Testosterone responses to competition in men are related to facial masculinity

Nicholas Pound1,* , Ian S. Penton-Voak2 and Alison K. Surridge3,†

1 Department of Psychology, School of Social Sciences, Brunel University, Uxbridge UB8 3PH, UK
2 Department of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK
3 School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Relationships between androgens and the size of sexually dimorphic male traits have been demonstrated in several non-human species. It is often assumed that a similar relationship exists for human male faces, but clear evidence of an association between circulating testosterone levels and the size of masculine facial traits in adulthood is absent. Here we demonstrate that, after experimentally determined success in a competitive task, men with more a masculine facial structure showed higher levels of circulating testosterone than men with less masculine faces. In participants randomly allocated to a ‘winning’ condition, testosterone was elevated relative to pre-task levels at 5 and 20 min post-task. In a control group of participants allocated to a ‘losing’ condition there were no significant differences between pre- and post-task testosterone. An index of facial masculinity based on the measurement of sexually dimorphic facial traits was not associated with pre-task (baseline) testosterone levels, but was associated with testosterone levels 5 and 20 min after success in the competitive task. These findings indicate that a man’s facial structure may afford important information about the functioning of his endocrine system.

Keywords: testosterone; facial masculinity; competition; sexual dimorphism

1. INTRODUCTION

Testosterone-dependent secondary sexual characteristics are demonstrably important in mate choice (Andersson 1994). Relationships between androgens and sexual displays have been demonstrated in several non-human species (Roberts et al. 2004), and it has been proposed that a similar underlying biological process may be useful in explaining women’s preferences for sexually dimorphic facial characteristics in men (see Rhodes 2006, for a review). A crucial underlying assumption of much work investigating women’s preferences for sexually dimorphic facial traits is the existence of a relationship between circulating testosterone levels and facial masculinity. Testosterone administration causes craniofacial growth in adolescents (Verdonck et al. 1999), but direct evidence of an association between circulating testosterone levels and masculine facial trait size in adulthood is absent. Moreover, evidence for an association between circulating testosterone levels and perceived masculinity in adult male faces is equivocal (Neave et al. 2003; Penton-Voak & Chen 2004; Roney et al. 2006).

To date, studies investigating relationships between appearance and testosterone (Neave et al. 2003; Penton-Voak & Chen 2004; Roney et al. 2006) have used a single measure of testosterone from each participant, failing to take into account that testosterone levels in men are dynamic, exhibiting both diurnal variability and fluctuations in response to social and psychological events. For example, circulating testosterone levels in men are affected by watching arousing videos (Hellhammer et al. 1985), brief interactions with women (Roney et al. 2003, 2007) and being insulted (Cohen et al. 1996). Moreover, Neave et al. (2003), Penton-Voak & Chen (2004) and Roney et al. (2006) did not control for salivary blood contamination that can elevate salivary testosterone levels (Kivlighan et al. 2004) and could introduce unsystematic error, or even systematic bias if its causes (e.g. gingivitis and/or micro-injury to the oral mucosa) vary systematically with other participant characteristics.

Given that testosterone levels are dynamic, total tissue exposure to the hormone’s masculinising effects may depend on the magnitude of hormonal responses to events and not just baseline levels, e.g. average plasma testosterone concentrations over time may be more closely associated with post-event testosterone levels than with baseline levels. Analogous evidence exists with respect to the importance of blood glucose dynamics. Although chronic hyperglycaemia, due to elevated baseline (fasting) blood glucose, is a continuous risk factor for cardiovascular disease (Coutinho et al. 1999) and microvascular complications (Stratton et al. 2000), postprandial hyperglycaemia is an independent risk factor (Ceriello 2000). In type 2 diabetes patients, glycosylated haemoglobin levels (HbA1c), which reflect average blood glucose concentrations over time, are more closely associated with postprandial plasma glucose, than with fasting or preprandial levels (Avignon et al. 1997). Moreover, coronary heart disease (Donahue et al. 1987) and mortality (DECODE 1999) are better predicted by postprandial than fasting plasma glucose.

