Psychoneuroendocrinology 93 (2018) 90–102

Contents lists available at ScienceDirect

Psychoneuroendocrinology

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

The role of mating context and fecundability in women's preferences for men's facial masculinity and beardedness

Barnaby J.W. Dixson[a,*], Khandis R. Blake[b], Thomas F. Denson[c], Amany Gooda-Vossos[b], Siobhan M. O'Dean[c], Danielle Sulikowski[d], Markus J. Rantala[e], Robert C. Brooks[b]

[a] School of Psychology, The University of Queensland, Brisbane, Queensland, Australia
[b] Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, The University of New South Wales, Kensington, Sydney 2052 NSW, Australia
[c] School of Psychology, University of New South Wales, UK
[d] School of Psychology, Charles Sturt University, Bathurst, NSW, Australia
[e] Turku Brain and Mind Center, Section of Department of Biology, FI-20014, University of Turku, Finland

ARTICLE INFO

Keywords:
Facial attractiveness
Menstrual cycle
Facial hair
Facial masculinity

ABSTRACT

The ovulatory shift hypothesis proposes that women's preferences for masculine physical and behavioral traits are greater at the peri-ovulatory period than at other points of the menstrual cycle. However, many previous studies used self-reported menstrual cycle data to estimate fecundability rather than confirming the peri-ovulatory phase hormonally. Here we report two studies and three analyses revisiting the ovulatory shift hypothesis with respect to both facial masculinity and beardedness. In Study 1, a large sample of female participants ($N = 2{,}161$) self-reported their cycle phase and provided ratings for faces varying in beardedness (clean-shaven, light stubble, heavy stubble, full beards) and masculinity (−50%, −25%, natural, +25% and +50%) in a between-subjects design. In Study 2, 68 women provided the same ratings data, in a within-subjects design in which fertility was confirmed via luteinising hormone (LH) tests and analysed categorically. In Study 2, we also measured salivary estradiol (E) and progesterone (P) at the low and high fertility phases of the menstrual cycle among 36 of these women and tested whether shifts in E, P or E:P ratios predicted face preferences. Preferences for facial masculinity and beardedness did not vary as predicted with fecundability in Study 1, or with respect to fertility as confirmed via LH in Study 2. However, consistent with the ovulatory shift hypothesis, increasing E (associated with cyclical increases in fecundability) predicted increases in preferences for relatively more masculine faces; while high P (associated with cyclical decreases in fecundability) predicted increases in preferences for relatively more feminine faces. We also found an interaction between E and preferences for facial masculinity and beardedness, such that stubble was more attractive on un-manipulated than more masculine faces among women with high E. We consider discrepancies between our findings and those of other recent studies and suggest that closer scrutiny of the stimuli used to measure masculinity preferences across studies may help account for the many conflicting findings that have recently appeared regarding cycle phase preference shifts for facial masculinity.

1. Introduction

The ovulatory shift hypothesis proposes that peri-ovulatory increases in women's sexual desire occur in response to male phenotypic and behavioral traits (Gangestad and Haselton, 2015). For example, at the peri-ovulatory phase women prefer men with more masculine facial features, including defined brows, deeply set and narrow eyes, thin lips, robust midface, and a square jaw (Penton-Voak et al., 1999; Penton-Voak and Perrett, 2000; Little and Jones, 2012, Little et al., 2008). Facial masculinity is androgen dependent (Whitehouse et al., 2015), and is positively associated with men's current health (Rhodes et al., 2003), past disease resistance (Thornhill and Gangestad, 2006), immune response (Rantala et al., 2012), physical strength (Windhager et al., 2011), social rank (Geniole et al., 2015), and mating success (Hill et al., 2013). However, investment in androgen dependent traits that are associated with mating effort may compromise paternal investment (Muller, 2017), so that masculine men may be costly as long-term partners. Facially masculine men report having more short-term than

---

[*] Corresponding author.
E-mail address: b.dixson@uq.edu.au (B.J.W. Dixson).

https://doi.org/10.1016/j.psyneuen.2018.04.007
Received 9 August 2017; Received in revised form 6 March 2018; Accepted 9 April 2018

long-term sexual partners (Rhodes et al., 2005) and women accurately assigned higher sexual infidelity to facially masculine men (Rhodes et al., 2013), which may explain some of the variation in women's facial masculinity preferences (Kruger, 2006; Perrett et al., 1998). However, women's preferences for facial masculinity were highest at the peri-ovulatory phase of the menstrual cycle (Penton-Voak et al., 1999; Penton-Voak and Perrett, 2000; Little and Jones, 2012, Little et al., 2008), suggesting that the costs of masculinity are sometimes bypassed when heritable benefits to offspring may be gained.

Like facial masculinity, beardedness is sexually dimorphic (Trotter, 1922), androgen dependent (Randall, 2008) and enhances ratings of men's masculinity, age, social dominance, and aggressiveness (Dixson and Vasey, 2012; Dixson and Brooks, 2013; Geniole and McCormick, 2015; Muscarella and Cunningham, 1996; Neave and Shields, 2008). Bearded men also report feeling more masculine (Wood, 1986), endorse masculine gender roles (Oldmeadow and Dixson, 2016), and have higher serum testosterone (Knussman and Christiansen, 1988). Although craniofacial masculinity and beardedness are both androgen dependent, they develop under different androgenic processes. Facial masculinity emerges as testosterone binds to androgen receptors that promote skeletal growth, beginning during fetal development (Whitehouse et al., 2015), becoming elaborated upon under the actions of testosterone during adolescence (Marečková et al., 2011), and is fully developed at adulthood (Penton-Voak and Chen, 2004). Beardedness requires the conversion of testosterone to dihydrotestosterone via 5 alpha reductase activity within hair follicles to stimulate the growth of facial hair (Farthing et al., 1982; Randall, 2008), which suggest facial masculinity varies, to some extent, independently of the capacity to grow a full beard and could signal different or convergent components of quality (Dixson et al., 2016).

