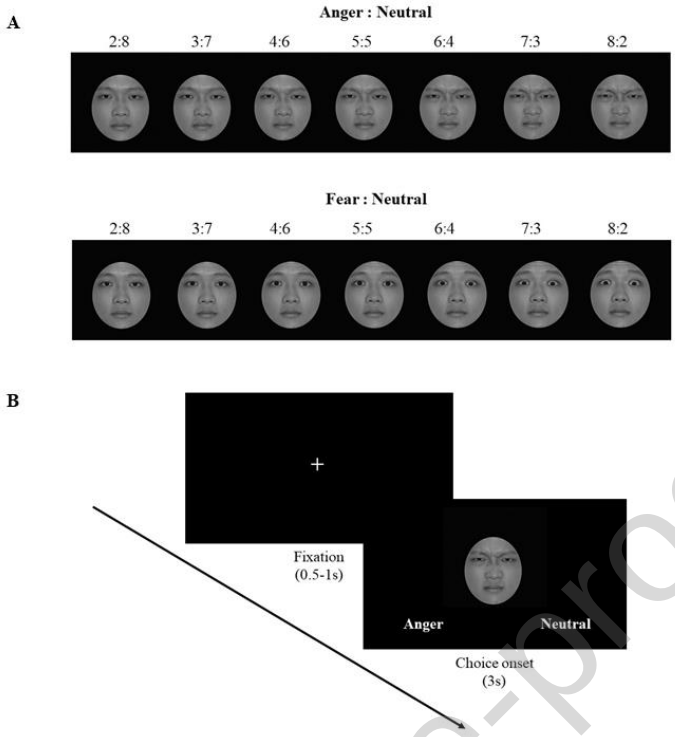


**Figure 1.** (A) Sample task stimuli. (B) Emotion recognition task

## 2.5 Data Analysis

First, we combined anger-neutral and fear-neutral trials into a mixture-neutral blend, where anger and fear were unified into a mixture (i.e., a mixture-neutral blend). Next, we performed mixed-effects logistic regression analysis, signal detection analysis, and drift-diffusion modeling for mixture-neutral blend, anger-neutral blend, and fear-neutral blend, respectively. We primarily focus on the anger-neutral blend, while the fear-neutral blend and mixture-neutral blend serve as control conditions to complement the findings of the anger-neutral blend.

### 2.5.1 Emotion recognition

Ninety-one trials (0.9% of the total trials) were excluded from the analysis because participants did not respond or responded within 300 ms (Brennan & Baskin-Sommers, 2020). We performed a mixed-effect logistic regression analysis on emotion recognition using R (R version 3.5.1; http://www.r-project.org/). The glmer function in the lme4 package was used for parameter estimates (Bates et al., 2014), and the *p*-values were obtained using the lmerTest:: Summary function in Satterthwaite approximations (Kuznetsova et al., 2015). In addition, we reported odds ratios (ORs) with 95% confidence intervals (95%CIs) for behavioral analyses.

In the mixture-neutral blend, we first performed a mixed-effects logistic regression to exam the effects of treatment, morph level, emotion type (fear-neutral vs. anger-neutral) and their interactions on emotional perception. In the anger-neutral blend and fear-neutral blend, the mixed-effects logistic regression examined the effect of testosterone, morph level, and their interaction on emotional perception. Specifically, we expected participants' selection of expression to shift from neutral to anger with increasing morph level. In each model, the dependent variable was participant's choice (i.e. choosing neutral or choosing anger/fear).

In the model, the fixed effect factor was treatment (categorical), and the fixed effect predictor was morph level (continuous). Additionally, subject ID (*i*; random item: *u*) and face ID (*j*; random item: *w*) were entered into the model as random effect factors. To meet the maximum random structure, we added the random subject intercepts, random face intercepts and by-subject random slope to the model. The model consisted of the following variables across the *k* morph levels for participant *i* with recognizing the actor's face *j*:

**Level 1:**

$$\text{logit}(p(anger/fear))_{ijk} = \beta_{0ij} + \beta_{1ij} * \text{morph level}_k + \beta_2 * \text{group}_k + \beta_3 * \text{morph level}_k * \text{group}_k + \varepsilon_{ijk}$$

**Level 2:**

$$\beta_{0ij} = \gamma_{00} + u_{0i} + w_{0j}$$

$$\beta_{1ij} = \gamma_{10} + u_{1i}$$

$$u_{0i} \sim N(0, \sigma_{u0}^2)$$

$$w_{0j} \sim N(0, \sigma_{w0}^2)$$

$$u_{1i} \sim N(0, \sigma_{u1}^2)$$

### 2.5.2 Signal detection analysis

Signal detection theory is a psychophysical approach to conceptualize the process of facial emotion recognition (Gillespie et al., 2015). Most decisions made by an individual contain some degree of uncertainty (McNicol, 2005), whereby people make decisions based on available evidence. Signal detection theory attempts to explain this process. It reflects the two comprehensive psychological processes of decision-making, the sensory process and the cognitive decision-making process, by providing separate measures of performance in decision making (Krantz, 1969). The sensory process corresponds to sensitivity ($d'$), which determines the observer's ability to select the correct stimulus while avoiding the wrong stimulus (Tsoi et al., 2008). An increase in the value of $d'$ refers to greater sensitivity to a given signal or expression. The response criterion ($\beta$) reflects an individual's inclination to make a certain decision based on evidence obtained from the sensory process. Criterion ($\beta$) is primarily used to measure cognitive decision-making processes (Gillespie et al., 2015).