Consistent with the ‘challenge hypothesis’ (Wingfield et al. 1990) men respond to competitive situations with increased testosterone levels (for a review see
Archer 2006). In men, testosterone levels are affected by the outcome of competition, increasing in winners relative to losers in physical (Mazur & Lamb 1990) and non-physical (Gladue et al. 1989; Mazur et al. 1992; McCaul et al. 1992) contests. In certain circumstances, testosterone levels rise in anticipation of a contest (e.g. Booth et al. 1989; Mazur et al. 1992). Moreover, direct participation in competition is not necessary to elicit testosterone responses since vicarious experiences of winning and losing can have similar effects (Bernhardt et al. 1998). There is some evidence for individual differences between men in the magnitude of their testosterone responses to contests (Cohen et al. 1996; Edwards et al. 2005), and the present experiment was conducted to determine whether such individual differences are related to an aspect of anatomical masculinization, i.e. the expression of sexually dimorphic characteristics in the face.

There is evidence that individual differences in behavioural responses to certain stimuli are associated with another anatomical trait, second-to-fourth digit ratio (2D : 4D), which is sexually dimorphic (Manning et al. 2000) and a putative marker for prenatal testosterone (Manning 2002). Millet & Dewitte (2007) found that, after viewing an ‘aggressive’ music video, men with more masculine (lower) 2D : 4D ratios gave more aggressive responses to a series of hypothetical provocation scenarios. Similarly, Van den Bergh & Dewitte (2006) found that exposure to images of attractive women produced more pronounced changes in behaviour in an ultimatum game in men with masculine (lower) 2D : 4D ratios. Neither study, however, measured testosterone levels and meta-analysis has indicated that in the normal adult population 2D : 4D is not associated with circulating testosterone levels (Hönekopp et al. 2007). Moreover, 2D : 4D is unrelated to facialmetric measures of masculinity in men (Burriss et al. 2007). Crucially, then, definitive evidence of an association between circulating testosterone and facial masculinization in adult men has yet to be found, even though a large literature on human facial preferences is based on this assumption (see Rhodes 2006 for a review).

Previous research using a variety of procedures has demonstrated that participation in competition increases testosterone levels, and these increases tend to be greater in winners than in losers (for a review see Archer 2006). Consequently, the a priori aim of this research was to examine the associations between facial structure and the magnitude of the testosterone increases that occur in response to success in competition. Accordingly, we tested whether there was an association between an index of facial masculinity (derived from measurements of sexually dimorphic facial characteristics) and the magnitude of men’s testosterone responses to success in a competitive task. We used a task similar to that used by Gladue et al. (1989), i.e. a non-physical laboratory contest in which participants are experimentally allocated to either a winning or losing condition by adjusting feedback regarding task performance. However, the task used in the present study also involved an element of vicarious participation in a contest, since participants were asked to predict the outcome of wrestling bouts.

2. MATERIAL AND METHODS

(a) Measurement of salivary testosterone and blood contamination

Male participants (n=57; mean age 21.7±2.5 years) each provided a series of saliva specimens, two before and two after participating in a competitive task according to a schedule similar to that used by Mazur et al. (1997). Specimens were collected at 10 (T1) and 5 (T2) min before participating in the task, and 5 (T3) and 20 (T4) min after completing the task. The competitive task took approximately 5 min to complete. To avoid anticipatory effects on testosterone levels (e.g. Booth et al. 1989; Mazur et al. 1992, 1997), the competitive aspect of the task was explained fully to participants after they had provided their second saliva specimen (T2—5 min before the task). All participants were tested in isolation. To minimize variability due to diurnal fluctuations in circulating testosterone (Dabbs 1990), all experimental sessions took place between 12.30 and 16.00, and for the specimens collected for this experiment there were no significant associations between testosterone concentrations and specimen production times. To stimulate saliva production, participants chewed sugar-free gum and then delivered 5 ml saliva into a polypropylene cryotube (Nalge Nunc International, Rochester, NY). Specimens were frozen and stored at −20°C and then assayed in duplicate using a competitive enzyme immunoassay for testosterone (Salimetrics LLC, State College, PA). The average intra-assay coefficient of variation was 8.9 per cent and inter-assay coefficient was 10.3 per cent. To assess blood contamination, a competitive enzyme immunoassay (Salimetrics LLC) was used to determine levels of transferrin in each saliva specimen. For this, the average intra-assay coefficient of variation was 6.5 per cent and inter-assay coefficient was 5.5 per cent.