Facial hair enhances the appearance of testosterone dependent facial traits, such as overall facial length and jaw size which, in turn, augments judgments of masculinity and dominance (Dixson et al., 2017a; Sherlock et al., 2017). Facial masculinity and beardedness also interact to determine women's attractiveness judgments of men's faces, so that slightly less masculine faces are judged as more attractive when bearded than highly masculine faces, possibly because beards mask the less masculine facial cues that may not enhance male facial attractiveness (Dixson et al., 2016). While highly masculine faces and full beards in combination may not enhance attractiveness due to appearing overly masculine, dominant and aggressive, they may be more attractive when considering short-term rather relationships and when fertility is highest. However, whether this interaction between facial masculinity and beardedness on attractiveness judgments varies over the menstrual cycle is unknown.

Although initial research provided compelling evidence for ovulatory shifts in women's mate preferences (Gangestad and Thornhill, 2008), recent studies did not find ovulatory shifts in preferences for facial masculinity (Harris, 2011, 2013; Zietsch et al., 2015) or beardedness (Dixson and Brooks, 2013; Dixson et al., 2013; Dixson and Rantala, 2016, 2017). Evidence from two meta-analyses were also mixed. Wood et al. (2014) concluded that there were no ovulatory shifts in women's mate preferences for masculinity. However, their meta-analyses estimated that the effect size (g) for masculinity preference shifts was 0.08, with a 95% CI spanning −0.01–0.16, which only just includes 0 and does not constitute strong evidence in favor of the null hypothesis. Additionally, the studies included in the estimate of cycle shift effects on masculinity preferences (k = 38) combined attractiveness judgements for faces, bodies, trait descriptions, and voices. If only studies assessing preference shifts for masculine facial shape are considered (k = 28), the estimated mean effect size more than doubles to 0.19. Gildersleeve et al. (2014) did assess cyclical preference shifts for facial masculinity, specifically, and found the significant predicted shift with an estimated effect size (g) of 0.13 overall, increasing to 0.19 for short-term contexts. Both meta-analyses reported significant cycle phase shifts across other traits relevant to the ovulatory shift hypothesis (including facial symmetry), which are not examined in the current study.

However, many of these studies were criticized for employing self-reported menstrual cycle data and variable computations of the peri-ovulatory phase in their analyses (Harris et al., 2014; Wood and Carden, 2014). Using self-reported recollected dates of menstrual bleeding may not generate accurate estimations of current fecundability (Small et al., 2007), owing to natural variation within healthy and regularly cycling women in menstrual cycle lengths (Jukic et al., 2008) and hormone levels (Jasienska and Jasienski, 2008). These natural differences reflect development in utero (Jasienska et al., 2006b), genetic differences (Jasienska et al., 2006a), body fat distribution (Ziomkiewicz et al., 2008), lifestyle factors (Jasienska, 2003) and age related changes in hormones (Lipson and Ellison, 1992). Statistical simulations suggest that between-subject designs, indirect counting methods, and low statistical power have contributed to mixed findings in past ovulatory shift research (Gangestad et al., 2016). Indirect counting methods do not predict hormonally verified peri-ovulatory periods with greater than 60% accuracy (i.e. these methods typically result in fertile window estimates where no more than 60% of the days are actually in the period of increased fecundability, Blake et al., 2016).

Peri-ovulatory increases in women's sexual desire coincide with rises in estradiol (E) and lower progesterone (P) levels (Roney and Simmons, 2013; Jones et al., 2018a). These hormonal changes may also underpin aspects of women's physical attractiveness (Puts et al., 2013), assertiveness (Blake et al., 2017a,b) and mate preferences (Gangestad and Haselton, 2015). Mid-cycle levels of E were positively associated with between-subject and within-subject preferences for facial masculinity (Roney and Simmons, 2008; Roney et al., 2011; Ditzen et al., 2017). However, two studies employing within-subject designs did not report effects of E or P and instead found that testosterone levels were associated with preferences for facial masculinity (Bobst et al., 2014; Welling et al., 2007). Three recent studies employing between-subject designs also found no association between E and preferences for facial masculinity for either short-term or long-term relationships (Marcinkowska et al., 2016; Escasa-Dorne et al., 2017; Jones et al., 2018b). Additional studies testing associations between women's hormone levels over the menstrual cycle and their mate preferences would therefore be valuable.

Here, following recent methodological recommendations targeted at reducing inconsistent findings across tests of the ovulatory-shift hypothesis (Gangestad et al., 2016), we tested whether women's preferences for facial masculinity and beardedness vary with fecundability using three methods varying in expected reliability: with the fertile window estimated via self-report of recent menstruation commencement dates; via detection of luteinising hormone peaks to define the fertile window; and via measures of salivary estradiol and progesterone representing continuous variations in fecundability across the cycle. In Study 1, we used a large between-subjects design among 2161 women who provided sexual attractiveness ratings when considering a short-term relationship for stimuli varying in facial hair (clean-shaven, light stubble, heavy stubble, full beards) and facial masculinity (−50%, −25%, natural, +25% and +50%). This sample size is almost twice the 1213 participants recommended by Gangestad and colleagues (2016, for 80% power to detect a medium effect size of $d = 0.5$). In Study 2a, we used a within-subject design in which the peri-ovulatory period was confirmed via LH tests among 68 women. This sample size is again larger than the 48 participants recommended for 80% power to detect a medium effect size of $d = 0.5$ (Gangestad et al., 2016). We also collected salivary E and P at the low and high fertility phases of the menstrual cycle among 36 of these women to test whether shifts in E, P or the E:P ratio predicts preferences (Study 2b). This sample size is larger than some past within-subject studies reporting significant associations between estradiol and women's facial masculinity preferences (e.g. Roney et al., 2011).