We employed signal detection analysis to examine whether testosterone influenced sensitivity $d'$ and the criterion $\beta$ in the mixture-neutral blend, anger-neutral blend, and fear-neutral blend (Graham et al., 2007). Among these three blends, emotional facial expressions (anger vs. fear vs. mixture) as signals and neutral facial expressions as noise defined four types of response in this task. When the intensity of anger/fear/mixture was strong (morph levels: 60%-80%), the participants chose anger/fear/mixture as Hit, and neutral as Miss. When the intensity of neutral was strong (morph levels: 20%-40%), anger/fear/mixture was selected as False Alarm, and neutral as Correct Rejection. The moderate level (morph level of 50%) was excluded from the analysis because the trials in this level could not be classified as Hit or False Alarm. Accordingly, we calculated the hit rate and the false alarm rate of each participant, and obtained the corresponding $O$ value and $Z$ value through transformation. Then the $d'$ and $\beta$ of each participant were derived (see the formula as below). Finally, an independent samples $t$ test was used to examine group differences.

$$P(Hit) = \frac{N(Hit)}{N(Hit) + N(Miss)}$$

$$P(False\ Alarm) = \frac{N(False\ Alarm)}{N(False\ Alarm) + N(Correct\ Rejection)}$$

$$d' = Z(Hit) - Z(False\ Alarm)$$

$$\beta = O(Hit)/O(False\ Alarm)$$

### 2.5.3 Hierarchical Drift Diffusion modeling

HDDM has been widely used to estimate 2-choice decision-making processes (Ratcliff

& McKoon, 2008), in particular in evaluating dynamic decision processes involving perception and decision-making (Ratcliff et al., 2016). We combined decision data with choice reaction time data using a hierarchical drift-diffusion model in the Python toolbox HDDM 0.6.0 (Wiecki et al., 2013). The Hierarchical Bayesian framework is an ideal analytical method for this study because it can recover parameters with fewer trials and the estimates are less susceptible to outliers, thereby improving statistical power (Wiecki et al., 2013). The model parameters for each participant were extracted from group-level distributions, and both group and participant level parameters were estimated using Bayesian Markov Chain Monte Carlo (MCMC) sampling. Decision threshold ($a$), drift rate ($v$), non-decision time ($t$) and initial bias ($z$) were parameterized to fit reaction time and participant's choice in this model. Threshold ($a$) represented the degree of separation of the decision boundary, which reflected the extent of prudence in decision-making processes. Smaller values of the boundary reflect a more impulsive decision-making style (i.e., faster, less accurate). The drift rate ($v$) was considered as the rate of evidence accumulation and provided an estimate of the speed of the evidence accumulation process. The larger the drift rate, the faster the individual's effective accumulation of emotional information. Non-decision time ($t$) referred to the reaction time phases unrelated to information accumulation processes, such as perceptual and motor response execution. The initial bias ($z$) was the starting point of the value accumulation process. A value away from 0.5 reflects a preference for Option A (anger; if > 0.5) or a preference for Option B (neutral; if < 0.5) (Brennan & Baskin-Sommers, 2020).

In the present study, we used HDDM to model the trial-by-trial RTs and binary choice to examine how participants accumulated evidence for emotional faces at different morph

levels during decision-making. Here, the choices of angry/fearful/mixed expressions were coded as "1", and the choices as neutral expressions was coded as "0". Reaction times (RT) shorter than 0.3 s were excluded (less than 1% of all trials)(Brennan & Baskin-Sommers, 2020). We estimated models with 15,000 samples and the first 5000 samples were discarded as burn in (Tipples, 2018). In addition, the model's deviation information criterion (DIC) was used to evaluate the goodness of model fit to help us choose the best model. The lower DIC value indicated the better model fit (Spiegelhalter et al., 2002), and a difference of more than 10 DIC values was considered as statistically significant. Estimating the model in a Bayesian framework allows us to perform significance tests directly on the posterior.

### 2.6. Power simulations

The sample size for this research was determined by power to detect between-group differences in a pharmacological treatment experiment and was the largest possible within the experiment's budget and resource constraints. The power simulations conducted after data collection indicate that, with our intended sample size of 120 participants and employing a within-subjects approach, the experiment achieved 80% power to detect a three-way interaction term equivalent to log(B) = 0.7, OR = 2.0, which corresponds to a moderate effect size (Figure S1, see Supplemental Materials).

### 2.7. Data availability and pre-printing

Data and materials of this study are available on OSF at https://osf.io/k59m7/. The study

design and analysis were not pre-registered.

## 3. Results

### 3.1 Emotion Recognition using Mixed-Effects Logistic Regression Analysis

In the mixture-neutral blend, there was a significant main effect of treatment (OR = 1.67, *95%CI* [1.07, 2.60], $p < 0.05$) and a significant two-way interaction between treatment and morph level (OR = 0.40, *95%CI* [0.17, 0.94], $p < 0.05$) (see Fig. 2A). But we did not find a significant three-way interaction (OR = 1.67, *95%CI* [0.71, 3.94], $p = 0.237$)). Although no significant three-way interaction was observed in the mixture-neutral blend, we considered it essential to conduct separate analyses for the anger-neutral and fear-neutral blends. Given the biological and psychological contexts, anger and fear as distinct emotions may possess varying physiological and cognitive mechanisms. Consequently, they might interact differently with the treatment. To elucidate any nuanced differences in these interactions, we opted to further investigate the interactions between each emotion and the treatment, maintaining this analytical approach in subsequent methodologies.

In the anger-neutral blend, there was a significant main effect of morph level (*OR* = 2.18, *95%CI* [2.05, 2.32], $p < 0.001$) such that participants were more likely to recognize the facial expressions as angry expressions with increasing morph level (i.e., higher intensity of anger). The main effect of treatment was significant (*OR* =1.52, *95%CI* [1.03, 2.25], $p < 0.01$), indicating that men given testosterone identified expressions as anger at a higher rate compared to the placebo group. Importantly, the interaction between treatment and morph level was

significant (*OR* = 0.91, *95%CI* [0.84, 0.99], $p < 0.01$). Specifically, participants who received the testosterone treatment were more likely to classify images with lower anger intensity as 'angry' compared with those on the placebo. Conversely, they were more inclined to classify images with higher anger intensity as 'neutral' than those receiving the placebo (see Fig. 2B).

In the fear-neutral blend, there was a significant main effect of morph level (*OR* = 2.93, *95%CI* [2.67, 3.23], $p < 0.001$), indicating that the proportion of expressions recognized as fear increased with increasing intensity of the fearful emotion. Neither the main effect of treatment (*OR* = 1.23, *95%CI* [0.74, 2.04], $p = 0.427$) nor the interaction between treatment and morph level (*OR* = 0.96, *95%CI* [0.84, 1.09], $p = 0.492$) was significant (see Fig. 2C).