(b) Competitive task

The competitive task involved predicting the outcome of Sumo wrestling bouts based on fictitious information about the wrestlers. The task involved notional direct competition with other participants (to be most successful in predicting the outcome of bouts) and vicarious experiences of being involved in competition (watching a chosen wrestler compete). Success or failure (i.e. ‘winning’ or ‘losing’) in the task was experimentally determined. Of the 57 participants, 10 were selected randomly and allocated to a losing control group while the remaining 47 were allocated to a winning condition. Allocation to the losing group was limited since the purpose of this control group was simply to validate the experimental procedure (i.e. demonstrate that participation in the task caused changes in the testosterone levels). The main aim of this research was to examine the associations between facial structure and that magnitude of the testosterone increases that occur in response to success in competition, rather than compare associations between testosterone and facial structure in experimentally allocated ‘winners’ and ‘losers’.

Stimuli were presented on a computer monitor, and during the task participants were asked to predict the outcome of a series of six wrestling (Sumo) bouts based on information about the wrestlers (e.g. height, weight, reach) who were identified by letter codes. After making each prediction, participants were shown a short (<30 s) video of a Sumo bout with audio commentary. Prior to the start of each bout, an arrow was used to label a wrestler as the one chosen by the participant. However, a participant’s success or
failure in predicting the outcome was experimentally allocated with either the winning or losing wrestler marked as chosen. Wrestlers were distinguishable by the colour of their mawashi (belt worn by each wrestler). In the winning condition, participants were told, and shown videos indicating, that they had successfully predicted the outcome of five out of six bouts. In the losing condition, the success rate given was just one out of six bouts.

(c) Facial masculinity index
Participants’ faces were photographed in a standing position, with a neutral expression, using a digital camera (Canon EOS 350D) at a resolution of 1629×2304 pixels, with bilateral illumination (Portaflash DL 1000). Where necessary to reveal the hairline, hair was pulled back with a hairband. Five facial dimensions (ratios) previously shown to be sexually dimorphic (Penton-Voak et al. 2001) were measured. These ratios were: (i) eye size (horizontal inter-exocanthial distance/inter-endocanthial distance), (ii) lower face/face height (vertical distance from mean pupil height to gnathion approximation/vertical distance from trichion to gnathion approximation), (iii) cheekbone prominence (horizontal distance between most outward projecting points on the face at or below the eyes/vertical distance between left and right gonion approximations), (iv) face width/lower face height (horizontal distance between most outward projecting points on the face at or below the eyes/vertical distance from mean pupil height to gnathion approximation), and (v) mean eyebrow height (mean vertical distance from pupil to inferior aspect of brow/vertical distance from trichion to gnathion approximation). Measurements were made as previously described (Penton-Voak et al. 2001) using University of Texas Health Science Center, San Antonio (UTHSCSA) ImageTool to record landmark locations.

For each dimension, measures were converted to standardized (z) scores and a composite facial masculinity index was computed as the sum of these z-scores (oriented such that high scores are masculine for each dimension). Measurements were made in duplicate by two independent research assistants blind to the testosterone status of the participants. The duplicate measurements were highly correlated (r=0.98) and the mean of these was used in analyses.

(d) Perceived masculinity and dominance ratings
An independent sample of 72 raters (61 females and 11 males; mean age 19.2 years; s.d. 2.7 years) rated the participants’ faces for perceived masculinity. A further sample of 29 raters (18 females and 11 males; mean age 24.8 years; s.d. 6.9 years) rated the faces for perceived dominance. Digital photographs were cropped to remove hair and clothing and then presented to raters in random order on a computer screen. Raters were asked to rate each face on a 7-point Likert scale from 1 (not at all masculine/dominant) to 7 (very masculine/dominant). Raters exhibited a high degree of consensus for perceived masculinity (Cronbach’s α=0.96) and dominance (α=0.85), so mean ratings for each face were used in analyses.