2. Materials and procedure

2.1. Study 1: large between-subjects study

2.1.1. Photographic stimuli

Thirty-six men (mean age ± SD = 27.08 ± 5.61 years) of European descent were photographed when clean-shaven, with five days of regrowth (light stubble), 10 days of regrowth (heavy stubble) and at least four weeks of untrimmed growth (full beard) posing neutral facial expressions in front and profile view using a Canon digital camera (8.0 megapixels resolution), 150 cm from the participant under controlled lighting. The clean-shaven versions of these faces had been measured for craniofacial masculinity as part of another study using geometric morphometrics, which used 164 facial landmarks to create an objective measure masculinity that was validated with perceptual ratings (Dixson et al., 2017a). For the present study, we selected 16 men (mean age ± SD = 23.95 ± 3.43 years, range 20–31) that were of intermediate levels of masculinity from the total set of 36 faces, to which we applied manipulations of facial masculinity.

2.1.2. Masculinity manipulation

Facial masculinity was manipulated via JPsychomorph software (Tiddeman et al., 2001). A sexual dimorphism continuum was defined as the vector difference between an average male and an average female face, created by averaging 50 Caucasian male and 50 female face images, respectively, not including the stimulus identities of the current study. The average male and female faces were matched for overall color content using the Match Color tool in Photoshop (vCS5.1). This ensured that morphs created using this continuum would not differ in overall hue from their original image, but permitted variation of local color cues that likely contribute to perceived facial structure. By local color cues, we refer to local relative differences in color between male and female faces. For example, the level of darkness/brightness between the cheeks and the rest of the face, likely differs between the prototype male and female faces. These differences in relative patterns are preserved, even though the overall hues are matched. These local differences contribute to the apparent 3D shape differences between the faces. It is these local color cues, that are then permitted to vary in manipulated stimuli, capturing as much relevant variation in apparent 3D shape of the stimuli as possible (Lee and Perrett, 2000).

For each stimulus identity, the four variants (clean-shaven, light stubble, heavy stubble and full beard) were each then morphed in JPsychomorph to create two images in which masculinity was increased by 25% and 50% (by morphing parallel to the male-female vector, in the direction of the average male face) and decreased by 25% and 50% (by morphing parallel to the male-female vector, in the direction of the average female face), respectively. The resultant morphs were then refined in Photoshop to ensure each had sharp edges at the sides of the neck, smooth pupils (by replacing irises in the morphs with irises from the original image) and were presented on a consistent background color. Removal of artifacts around the neck and eyes ensured the morphs looked as much like un-manipulated photographs as the original images. Each image measured 1458 × 2292 pixels (Fig. 1).

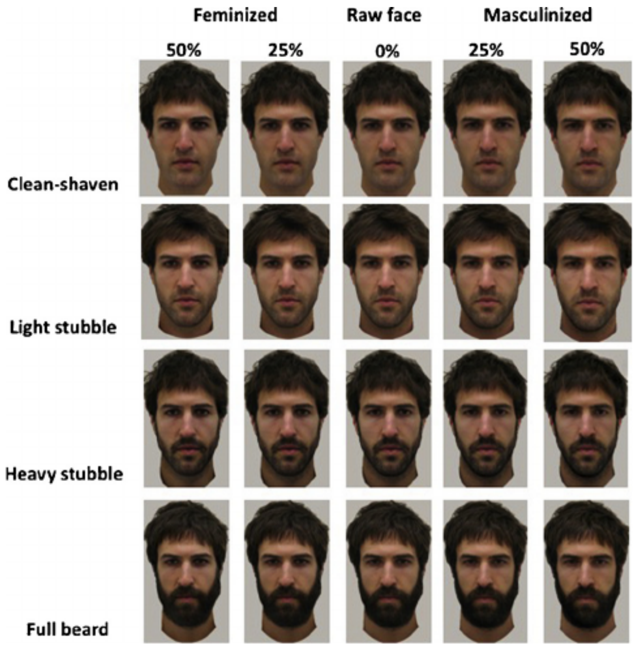

Fig. 1. One of the sixteen individuals (i.e. models) used as stimuli in study 1. The same individual is depicted in four levels of facial hair: clean-shaven, five days of natural growth (light stubble), ten days of natural growth (heavy stubble) and fully bearded. Each level of facial hair was manipulated to appear less (−50% and −25%) masculine, more (25% and 50%) masculine and raw or un-manipulated (0%).

2.1.3. Experimental procedure

Studies were completed online (www.socsci.com). On entry to the website, participants first provided consent and were then assigned to rate faces for sexual attractiveness when considering a short-term relationship using a Likert scale (0 = very unattractive to 10 = very attractive). Participants were given a written instruction taken from Little and Jones (2012), as follows: "Imagine you are looking for the type of person who would be attractive in a short-term relationship. This implies that the relationship may not last a long time. Examples of this type of relationship would include a single date accepted on the spur of the moment and the possibility of a one-night stand." After reading the description, participants rated a total of 16 faces that were randomly selected from the full stimulus set. Thus, one face was drawn at random, without replacement, from each of the 16 male models (i.e. the man whose face was in the photograph), so that the amount of facial hair (clean-shaven, light stubble, heavy stubble or full beard) and the degree of masculinity (−50%, −25%, neutral (raw face), +25%, +50%) was fully randomized in the presentation for each participant.

2.1.4. Participants

Participants were a sub-sample of data from a larger study (Dixson et al., 2016) in which a total of 8698 heterosexual women, the majority of whom were of European descent, completed ratings of the facial stimuli. In the current analyses, we were interested in testing whether fertility is positively associated with preferences for facial masculinity and beardedness when judging for short-term sexual attractiveness. We first removed 179 non-heterosexual participants, leaving 8519 participants who were predominantly or exclusively heterosexual. Of these participants, 194 were eliminated because they were currently pregnant and 2650 because they were on hormonal contraceptives. From the remaining 5675 participants, we eliminated 123 because they were younger than 18 years old, and 3391 because they were over 30 years old to lessen the chance of participants having irregular ovulatory cycles (DeBruine et al., 2010a). This left a sample of 2161 (Mean age = 24.91 years, SD = 3.47) for analyses. However, we repeated the analyses without restricting for age among premenopausal, non-pregnant participants who were not using hormonal contraceptives (n = 5675) and these results are reported in the electronic supplementary materials (Table S1).