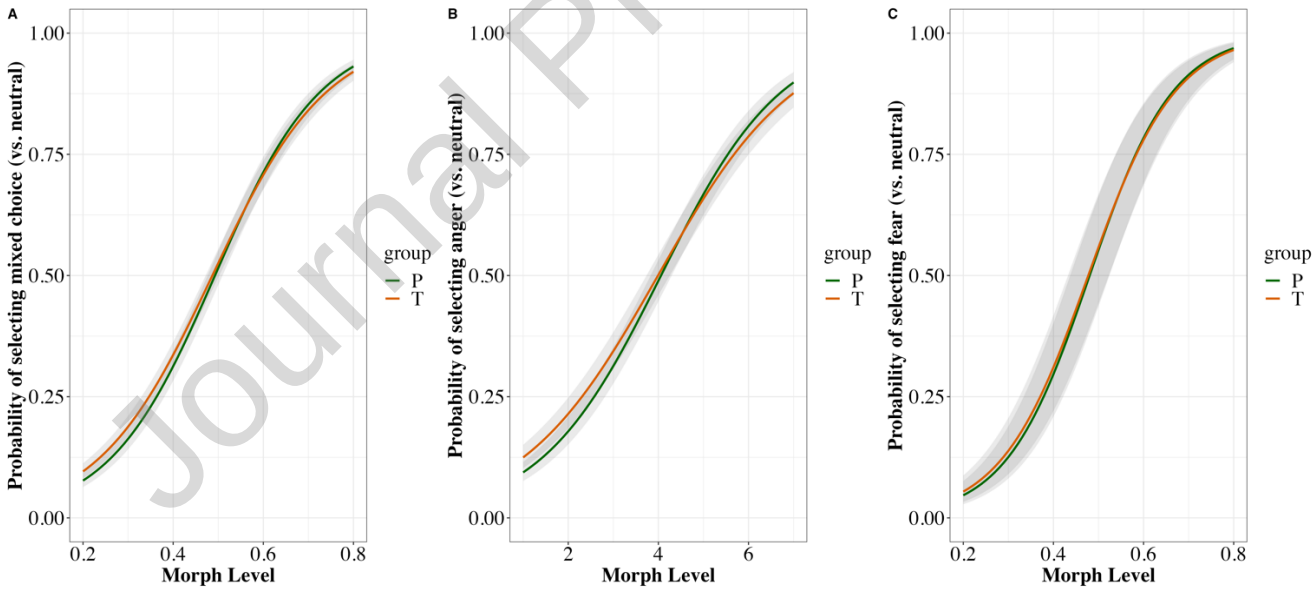


Figure 2. The effect of treatment on the relationship between morph level and the probability of recognizing the (A) mixed facial expressions, (B) anger facial expressions and (C) fear facial expressions.

### 3.2 Signal Detection Analysis

In the mixture-neutral blend, we found that the sensitivity *d'* in the testosterone group ($M$ = 1.78, $SD$ = 0.60) was significantly lower than that in the placebo group ($M$ = 1.94, $SD$ = 0.55; $t$ = 1.98, $p < 0.05$, $d$ = -0.15, *95% CI* [0.00, 0.31]) (see Fig. 3A). We did not find any significant differences in $\beta$ ($t$ = 0.87, $p$ = 0.385) between the two groups(see Fig. 3B).

In the anger-neutral blend, an independent sample *t*-test revealed that the sensitivity *d'* in the testosterone group ($M$ = 1.50, $SD$ = 0.50) was significantly lower than that in the placebo group ($M$ = 1.74, $SD$ = 0.41; $t$ = -2.71, $p < 0.01$, $d$ = 0.5, *95% CI* [0.06, 0.41]) (see Fig. 3C). There was no significant difference in the response criterion $\beta$ between the two groups ($t$ = -0.68, $p$ = 0.50). The results suggested that testosterone reduced individuals' sensory sensitivity between angry and neutral emotional expressions, while the judgment criterion being strict or lax in judging emotions was not affected by the treatment (see Fig. 3D).

In the fear-neutral blend, we did not find any significant differences in *d'* ($t$ = 0.57, $p$ = 0.57) and $\beta$ ($t$ = 0.57, $p$ = 0.57) between the two groups.