(e) Statistical analyses
Since testosterone concentrations vary across an order of magnitude (34–247 pg ml⁻¹ in the present study) and are positively skewed (skewness/standard error of skewness greater than 2), statistical tests involving testosterone concentrations have been performed on log-transformed values. All analyses were carried out using SPSS 13.0. To validate the experimental task (i.e. to determine whether being allocated to the winning condition caused an increase in testosterone), mean pre-task (T1 and T2) and post-task (T3 and T4) testosterone levels were calculated for each participant and analysed using a 2 (win/lose)×2 (pre/post) repeated-measures ANOVA. Post hoc comparisons between mean pre-task (T1 and T2) and post-task testosterone levels were conducted using paired sample t-tests. All tests were two tailed with α=0.05. Pearson product-moment correlation coefficients were calculated to assess zero-order relationships between testosterone levels and both the facial masculinity index and masculinity/dominance ratings. In addition, partial correlation coefficients were calculated to examine linear relationships between variables while controlling for the level of blood contamination in the specimens concerned.

3. RESULTS

(a) Facial masculinity index and testosterone
Participation in the competitive task affected testosterone levels (figure 1). Overall, in winners (n=47), post-task testosterone (mean T3 and T4) was on average 14.6 per cent higher than pre-task testosterone (mean T1 and T2), while in the control group of losers (n=10) it was 3.3 per cent lower. This condition (win/lose)×specimen time (pre/post) interaction was significant (repeated-measures ANOVA, d.f.=1.55; F=4.59; p=0.037). Paired sample t-tests revealed that post-task testosterone (mean T3 and T4) was significantly higher than pre-task testosterone (mean T1 and T2) in winners (t=3.45, d.f.=46, p=0.0012). Significant elevation relative to pre-task (mean T1 and T2) was seen at both T3 (t=3.59, d.f.=46, p=0.00081) and T4 (t=2.79, d.f.=46, p=0.0077). In the control group of losers there were no significant differences between pre- and post-task testosterone.

As shown in table 1, in winners (n=47) the facial masculinity index was not associated with pre-task testosterone levels (mean T1 and T2) but as shown in figure 2, it was positively associated (r=0.36, n=47, p=0.013) with post-task testosterone (mean T3 and T4). The association was present for specimens collected at T3—5 min (r=0.39, n=47, p=0.006) and T4—20 min
Table 1. Associations between facial masculinity index and measures of pre- and post-task testosterone (log transformed) for men allocated to the ‘winning’ condition.

<table>
<thead>
<tr>
<th></th>
<th>correlation with facial masculinity index (n=47)</th>
<th>partial correlation with facial masculinity index controlling for blood contamination (n=47)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean pre-task testosterone</td>
<td>$r=0.19$; n.s.</td>
<td>$r=0.20$; n.s.</td>
</tr>
<tr>
<td>T1</td>
<td>$r=0.23$; n.s.</td>
<td>$r=0.24$; n.s.</td>
</tr>
<tr>
<td>T2</td>
<td>$r=0.14$; n.s.</td>
<td>$r=0.14$; n.s.</td>
</tr>
<tr>
<td>mean post-task testosterone</td>
<td>$r=0.36$; p=0.013</td>
<td>$r=0.37$; p=0.011</td>
</tr>
<tr>
<td>T3</td>
<td>$r=0.39$; p=0.006</td>
<td>$r=0.40$; p=0.006</td>
</tr>
<tr>
<td>T4</td>
<td>$r=0.29$; p=0.048</td>
<td>$r=0.31$; p=0.029</td>
</tr>
</tbody>
</table>

Figure 2. Association between post-task testosterone (mean T1 and T4) and facial masculinity index for men allocated to the ‘winning’ condition; $r=0.36$; $p=0.013$.

$(r=0.29$, $n=47$, $p=0.048)$ after the competitive task. Additionally, controlling for blood contamination, partial correlations between the facial masculinity index and testosterone remained significant at T3—5 min (partial correlation $r=0.37$, d.f. = 44, $p=0.011$) and T4—20 min (partial correlation $r=0.31$, d.f. = 44, $p=0.029$).