2.1.5. Characterising fertility

Participant's fertility was assigned using self-reported information on menstrual cycles wherein they stated when the onset of their last menstrual bleeding occurred (in days). Fertility was assigned using a continuous measure of the likelihood of conception on a given day of the menstrual cycle (Table 1; Wilcox et al., 2001). While this forward counting approach is problematic for assessing actual fertility (Blake

Table 1
GLMM including facial hair, facial masculinity, and likelihood of conception on women's short-term attractiveness ratings.

	F	df	P
Intercept	181.557	1,15.549	< 0.001
Facial hair	35.073	3, 25037.821	< 0.001
Facial masculinity	22.751	4, 25033.561	< 0.001
Likelihood of conception	0.713	1, 1647.091	0.398
Facial hair x facial masculinity	1.966	12, 25041.370	0.023
Facial hair x likelihood of conception	0.571	3, 25026.732	0.634
Facial masculinity x likelihood of conception	0.489	4, 25034.994	0.744
Facial hair x facial masculinity x likelihood of conception	0.915	12, 25036.142	0.531

et al., 2016), the method has an approximate validity of 0.43, (where validity is the correlation co-efficient between actual fertility and assigned fertility; Gangestad et al., 2016).

2.1.6. Statistical analyses

Ratings of short-term sexual attractiveness was the dependent variable in a General Linear Mixed Model where facial hair (clean-shaven, light stubble, heavy stubble, full beard) and masculinity (+50%, +25%, un-manipulated, −25%, −50%) were fixed effects, fertility was a covariate, and participant and model (the man whose face was in the photograph) were fitted as random intercepts.

3. Results

There was a main effect of facial hair (Table 1). Heavy stubble was most attractive followed by light stubble, then full beards and clean-shaven faces were least attractive. There was also a significant main effect of facial masculinity, which reflects the extremes of high (+50%) and low (−50%) were judged as least attractive (Fig. 2). There was a significant facial hair × facial masculinity interaction (Table 1), so that preferences were higher for faces with light stubble that were masculinised over feminized versions of the faces. Heavy stubble followed a similar pattern, while preferences for full beards were stronger when rating slightly less masculine and un-manipulated than when masculinised. Clean-shaven faces were rated least attractive when highly feminized and most attractive in the un-manipulated conditions (Fig. 2). There was no main effect or any interactions involving likelihood of conception on women's preferences (Table 1). These results were also found in our analyses including a larger sample size in including participants of all ages (See ESM Table S1).

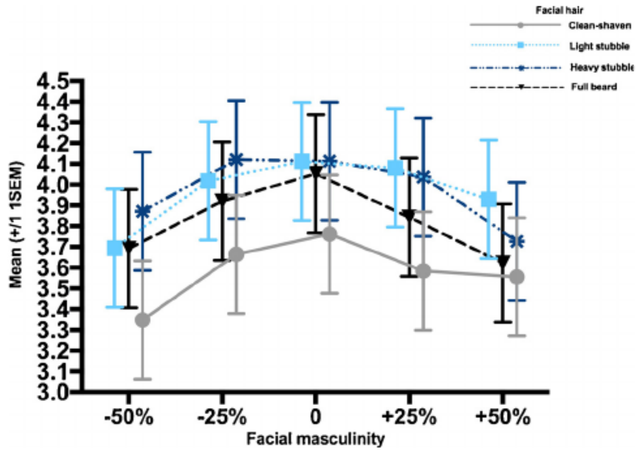

Fig. 2. Mean (± 1 SEM) ratings of sexual attractiveness ratings when considering a short-term relationship for faces varying in beardedness and facial masculinity. The separate lines depict the four levels of facial hair as they were rated across each level of facial masculinity (+/− 50% and 25%, with '0' representing the 'raw' or un-manipulated faces). Ratings for clean-shaven faces are shown on a grey line with a circular symbol, light stubble on a light blue dotted line with a square symbol, heavy stubble on a dark blue dashed and dotted line with an asterisk symbol, and full beardedness on a black line with an inverted triangular symbol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1. Study 2A: a within-subjects study with LH measures

3.1.1. Materials and procedure

3.1.1.1. Participant pre-screening. Exclusion criteria for participating in the study included use of hormonal birth control (current or within the past two months; e.g., birth control pills, Norplant, vaginal ring, birth control patch, Depo-Provera, Mirena IUD); highly irregular menstrual cycles; pregnancy/breastfeeding (current or recent); immune, cardiovascular, metabolic, or kidney disorders; anabolic steroid use; cancer/tumors; recreational drug use within the past 30 days; and smoking or alcohol use within the past 12 h. Participants refrained from consuming caffeine or eating/drinking anything except water within one hour of their session.

3.1.1.2. Menstrual cycle self-report. After pre-screening, participants completed a questionnaire on their average menstrual cycle length, their confidence in that length, the date of their last menses onset, and the date of the predicted onset of their next menses. Women indicated their typical menstrual cycle length on a 13-point scale ('23 days or under', then in one-day intervals to '35 days or over') then reported their confidence in that length on a 9-point scale ('1′ = not at all confident' to '9′ = very confident'). We used this information to determine when each woman should attend her non-fertile session, and when to begin her LH tests for her fertile session. Participants were then randomly allocated to attend their first session when fertile or non-fertile. The mean days on which the non-fertile sessions occurred were as follows: Early follicular non-fertile = 2.50 days, SD = 1.73 and the non-fertile luteal sessions = 25.84 days, SD = 3.49. We aimed to schedule sessions a few days after or a few days prior to menstruation.