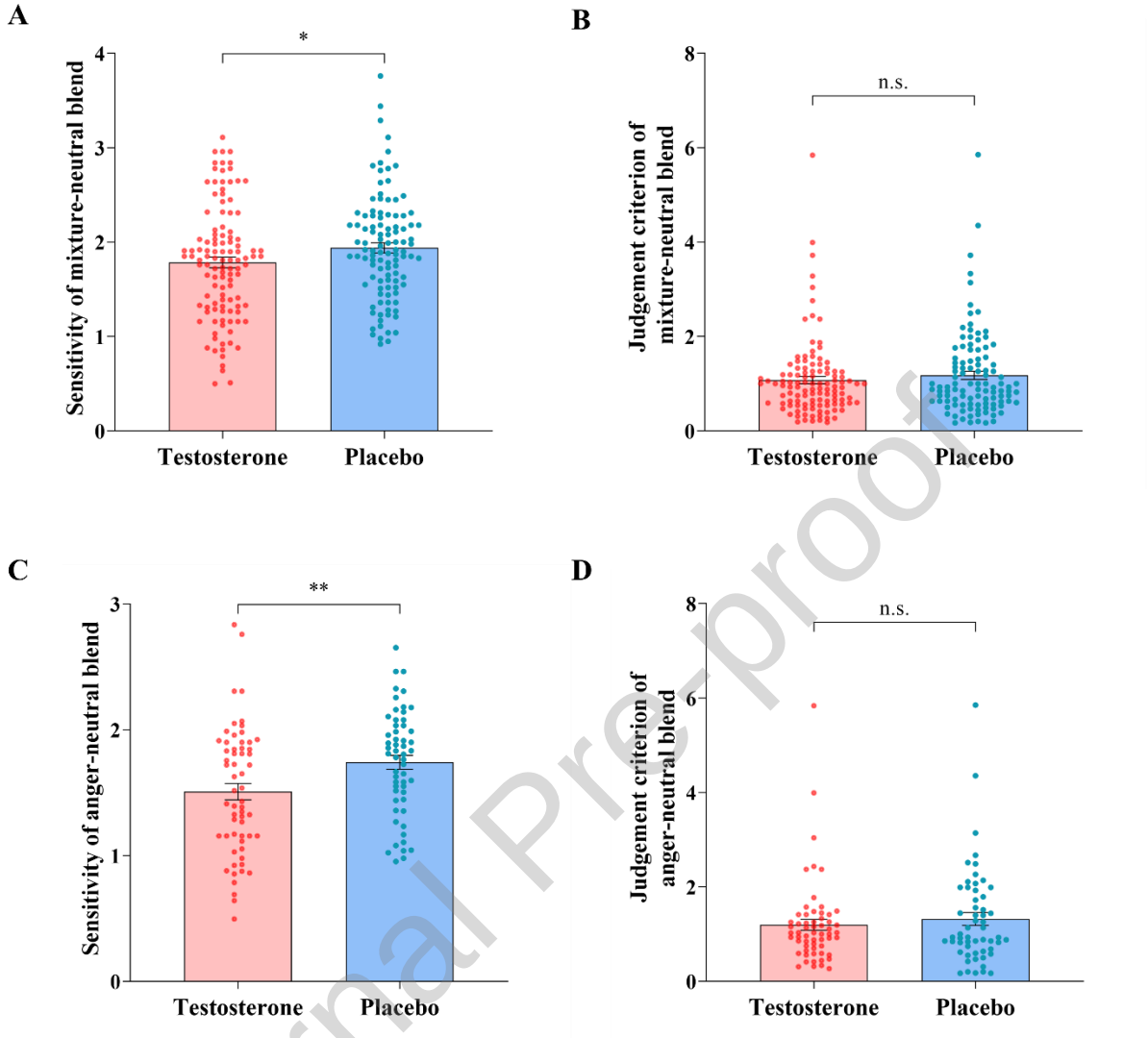


Figure 3. The sensitivity (*d'*) of testosterone group and placebo group to (A) mixed facial expressions and (C) angry facial expressions. The greater the *d'*, the more sensitive. $D' = Z$ (hit rate) – $Z$ (false alarm rate), where function $Z$(p), $p \in [0,1]$. The judgment criterion ($\beta$) of testosterone group and placebo group to (B) mixed facial expressions, (D) angry facial expressions. The greater the $\beta$, the stricter the judgment. $B = O$ (hit rate) / $O$ (false alarm rate). *$p < 0.05$, **$p < 0.01$

### 3.3 Hierarchical Drift Diffusion Modeling

In the mixture-neutral blend, participants who received testosterone showed lower drift

rate compared to those who received placebo (posterior probability of group difference in drift rate > 0.98; higher values indicate a larger difference between conditions; see Fig. 4A and 4B). There was no significant difference in initial bias and non-decision time between the two groups.

In the anger-neutral blend, we fist compared the models by computing the values of DIC and found that the best model (model 7, DIC = 16871; see Supplementary Table S1) consisted of the drift rate ($v$), non-decision time ($t$) and initial bias ($z$) varying across anger condition and neutral conditions, and the drift rate ($v$), non-decision time ($t$), and initial bias ($z$) were modulated by trial-by-trial decision RTs and binary choice. The result of the model showed that participants who received testosterone showed lower drift rate compared to those who received placebo (posterior probability of group difference in drift rate > 0.99; higher values indicate a larger difference between conditions; see Fig. 4C and 4D). There was no significant difference in initial bias and non-decision time between the two groups. These results suggested that individuals in the placebo group were more effective at accumulating anger-related information with increasing morph levels of emotional face, and testosterone administration resulted in a lower rate of accumulation of the information (placebo: mean ± SD = 4.66 ± 0.41; testosterone: mean ± SD = 4.04 ± 0.55; $b$ = -0.62, $SE$ = 0.06, $t$ = -10.24, $p$ < 0.001).

In the fear-neutral blend, we did not observe a significant effect of treatment on participants' evidence accumulation on fearful expressions (see Fig. 4E and 4F), there was no significant difference on the other two parameters between groups.