The purpose of allocating 10 participants to a losing condition was simply to create a control group to validate the experimental procedure (i.e. demonstrate that participation caused changes in testosterone levels). Consequently, there were not sufficient men in this group ($n=10$) to effectively examine associations between facial structure and testosterone levels after losing. However, it is worth noting that there were no significant associations between the facial masculinity and either pre- or post-task testosterone in losers (all $p>0.2$). Moreover, while there was a significant zero-order correlation between mean pre-task testosterone and the facial masculinity index for the sample as a whole ($r=0.27$, $n=57$, $p<0.05$), once blood contamination was controlled for the partial correlation was non-significant ($p>0.05$).

**4. DISCUSSION**

The present study provides clear evidence of an association between circulating levels of testosterone and facial structure in young adulthood. Since testosterone causes craniofacial growth (Verdonck et al. 1999), exposure over time to large and frequent testosterone responses to competitive situations could produce greater cumulative exposure in facial tissues to the hormone’s masculinising effects leading to an association analogous to that which exists between postprandial blood glucose spikes and tissue damage. Intermittent spikes in hormone levels may have more pronounced effects than chronic elevation of baseline levels, to which rapid tissue habituation may occur. The fact that the observed association is only present for testosterone levels measured after success in a competitive task demonstrates the importance of studying testosterone dynamics and not just baseline levels.

The lack of significant associations between testosterone measures and either perceived masculinity or dominance is consistent with the findings of Neave et al. (2003) but not with more recent findings (Penton-Voak & Chen 2004; Roney et al. 2006). Associations in the range 0.2–0.3 were found but they merely approached significance. To place the present findings in context, the three previous studies involved sample sizes of $n=48$, 50 and 38, respectively. The present study, with a sample size of $n=57$, had power of 85.3 per cent to detect a positive association between baseline testosterone and perceived masculinity of the magnitude ($r=0.34$) reported by Roney et al. (2006). Moreover, the present study involved the collection of multiple specimens and statistical control for levels of blood contamination. Given the observed association between perceived dominance and blood contamination, previous reports of associations between perceived masculinity/dominance and baseline testosterone could be blood contamination artefacts, or the...
discrepant findings could be attributable to endocrine differences between study populations.

The issue as to whether there is an association between testosterone and perceived masculinity still remains rather equivocal, potentially raising some questions about the validity of perceived masculinity measures to assess measured, structural facial sexual dimorphism. Although structural and rated masculinity share some characteristics (DeBruine et al. 2006), our data indicate that they are not always interchangeable concepts. Neither perceived masculinity nor dominance was significantly associated with the facial masculinity index. One possibility is that raters may attribute ‘masculine’ ratings to faces they find attractive irrespective of the objective sexual dimorphism, due to stereotypical associations between the term ‘masculinity’ and attractiveness.

These findings confirm an important underlying assumption of current biological theories of male facial attractiveness in humans. Female preferences for masculinity in male faces vary systematically across the menstrual cycle (Penton-Voak et al. 1999; Penton-Voak & Perrett 2000; Johnston et al. 2001; Jones et al. 2005; Welling et al. 2007; Roney & Simmons 2008), according to self-perceived (Little et al. 2001) and other rated (Penton-Voak et al. 2003) attractiveness, female partnership status (Little et al. 2002), female sex drive (Welling et al. 2008), hypothetical environmental ‘harshness’ (Little et al. 2007) and female waist-to-hip ratio (Penton-Voak et al. 2003). Since testosterone can have deleterious effects on the immune system in humans and other mammals (Angele et al. 2000; Messingham et al. 2001), sexually dimorphic traits (such as masculine facial features) may signal immunocompetence and developmental stability to potential mates (Folstad & Karter 1992). In light of such models, systematic variation in female preferences for masculinity in male faces have been argued to represent adaptive strategic pluralism in mate choice in which cues to immunocompetence (i.e. androgen-dependent dimorphic facial traits) are traded off against cues to prosociality that are associated with low testosterone (Gangestad & Simpson 2000; Penton-Voak & Perrett 2001).