3.1.1.3. LH testing procedure. Participants used commercially available urinary LH tests (Blue Cross Bio-Medical Co. LTD, CE/FDA Registered), which detect LH surges with > 99% accuracy at 25 ml U/ml sensitivity (Blake et al., 2016), and tested daily until a positive surge was detected or for a maximum of 10 days. We instructed participants to begin testing 18 days prior to their next predicted cycle onset and to test between 10am to 8pm, reporting the result immediately to the research team via SMS or email. If no result was reported, we followed up with the participant either that day or the next day via SMS or email. We also sent emails prior to the first testing day to remind women to begin testing. Once a positive LH surge was detected, fertile laboratory sessions occurred within 48 h. A review of 38 studies revealed that ovulation occurs on the day of the LH among only a minority women, whereas for the majority ovulate +2 days after a surge (Blake et al., 2016). In the current study, the majority of women were tested 1 day after the LH surge, suggesting they were within the peri-ovulatory period. Thus, 15 (22%) participated on the day of the LH surge, 43 (63%) one day after and 10 (15%) two days after. To control for diurnal progesterone and estradiol variation, laboratory sessions were conducted between 1200 h–1800 h.

3.1.1.4. Laboratory sessions. Participants first rinsed their mouth out with water, then completed the hormone pre-screening questionnaire, and provided a saliva sample via passive drool. The fertile and non-fertile session were identical, with the exception that all fertile sessions included an LH test at the lab to confirm participants were in their fertile phase.

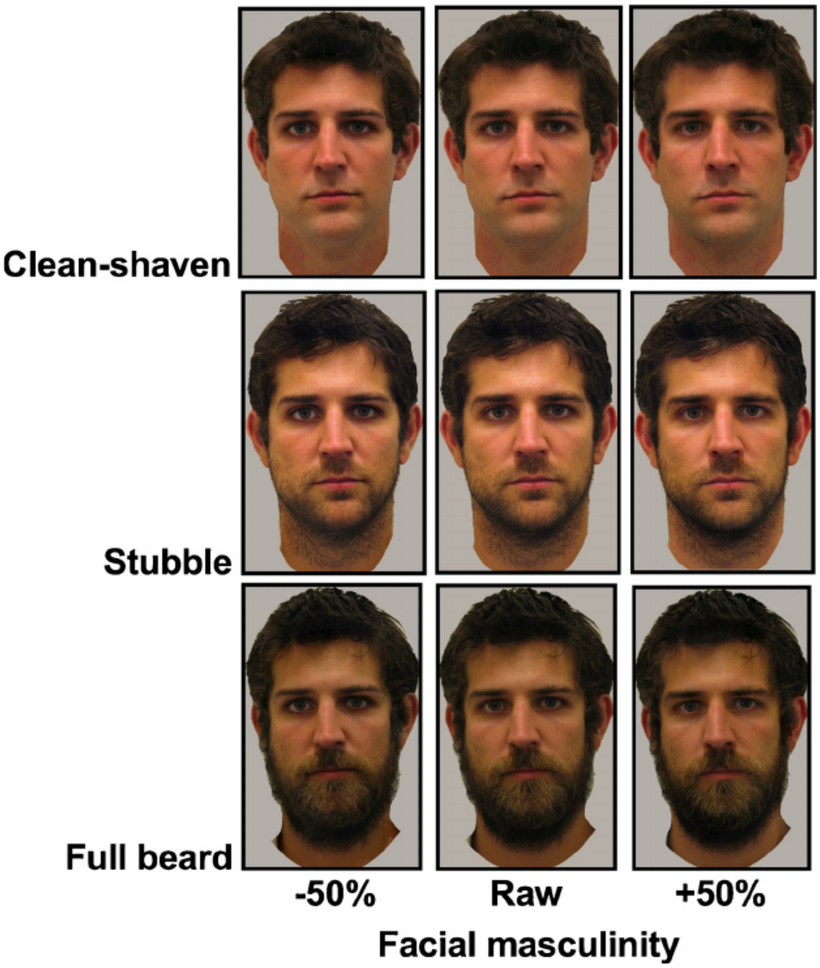

Fig. 3. Examples of the stimuli used in study 2. Images show the same individual in three categories of facial hair (clean-shaven, heavy stubble and fully bearded) manipulated to appear +/−50% masculine and un-manipulated (i.e. raw).

3.1.1.5. Hormone assessment. Saliva samples were stored at −20 °C and analysed by Dr. Clemens Kirschbaum's professional reference laboratory in Dresden, Germany. After thawing, samples were centrifuged at 3000 rpm for five minutes, which resulted in a clear supernatant of low viscosity. Salivary progesterone and estradiol concentrations were measured using commercially available chemiluminescence-immuno-assays with high sensitivity (IBL International, Hamburg, Germany). Mean (+/− 1SD) estradiol at the non-fertile phase was 6.62 (+/− 3.39) and the fertile was 9.00 (+/− 5.98). Mean (+/− 1SD) progesterone at the non-fertile phase was 187.55 (+/− 140.81) and the fertile was 163.76 (+/− 164.66).

3.1.1.6. Photographic stimuli. The same stimuli as in Study 1 were used, except to shorten the experiment 16 male models were presented when clean-shaven, with heavy stubble, and full beards when masculinized (+50%, un-manipulated, and −50%, see Fig. 3). Heavy stubble was chosen rather than light stubble as it has been judged as more sexually attractive and more masculine than light stubble (Dixson and Brooks, 2013; Janif et al., 2014), which studies of the density and distribution of facial hair suggest is due to the uneven density and distribution of light stubble (Dixson and Rantala, 2016). We also elected to use +/−50% masculinity rather than +/−25% as in study 1 in order to keep our study consistent with past laboratory-based experiments testing for ovulatory shifts in women's preferences for facial masculinity (i.e. Little and Jones, 2012).

3.1.1.7. Experimental procedure. Participants first provided consent to complete ratings of men's facial attractiveness. Prior to providing ratings, participants were given the same written instructions as in Study 1 guiding them in how to apply the Likert scales for rating short-term attractiveness, which were taken from Little and Jones (2012). The definition of a long-term relationship was also taken from Little and Jones (2012) and asked participants to imagine they were looking for the type of person who would be attractive in a long-term relationship. Examples of this type of relationship include someone you may want to move in with, settle down and, at some point, wish to marry, or enter into a relationship on similar grounds as marriage (Little and Jones, 2012). Participants were asked to rate faces for sexual attractiveness when considering a short-term relationship and a long-term relationship using a Likert scale (0 = very unattractive to 5 = very attractive) for both dependent variables. Participants made ratings of long-term and short-term attractiveness simultaneously.