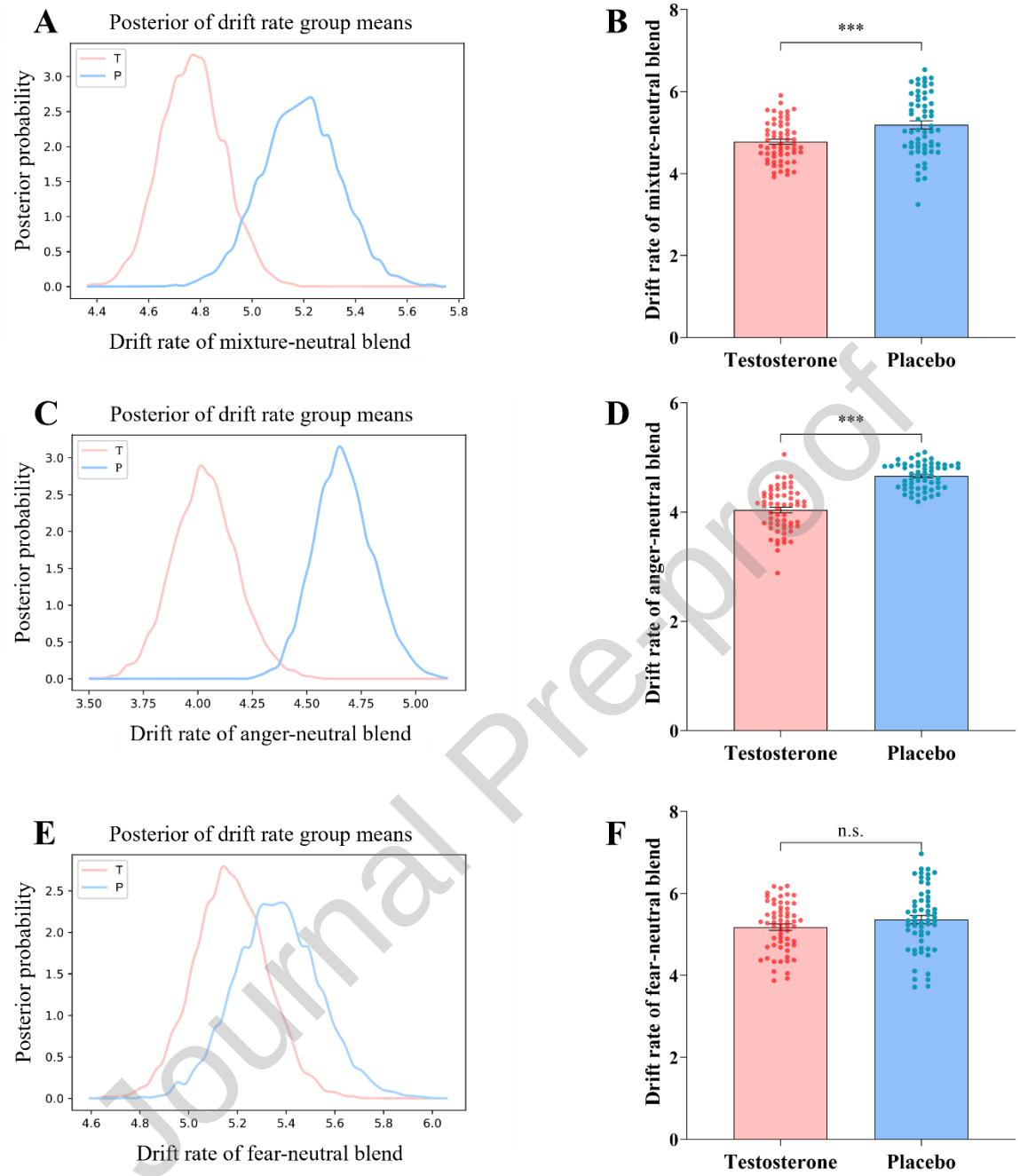




Figure 4. (A) Posterior density plot of the group means of the 2 different drift-rates *v* as produced by the hddm.analyze.plot_posterior_nodes() function in the (A) mixture-neutral blend, (C) anger-neutral blend and (E) fear-neutral blend. Regions of high probability are more credible than those of low probability. Bar graphs depict parameters of drift rate in testosterone and placebo groups, with better performance in males who received placebo than males who received testosterone in the (A) mixture-neutral blend and (C) anger-neutral blend. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

## 4. Discussion

In the present study, we tested the effect of exogenous testosterone on men's sensitivity in identifying facial threat expressions using an emotion recognition task. In the regression analysis that includes both anger and fear trials, we found a significant two-way interaction of testosterone administration and morph level. This two-way interaction provided a rationale for conducting signal detection analysis and drift diffusion models by integrating anger-neutral and fear-neutral trials into a unified analysis, without making a distinction between emotion type (anger or fear) in these analyses. The findings revealed that testosterone significantly decreased the drift rate and significantly reduced sensitivity (*d'*). Although we did not find a statistically significant three-way interaction in this exploratory analysis, this was like due to insufficient power to detect an attenuated interaction (effect present in one condition but not the other).

Despite not finding a statistically significant three-way interaction, we proceeded to conduct separate exploratory analyses for anger and fear trials, drawing inspiration from the findings of relevant studies (Adams Jr et al., 2003; Derntl et al., 2009; Van Honk & JLG Schutter, 2007) and from statisticians' recommendations to focus on effect estimation rather than statistical significance alone (Cumming, 2014). First, results of the mixed-effect logistic regression showed that participants were more likely to recognize anger with increasing morph level; crucially, this effect of morph level on anger recognition was significantly weaker in the testosterone group relative to the placebo group. Second, the sensitivity index of the signal detection analysis directly demonstrated that testosterone reduced individual's sensitivity to angry facial expressions. Last, the results from HDDM showed that testosterone administration

reduced the rate of evidence accumulation for angry facial expressions. These findings extend previous research conducted in women, which provided initial evidence that testosterone reduces sensitivity for consciously recognizing facial threat in a sample of 16 volunteers (Van Honk & Schutter, 2007). The present experiment demonstrates a similar effect in men using more rigorous methods, including a larger sample size and computational modeling approaches to provide greater insight into the underlying mental process. The results of this approach strongly suggested that the observed two-way interaction was primarily influenced by anger trials. Based on this exploratory analysis combined with our primary analyses focused on anger, we tentatively conclude that testosterone administration reduced sensitivity to threat cues, primarily driven by a diminished sensitivity to angry emotional expressions. Nevertheless, we recommend follow-up studies with even larger sample sizes to draw more definitive conclusions about emotion specificity.

Our findings extend the challenge hypothesis in suggesting that reduced facial threat sensitivity is a plausible mechanism for testosterone's influence on aggression and dominant behaviors. For example, Geniole et al. (2019) found that men exhibit a *threat premium effect*: a behavioral tendency to cede resources to other men with high-threat faces compared to other men with low-threat faces; crucially, testosterone treatment reduced this threat premium effect. The findings from the present experiment suggest that elevated testosterone levels might reduce this threat premium behavioral effect by reducing perceived sensitivity to facially threatening faces, which in turn may lead to a misperception of others' dominance and an increase in one's own aggressive and dominant behaviors. In sum, decreased sensitivity to facial threat caused by exogenous testosterone might lead individuals to misestimate the threat posed by opponents

and therefore to engage in more confrontational and aggressive behaviors in competition. More broadly, our data provide an important theoretical advance in identifying a social cognitive mechanism that may guide testosterone's role in conflict escalation.