Hönekopp et al. (2007) noted that apparent associations between behaviour and an aspect of anatomical masculinity (2D : 4D) could emerge if anatomically more masculine men exhibit greater testosterone responses to an experimental intervention (which could affect subsequent behaviour). Moreover, McGlothlin et al. (2008) have recently shown that, in dark-eyed juncos (Junco hyemalis), testosterone increases in response to a GnRH challenge are positively associated with the size of a secondary sexual characteristic (a plumage ornament—‘tail white’). However, the present study is the first demonstration that an index of anatomical masculinity (in this case, facial structure) is positively associated with directly measured testosterone responses to an experimental intervention in humans. Furthermore, facial structure is known to have a significant influence on mate preferences, whereas, to our knowledge, this has not been demonstrated directly for 2D : 4D. If digit masculinity, such as facial masculinity, is associated with the size of testosterone responses to experimental interventions, then this could explain the findings that men with masculine 2D : 4D exhibit greater responses to arousing stimuli (e.g. Van den Bergh & Dewitte 2006; Mille & Dewitte 2007). Both these previous studies employed stimuli likely to provoke increases in circulating testosterone, but neither measured these responses directly. As noted above, there are some intriguing findings of individual differences between men in the magnitude of their testosterone responses to contests (e.g. Cohen et al. 1996; Edwards et al. 2005) and further research is needed to establish how these might relate to other phenotypic traits (e.g. 2D : 4D). Evidence of whether women exhibit similar responses in anticipation of, and following participation in, competition is inconsistent (e.g. Bateup et al. 2002; Kivlghan et al. 2005). It would, however, be interesting to investigate whether women with more masculine facial features exhibit more male-typical endocrine responses to competitive situations.

The present study replicates previous findings that the outcome of non-physical contests (Mazur et al. 1992; McCaul et al. 1992) and vicarious experiences of winning (Bernhardt et al. 1998) can affect testosterone levels in men. Individual differences in the magnitude of these responses have been observed previously (Cohen et al. 1996; Edwards et al. 2005) and the present findings indicate that these are related to degree of facial anatomical masculinization. Further research is needed to establish whether individual differences in testosterone responsiveness are stable over time, and how men who show larger testosterone increases after success in competition respond to fail. Individual differences in the magnitude of testosterone responses to competitive stimuli could reflect either genetic differences or facultative adjustments in life-history strategy in response to early experiences, or both. Work in non-human species has indicated that testosterone may have a role in mediating life-history trade-offs between reproductive and immune functions (Muehlenbein & Bribiescas 2005). Underlying ability to cope with the costs associated with elevated testosterone (e.g. immunocompetence) may permit the development of a more competitive life-history strategy associated with larger testosterone responses to competitive stimuli and the development of more masculine facial features. It seems unlikely, however, that social events in themselves could drive facial masculinization independent of the characteristics of an individual male. In non-experimental situations, competition outcomes are not randomly allocated. Consequently, the extent to which a particular male is exposed to the testosterone responses associated with success in competitive situations depends on the extent to which he is successful. Therefore, the association we report here supports the possibility that men’s facial structure could signal the adoption of a more competitive life-history strategy and perhaps even predict success in competition.

An alternative interpretation of these findings is that facial masculinity is systematically related to the salience of success in competition for participants. There is evidence that testosterone responses depend on a man experiencing mood elevation as a consequence of winning (McCaul et al. 1992) and degree of mood elevation may depend on the salience of the competitive encounter for the individual, and the extent to which they attribute their success to internal, rather than external causes. For example, Gonzalez-Bono et al. (1999) studied professional basketball players, and found that post-game testosterone levels in winners were negatively associated with the extent
to which individuals attributed victory to external causes (e.g. luck and referee’s decisions). Similarly, Edwards et al. (2005) found that the percentage increase in testosterone seen after participation in a soccer match (and winning) was positively associated with self- and other-rated connectedness with teammates. It is possible that, perhaps as a consequence of experiencing more success in social competition outside the laboratory, men with more masculine facial structure were more likely to attribute their success in the experimental task used here to internal causes (e.g. skill, knowledge, ability) and this proximate psychological mechanism led to greater increases in post-task testosterone.

This study was approved by the Brunel University School of Social Sciences Research Ethics Committee.

The authors would like to thank Amy Burgess and Hayley Boulton for their assistance, and three anonymous referees for their many suggested additions to the manuscript. This research was supported by a Brunel Research Enterprise and Initiative Fund (BRIEF) award to N.P.

REFERENCES


Hönkopp, J., Bartholdt, L., Beierb, 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, 313–321. (doi:10.1016/j.psyneu.2007.01.007)