After reading these instructions, participants rated all 144 stimulus images, which included the 16 male models at three level of facial hair (clean-shaven, heavy stubble or full beard) and three degrees of facial masculinity (−50%, un-manipulated, +50%). The order in which the faces were presented was fully randomized for each participant.

3.1.1.8. Participants. These data were collected as part of a larger two-wave study on fertility and behavior (Blake et al., 2016, 2017a,b). For current study, 146 women did at least one survey measuring their preferences for the facial stimuli, of whom 21 were removed from further analyses as they either never completed the LH test or their cycles became irregular and we withdrew them before they tested. In total, 125 people completed LH testing of which 87 of them (69.6%) had positive LH tests. Of those 87, 18 did not complete the survey at both fertile and non-fertile time points and were withdrawn from analyses. Finally, one participant's hormone samples were damaged in transit and their data were withdrawn. That left a final sample of 68 ($M_{age} = 22.07$, *SD* = 4.6, range 18–36 years) heterosexual women for analyses at the fertile and non-fertile phases of cycle, 34 of whom also provided saliva samples for hormone analyses that are reported in Study 2 B below.

3.1.1.9. Statistical analyses. Inter-rater reliability for the 16 faces within each category of manipulation (i.e. each level of facial hair and each level of facial masculinity) showed strong reliability (all Cronbach alphas ≥0.872; Table S2). Thus, we averaged ratings across the 16 stimulus images within each category separately for short-term and long-term attractiveness ratings. Ratings of attractiveness when judging a short-term and long-term relationship were dependent variables in a repeated-measures MANOVA where facial hair (clean-shaven, heavy stubble, full beards), masculinity (low masculinity, un-manipulated, high masculinity) and fertility (low, high) were the within-subject factors.

4. Results

The multivariate (combined over long- and short-term relationship contexts) and univariate main effects of facial hair, facial masculinity and fertility were all significant and there were no significant two- or three-way interactions in either the multivariate or univariate analyses (Table 2). The main effects are each considered below (and see Fig. 4).

4.1. Facial hair

Full-factorial paired contrasts revealed that heavy stubble received significantly higher ratings than both clean-shaven and full bearded faces for short-term relationships (all $p < 0.004$, uncorrected alpha), and long-term relationships (all $p \leq 0.031$, uncorrected alpha). Clean-shaven faces were significantly preferred to full bearded faces for short-term relationships (linear contrast, $p = 0.019$), but not for long-term relationships (linear contrast, $p = 0.086$). Overall the univariate simple effect of facial hair was significantly larger for short-term, compared to long-term, relationships (confirmed by entering relationship type as a within-subject factor and examining the relationship type by facial hair interaction, $F_{2,66} = 9.39$, $p < 0.001$, $\eta_p^2 = 0.222$).

4.2. Facial masculinity

Un-manipulated faces were rated as more attractive than both masculinized and feminized faces for short-term relationships (linear contrast, both $p \leq 0.001$), and long-term relationships (linear contrast, both $p < 0.001$), respectively. Feminine faces were significantly preferred to masculine faces for short-term judgements (linear contrast, $p = 0.012$), though not long-term judgements (linear contrast, $p = 0.119$).

4.2.1. Fertility

The main effect of fertility reflects that ratings overall were higher for the high, compared to the low, fertility phase of the menstrual cycle when considering both short-term (univariate $p = 0.005$) and long-term (univariate $p = 0.005$) relationships (Fig. 4).

Table 2
Multivariate repeated-measures ANOVA on the effects of facial hair, facial masculinity and fertility on attractiveness ratings.

	Within-subjects main effects											
	Facial hair[a]				Facial masculinity[a]				Fertility			
Wilks' Lambda	0.752				0.782				0.880			
	F	df	*P*	η_p^2	*F*	df	*P*	η_p^2	*F*	df	*P*	η_p^2
MANOVA	10.18	4, 266	< 0.001	0.133	8.71	4, 266	< 0.001	0.116	4.48	2,66	0.015	0.120
Long-term	12.94	1.5, 98.6	< 0.001	0.162	13.86	1.8, 118.9	< 0.001	0.171	8.33	1,67	0.005	0.111
Short-term	21.10	1.6, 106.9	< 0.001	0.239	16.86	1.8, 119.5	< 0.001	0.201	8.61	1,67	0.005	0.114
					Within-subjects 2-way interactions							
	Facial hair × Facial masculinity				Facial hair × Fertility[a]				Facial masculinity × Fertility			
Wilks' Lambda	0.978				0.977				0.986			
	F	df	*P*	η_p^2	*F*	df	*P*	η_p^2	*F*	df	*P*	η_p^2
MANOVA	0.75	8, 534	0.645	0.011	0.79	4, 266	0.535	0.012	0.49	4, 266	0.746	0.007
Long-term	0.69	4, 268	0.603	0.010	1.04	1.6, 106.6	0.344	0.015	0.39	2, 134	0.679	0.006
Short-term	1.27	4, 268	0.284	0.019	1.56	1.6, 108.3	0.217	0.023	0.47	2, 134	0.624	0.007
	Within-subjects 3-way interaction											
	Facial hair × Facial masculinity × Fertility											
Wilks' Lambda	0.975											
	F	df	*P*	η_p^2								
MANOVA	0.86	8, 534	0.552	0.013								
Long-term[a]	0.45	4, 268	0.774	0.007								
Short-term[a]	1.30	4, 268	0.270	0.019								

[a] Greenhouse-Geisser adjusted df (rounded to 1 decimal place).