The reason for decreased sensitivity to angry facial expressions following exogenous testosterone could involve one's own self-perceived dominance. A previous study found that exogenous testosterone increased men's perceptions of their own physical dominance compared to placebo (Welling et al., 2016). Furthermore, other studies suggested that men's dominance sensitivity was negatively correlated with their perception of their own dominance (Watkins et al., 2010), which has also been corroborated in a Chinese sample (Han et al., 2021). It Is worth noting that angry emotional faces were considered as a typical social dominance signal (Terburg et al., 2012). Therefore, in our study, the decrease in sensitivity to others' angry facial expressions caused by elevated testosterone levels might be due to the increase in self-perceived physical dominance. Future work can examine whether the effect of testosterone on anger sensitivity is mediated by self-perceived physical dominance.

Previous studies found that testosterone increased emotional responses (e.g., heart rate) (van Honk et al., 2001) and gaze fixation (Terburg et al., 2012) on angry facial expressions, which may look inconsistent with the findings of this study. The reason for the discrepancy could be due to the different levels of facial expression processing (Van Honk et al., 2005). Previous studies have focused on the sensory and motor components of emotional responses based on unconscious processing of emotional stimuli (before emotional sensory states). In contrast, the present study engages higher level of conscious recognition (in the emotional sensory state), focusing on the attention to the emotional state and emotional content of others

(Lane, 2000). At the neural level, conscious processing involves additional networks compared to unconscious processing. Structures such as the amygdala, thalamus, and superior colliculus are recruited to mediate the automatic and unconscious processing of basic emotional stimuli (Gainotti, 2012). Higher-level conscious processing of emotions is closely related to the cortico-pulvinar-cortical pathway, in which some structures connected or related to the pulvinar play an important role (Shobe, 2014). The role of testosterone in the two emotional processing levels may be different, the former may reflect the associations of testosterone level and the reinforcing qualities of these subthreshold (i.e., unconscious processing of emotional stimuli) angry faces (Wirth & Schultheiss, 2007), while the present study reflects the effect of testosterone on conscious vigilance to threatening stimuli.

We observed no significant effects of testosterone administration on fear perception in the separate analyses that examined fear trials. Some possible reasons are as follows. First of all, while fearful facial expressions have also been identified as threatening stimuli in previous studies (Anderson et al., 2013; LoBue & Rakison, 2013), they are not be the same source of threat as those represented by angry facial expressions. Specifically, angry facial expressions represent a direct personal threat to the observer, whereas fearful facial expressions may indicate indirect environmental threats which require further exploration (Grillon & Charney, 2011). In addition, evidence from adult fMRI studies suggest that differences in perception of anger and fear are engaged by different brain networks. Specifically, the amygdala was more responsive to fear compared to anger (Whalen et al., 2001), while angry facial expressions activated the insula, thalamus, cingulate, basal ganglia, and hippocampus (Strauss et al., 2005). Testosterone may also have different effects on different neural circuits. Future researchers can

investigate the neural mechanisms by which testosterone plays a role in threat perception.

The present findings have some clinical relevance. Recent studies employing drift diffusion modeling (DDM) have deepened our understanding of threat perception biases associated with social anxiety disorder (SAD). For instance, faster evidence accumulation during reward-based decisions has been observed in SAD patients, suggesting an enhanced reward motivation within this group (Dillon et al., 2022). Biased perceptual judgments in individuals with anxiety, influenced by prior knowledge of threats, have also been documented (Ozturk et al., 2023). Further contributing to this body of knowledge, an interactive effect of anxiety and social cues on aggression perception has been reported (Silva et al., 2023), and the heightened prioritization of danger-related social signals has been linked to threat-induced anxiety (Beaurenaut et al., 2023). These DDM investigations consistently underscore a pattern of altered threat processing in SAD, manifesting as hypersensitivity and an enhanced attentional focus on potential threats (Beaurenaut et al., 2023; Dillon et al., 2022; Silva et al., 2023). However, diverging from these patterns, our current findings indicate that administration of exogenous testosterone attenuates threat sensitivity during emotion recognition tasks. Thus, while the extant DDM literature establishes a clear association between social anxiety symptoms and perceptual biases towards threats, our study pioneers in demonstrating the therapeutic potential of testosterone in mitigating such biases by normalizing heightened threat reactivity in SAD patients. Previous studies have also shown that social anxiety disorder (SAD) is associated with lower levels of endogenous testosterone (Giltay et al., 2012). A follow-up pharmacological study in women demonstrated that single-dose testosterone administration reduced gaze avoidance in SAD (Enter et al., 2016). Consistent

with our current findings, this evidence supported the theories on the dominance-enhancing effects of exogenous testosterone. Furthermore, our results suggest a psychological mechanism for the social anxiolytic effects of exogenous testosterone (Moukheiber et al., 2012). That is, we found that elevated testosterone levels reduced the individual's perceived sensitivity to the dominance of others, which could lead to a decrease in avoidance behavior (e.g. submission) and an increase in approach behavior (e.g. dominance). Thus, the symptoms of social anxiety may be alleviated by raising testosterone levels, and this may be mediated by changes in threat perception. Future research could use testosterone as a pre-clinical treatment to treat or relieve social anxiety disorders and examine threat perception as a mechanism for anxiety reduction. Besides, to further unravel the intricacies of testosterone's therapeutic role, we advocate for combined pharmacological research with DDM methodologies, aiming to pinpoint the precise mechanisms by which testosterone moderates threat perception biases in social anxiety.