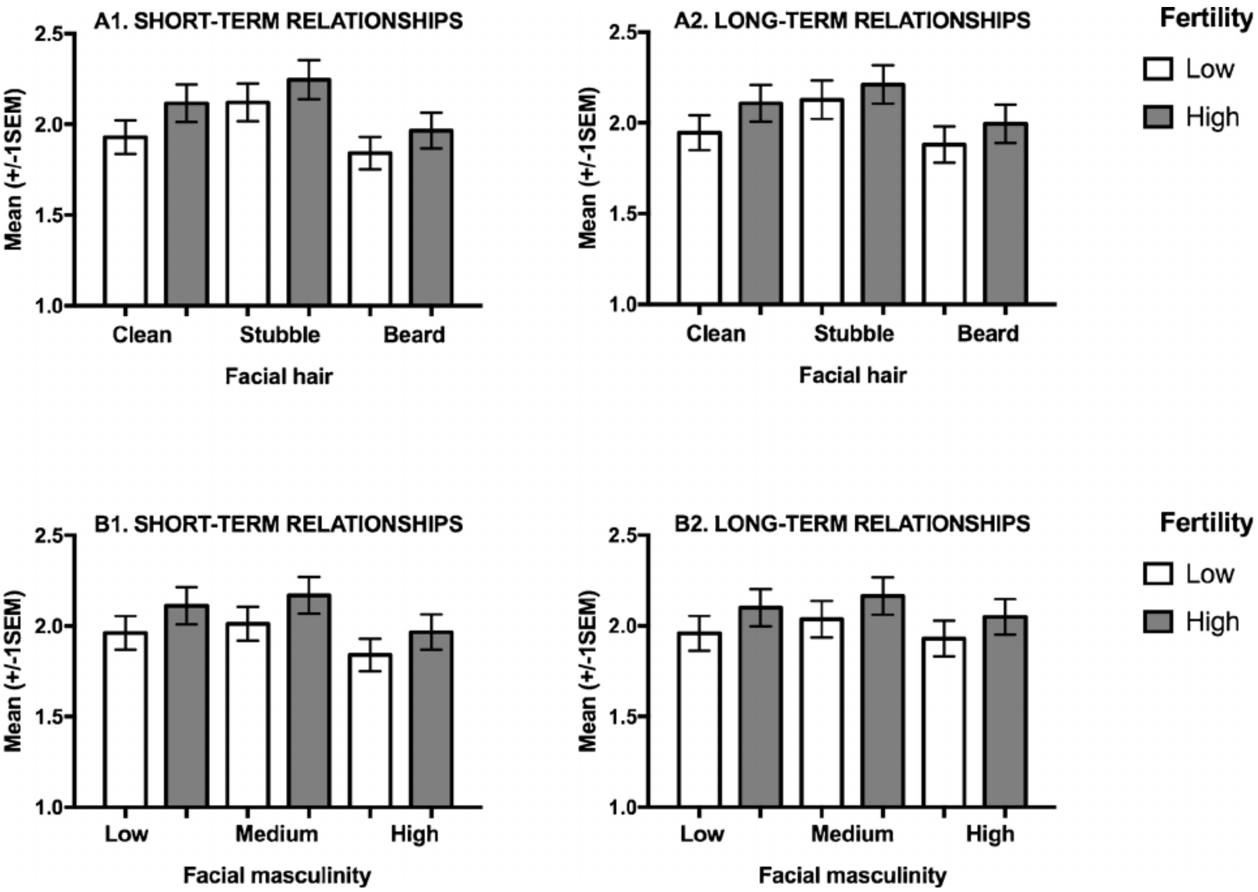

Fig. 4. The upper panels show the mean attractiveness ratings (+/−1 SEM) of faces varying in facial hair (clean-shaven, heavy stubble and fully bearded) for short-term (A1) and long-term relationships respectively (A2). The lower panels show the mean ratings (+/− 1 SEM) of faces manipulated to appear 50% less masculine (low), un-manipulated (medium) and 50% more masculine (high) for short-term (B1) and long-term (B2) relationships respectively. Open bars represent attractiveness ratings made at the low fertility phase and grey bars the fertile phase of the menstrual cycle.

4.2.2. Effects of fertility on relative preferences for facial masculinity and facial hair

We report no significant interaction between fertility and facial masculinity in our omnibus MANOVA. For the purposes of direct comparisons with previous studies reporting on effects of menstrual cycle phase on women's relative preferences for facial masculinity (which have most commonly used clean-shaven faces), we also conducted a MANOVA as described above, but without the facial hair variable, only including responses to clean shaven faces. Linear contrasts (applied separately to long-term and short-term ratings) were used to compare the effect of fertility on the relative preference for un-manipulated faces compared to feminized and masculinized faces, respectively, and for feminized faces compared to masculinized faces, directly. The ovulatory shift hypothesis would predict no significant effect of fertility for long-term relationships, but a significant increase in the fertile phase in relative preference for more masculine faces for short-term relationships. We observe no significant effects of fertility for either long-term or short-term relationships, with the largest effects (as measured by Cohen's d) counter to the predicted direction. For comparison purposes, we conducted parallel analyses examining the effect of fertility on relative preferences for facial hair, considering only the faces not manipulated for masculinity. The associated effect sizes from both sets of analyses are reported in Table 3.

4.3. Study 2B: within-subject variation in hormones and face preferences

4.3.1. Participants

Of the total 68 participants who completed preference tests at both the fertile and non-fertile phase of the menstrual cycle, 34 (M_{age} = 22.68, *SD* = 5.23, range 18–36 years) also provided saliva samples at the fertile and non-fertile sessions.

4.3.2. Statistical analyses

We used linear mixed regression models with maximum likelihood estimation in SPSS version 23 to analyze hormone effects. Linear mixed regression models are appropriate for analyzing nested data with correlated error terms (Twisk, 2006). In our data, there were four repeated effects. Observations (aggregated across stimuli subjects) were recorded for short- and long-term attractiveness judgements (level 1) across three masculinity conditions (feminized, masculinized, un-manipulated; level 2), at three levels of beard growth (bearded, clean-shaven, stubble; level-3), at fertile and non-fertile time points (Level-4), which were nested within women (Level-5). Thus, the four repeated effects were nested and were not treated separately in these analyses and we did not run any between-subject tests as the ovulatory shift hypothesis is best tested with within-subject analyses (Gangestad et al., 2016). However, we accounted for subject variation by including a random intercept in all models and tested for random regression slopes for all hormones, repeated effects, and their interactions and retained random slopes when $p < 0.10$. We used a first-order autoregressive error covariance matrix for the repeated effects and a variance components error covariance matrix for the random effect. We accounted for influential cases by excluding standardized residuals ≥ ± 3-*SD*s, which resulted in a small number of residual outliers being removed for estradiol (18/1224 residual outliers) and progesterone (11/1224 residual outliers). All hormone variables were log-transformed due to significant positive

Table 3
Effect sizes associated with relative preference shifts for facial masculinity across the menstrual cycle.