Our study design and results raise some issues that warrant further investigation. First, we only tested male participants in this study since pharmacokinetics of testosterone gel is only established among male participants (Eisenegger et al., 2013). Future studies are needed to replicate these results in a larger sample of women (Goetz et al., 2014). Second, the experimental stimuli viewed by the participants in this study were all from same-sex expressors. Although the influence of intersexual confounding factors could be ruled out, the influence of testosterone on the sex perception of expressors was ignored. Future research could consider adding women's facial expressions to examine whether there are differences in perceptual sensitivity. Third, different from the results of Carré et al. (2017), in which testosterone concentrations increased within 60 minutes and persisted for 3 hours after administration, our

experiments were performed 3 hours after Androgel administration (Wu et al., 2018). We encourage future studies to clarify the time course of testosterone gel administration and the underlying neural mechanisms. While our study focused on the effects of exogenous testosterone administration, it's crucial to note that these might differ from the behaviors driven by endogenous levels. The observed post-administration behaviors may not truly mirror those influenced by sustained endogenous testosterone. Our results mainly capture immediate responses to testosterone, missing potential long-term effects from consistent levels. Moreover, our controlled lab environment may not reflect the complexities of real-world settings where endogenous testosterone levels interact with various contextual factors. Finally, our results suggest that exogenous testosterone's effects on emotion recognition are primarily driven by anger but not fear. Although the sample size in this study is large relative to many prior studies and provided adequate power to examine our primary research questions, even higher-powered studies will be required to detect higher-order interactions in order to establish definitively that the effects are specific to anger.

In conclusion, by combining computational modeling and psychophysical approaches in an emotion recognition task, we found that exogenous testosterone administration reduces sensitivity to angry facial expressions. The decreased sensitivity to other's facial threats could lead individuals to misestimate others' dominance and thus increase one's own aggressive behavior, a social cognitive mechanism that extends the Challenge Hypothesis.

**Conflict of interest statement**

The authors declare that they had no conflict of interest with respect to their authorship or

the publication of the article.

## Role of the funding source

This work was supported by The Hong Kong Polytechnic University Interdisciplinary Large External Project (P0044755), APSS Fund (P0046091), Research Institute for Sports Science and Technology (P0043556), and Guangdong Natural Science Fund (2023A1515012362). The funding sources had no further role in the study design, data collection, analysis, interpretation, and decision to submit this manuscript for publication.

## Acknowledgements

The authors are grateful to Mr. Jianxin Ou and Mr. Junsong Lu for their help with the study.

## Reference

Adams Jr, R. B., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science*, *300*(5625), 1536-1536.

Anderson, A. K., Christorf, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. (2013). Neural correlates of the automatic processing of threat facial signals. In *Social neuroscience* (pp. 185-198). Psychology Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bird, B. M., Welling, L. L., Ortiz, T. L., Moreau, B. J., Hansen, S., Emond, M., Goldfarb, B., Bonin, P. L., & Carré, J. M. (2016). Effects of exogenous testosterone and mating context on men's preferences for female facial femininity. *Hormones and behavior*, *85*, 76-85.

Brennan, G. M., & Baskin-Sommers, A. R. (2020). Aggressive realism: More efficient processing of anger in physically aggressive individuals. *Psychological science*, *31*(5), 568-581.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), 365-376.

Carré, J. M., Putnam, S. K., & McCormick, C. M. (2009). Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. *Psychoneuroendocrinology*, *34*(4), 561-570.

Carré, J. M., Ruddick, E. L., Moreau, B. J., & Bird, B. M. (2017). Testosterone and human aggression. *The Wiley handbook of violence and aggression*, 1-14.

Cumming, G. (2014). The new statistics: Why and how. *Psychological science*, *25*(1), 7-29.

Derntl, B., Windischberger, C., Robinson, S., Kryspin-Exner, I., Gur, R. C., Moser, E., & Habel, U. (2009). Amygdala activity to fear and anger in healthy young males is associated with testosterone. *Psychoneuroendocrinology*, *34*(5), 687-693.

Eisenegger, C., von Eckardstein, A., Fehr, E., & von Eckardstein, S. (2013). Pharmacokinetics of testosterone and estradiol gel preparations in healthy young men. *Psychoneuroendocrinology*, *38*(2), 171-178.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203.

Enter, D., Spinhoven, P., & Roelofs, K. (2014). Alleviating social avoidance: Effects of single dose testosterone administration on approach–avoidance action. *Horm. Behav.*, *65*(4), 351-354.

Enter, D., Terburg, D., Harrewijn, A., Spinhoven, P., & Roelofs, K. (2016). Single dose testosterone administration alleviates gaze avoidance in women with Social Anxiety Disorder. *Psychoneuroendocrinology*, *63*, 26-33.

Gainotti, G. (2012). Unconscious processing of emotions and the right hemisphere. *Neuropsychologia*, *50*(2), 205-218.

Geniole, S., Bird, B., McVittie, J., Purcell, R., Archer, J., & Carré, J. (2020). Is testosterone linked to human aggression? A meta-analytic examination of the relationship between baseline, dynamic, and manipulated testosterone on human aggression. *Horm. Behav.*, *123*, 104644.

Geniole, S. N., MacDonell, E. T., & McCormick, C. M. (2017). The threat premium in economic bargaining. *Evolution and Human Behavior*, *38*(5), 572-582.

Geniole, S. N., Proietti, V., Bird, B. M., Ortiz, T. L., Bonin, P. L., Goldfarb, B., Watson, N. V., & Carré, J. M. (2019). Testosterone reduces the threat premium in competitive resource division. *Proceedings of the Royal Society B*, *286*(1903), 20190720.

Gillespie, S. M., Rotshtein, P., Satherley, R.-M., Beech, A. R., & Mitchell, I. J. (2015). Emotional expression

recognition and attribution bias among sexual and violent offenders: a signal detection analysis. *Front. Psychol.*, *6*, 595.

Giltay, E. J., Enter, D., Zitman, F. G., Penninx, B. W., van Pelt, J., Spinhoven, P., & Roelofs, K. (2012). Salivary testosterone: associations with depression, anxiety disorders, and antidepressant use in a large cohort study. *Journal of psychosomatic research*, *72*(3), 205-213.

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

Graham, R., Devinsky, O., & LaBar, K. S. (2007). Quantifying deficits in the perception of fear and anger in morphed facial expressions after bilateral amygdala damage. *Neuropsychologia*, *45*(1), 42-54.

Greenberg, N., & Crews, D. (1990). Endocrine and behavioral responses to aggression and social dominance in the green anole lizard, Anolis carolinensis. *General and comparative endocrinology*, *77*(2), 246-255.

Grillon, C., & Charney, D. R. (2011). In the face of fear: anxiety sensitizes defensive responses to fearful faces. *Psychophysiology*, *48*(12), 1745-1752.

Han, C., Watkins, C. D., Nan, Y., Ou, J., Lei, X., Li, X., & Wu, Y. (2021). Exogenous testosterone decreases men's sensitivity to vocal cues of male dominance. *Horm. Behav.*, *127*, 104871.

Hirschenhauser, K., Taborsky, M., Oliveira, T., Canario, A. V., & Oliveira, R. (2004). A test of the 'challenge hypothesis' in cichlid fish: simulated partner and territory intruder experiments. *Anim. Behav.*, *68*(4), 741-750.

Keltner, D. (2003). Expression and the Course of Life: Studies of Emotion, Personality, and Psychopathology from a Social-Functional Perspective. *Ann. N. Y. Acad. Sci.*, *1000*(1), 222-243.

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (2003). Facial expression of emotion.

Konovalov, A., Hu, J., & Ruff, C. C. (2018). Neurocomputational approaches to social behavior. *Current Opinion in Psychology*, *24*, 41-47.

Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological review*, *76*(3), 308.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest'. *R package version*, *2*(0), 734.

Lane, R. D. (2000). Neural correlates of conscious emotional experience.

LoBue, V., & Rakison, D. H. (2013). What we fear most: A developmental advantage for threat-relevant stimuli. *Developmental Review*, *33*(4), 285-303.

McNicol, D. (2005). *A primer of signal detection theory*. Psychology Press.

Moukheiber, A., Rautureau, G., Perez-Diaz, F., Jouvent, R., & Pelissolo, A. (2012). Gaze behaviour in social blushers. *Psychiatry research*, *200*(2-3), 614-619.

Muller, M. N., & Wrangham, R. W. (2004). Dominance, aggression and testosterone in wild chimpanzees: a test of the 'challenge hypothesis'. *Anim. Behav.*, *67*(1), 113-123.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review*, *9*(4), 278-311.

Peters, A. (2002). Testosterone and the trade-off between mating and paternal effort in extrapair-mating superb fairy-wrens. *Anim. Behav.*, *64*(1), 103-112.

Pichon, S., de Gelder, B., & Grèzes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage*, *47*(4), 1873-1883.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873-922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends Cognit. Sci.*, *20*(4), 260-281.

Rukavina, S., Sachsenweger, F., Jerg-Bretzke, L., Daucher, A. E., Traue, H. C., Walter, S., & Hoffmann, H. (2018). Testosterone and its influence on emotion recognition in young, healthy males. *Psychology*, *9*(07), 1814.

Shobe, E. R. (2014). Independent and collaborative contributions of the cerebral hemispheres to emotional processing. *Front. Hum. Neurosci.*, *8*, 230.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, *64*(4), 583-639.

Strauss, M., Makris, N., Aharon, I., Vangel, M. G., Goodman, J., Kennedy, D. N., Gasic, G., & Breiter, H. C. (2005). fMRI of sensitization to angry faces. *Neuroimage*, *26*(2), 389-413.

Terburg, D., Aarts, H., & van Honk, J. (2012). Testosterone affects gaze aversion from angry faces outside of conscious awareness. *Psychological science*, *23*(5), 459-463.

Tipples, J. (2018). Caution follows fear: Evidence from hierarchical drift diffusion modelling. *Emotion*, *18*(2), 237.

Trumble, B. C., Smith, E. A., O'Connor, K. A., Kaplan, H. S., & Gurven, M. D. (2014). Successful hunting increases testosterone and cortisol in a subsistence population. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1776), 20132876.

Tsoi, D. T., Lee, K.-H., Khokhar, W. A., Mir, N. U., Swalli, J. S., Gee, K. A., Pluck, G., & Woodruff, P. W. (2008). Is facial emotion recognition impairment in schizophrenia identical for different emotions? A signal detection analysis. *Schizophrenia research*, *99*(1-3), 263-269.

Van der Meij, L., Almela, M., Buunk, A. P., Fawcett, T. W., & Salvador, A. (2012). Men with elevated testosterone levels show more affiliative behaviours during interactions with women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1726), 202-208.

Van Honk, J., & JLG Schutter, D. (2007). Testosterone reduces conscious detection of signals serving social correction: Implications for antisocial behavior. *Psychological Science*, *18*(8), 663-667.

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*(3), 218-225.

Van Honk, J., & Schutter, D. (2007). Testosterone reduces conscious detection of signals serving social correction: Implications for antisocial behavior. *Psychological Science*, *18*(8), 663-667.

van Honk, J., Tuiten, A., Hermans, E., Putnam, P., Koppeschaar, H., Thijssen, J., Verbaten, R., & van Doornen, L. (2001). A single administration of testosterone induces cardiac accelerative responses to angry faces in healthy young women. *Behavioral neuroscience*, *115*(1), 238.

Vongas, J. G., & Al Hajj, R. (2017). The effects of competition and implicit power motive on men's testosterone, emotion recognition, and aggression. *Hormones and Behavior*, *92*, 57-71.

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

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.

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion*, *1*(1), 70.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in neuroinformatics*, 14.

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

Wu, Y., Clark, L., Zilioli, S., Eisenegger, C., Gillan, C. M., Deng, H., & Li, H. (2018). Single dose testosterone administration modulates emotional reactivity and counterfactual choice in healthy males. *Psychoneuroendocrinology*, *90*, 127-133.

Wudarczyk, O. A., Kohn, N., Bergs, R., Goerlich, K. S., Gur, R. E., Turetsky, B., Schneider, F., & Habel, U. (2016). Chemosensory anxiety cues enhance the perception of fearful faces–An fMRI study. *Neuroimage*, *143*, 214-222.

Declaration of Competing Interest

The authors declare that they had no conflict of interest with respect to their authorship or the publication of the article.

**Highlights**

- Exogenous testosterone attenuates sensitivity to facial threat.
- Testosterone group showed significantly reduced sensitivity to angry facial expressions.
- Testosterone has a causal effect on men’s perception of facial threat.
- The effect of testosterone to threat cues primarily driven by angry facial expressions.