Clean shaven faces

Comparison	Relationship type	Contrast estimate[a]	*P*	Cohen's *d*[a]
Feminine versus neutral	Long-term	0.003	0.943	< 0.01
Neutral versus masculine	Long-term	−0.030	0.533	−0.155
Feminine versus masculine	Long-term	−0.027	0.516	−0.155
Feminine versus neutral	Short-term	0.001	0.983	< 0.01
Neutral versus masculine	Short-term	−0.045	0.293	−0.255
Feminine versus masculine	Short-term	−0.044	0.291	−0.263

Faces with unmanipulated levels of masculinity

Comparison	Relationship type	Contrast estimate[b]	*P*	Cohen's *d*[b]
Clean shaven versus stubble	Long-term	−0.075	0.093	−0.419
Stubble versus full beard	Long-term	0.021	0.633	0.110
Clean shaven versus full beard	Long-term	−0.054	0.322	−0.247
Clean shaven versus stubble	Short-term	−0.077	0.137	−0.370
Stubble versus full beard	Short-term	0.018	0.720	0.090
Clean shaven versus full beard	Short-term	−0.059	0.313	−0.247

[a] Positive (negative) contrast estimates and effect sizes indicate an increase (decrease) in masculinity preference at the fertile phase of the cycle.

[b] Positive (negative) contrast estimates and effect sizes indicate an increase (decrease) in preference for more facial hair at the fertile phase of the cycle.

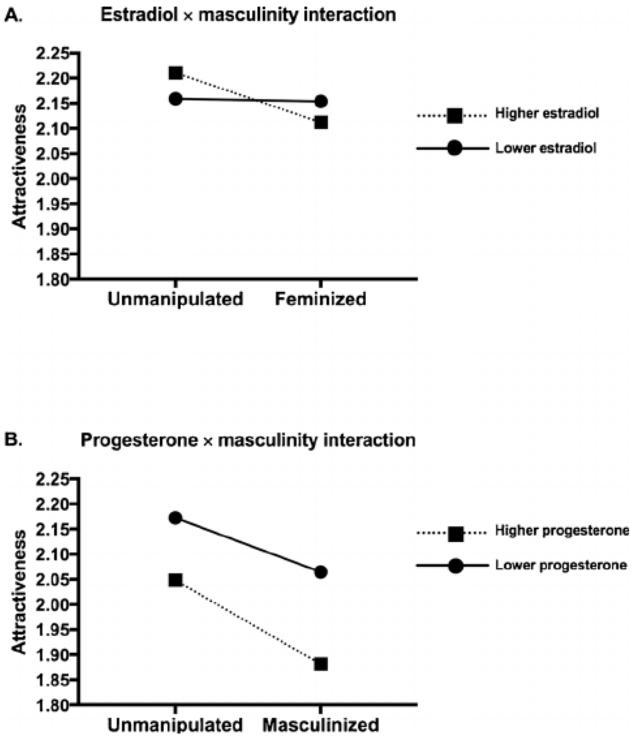

Fig. 5. Panel **A** depicts the interaction between estradiol and feminized facial shape on attractiveness judgements. The figure plots the relationship between estradiol and feminized versus un-manipulated facial shape on attractiveness judgements. Low and high estradiol values are 1-*SD* above and below the mean. Only the high estradiol slope is significant ($b = -0.10$, $t_{1204} = -2.95$, p 0.003). Panel **B** depicts the interaction between progesterone and masculinized facial shape on attractiveness judgements. The figure plots the relationship between progesterone and masculinized versus un-manipulated facial shape on attractiveness judgements. Low and high progesterone values are 1-SD above and below the mean. Both slopes are significantly different from zero, but the effect is greater at high values of progesterone (low progesterone: $b = -0.11$, $t_{1204} = -3.32$, $p = 0.001$; high progesterone: $b = -0.17$, $t_{1204} = -5.15$, $p < 0.001$).

skew and grand mean centered values for mixed model analysis.

We first tested the fixed zero-order effects of estradiol, E:P ratio, and progesterone in separate models, controlling for facial masculinity, beard growth, and relationship context (Step 1). We then entered two-way interactions between the hormone and repeated effects (Step 2). In Step 3, we entered the three-way interactions into the model and in Step 4 we tested the four-way interaction. Where interactions were significant, we examined the effect of the condition at values 1-*SD* below and above the mean for the hormone moderator (Aiken et al., 1991). After examining these effects, we repeated the estradiol models including progesterone as a covariate and vice versa (to test the main and interactive effects of each hormone when the influence of the other hormone was statistically controlled). We did not include the ratio of estradiol to progesterone in these final models due to multicollinearity with progesterone, $r(1222) = -0.78$, $p < 0.001$. No effects reported herein were substantially different when including each hormone as a covariate and details of these analyses are in the supplementary online materials.

These analyses reported within-subject effects. However, it is possible that hormone levels between women are associated with behavioral changes and mate preferences (Havlíček et al., 2015). To test this possibility, we repeated the analyses outlined above replacing estradiol and progesterone at each time-point with these same values averaged across time-points for each subject (i.e., constituting between-subject analyses). All steps and analytical procedures were consistent with those listed above, with one exception. For mean progesterone analyses, we excluded women whose non-fertile session was in the early follicular phase ($n = 4$), as this measurement provides little indication of peak progesterone throughout the menstrual cycle.

5. Results

5.1. Within-subject analyses

The models including estradiol: progesterone (E:P) ratio revealed no main or interactive effects of the E:P ratio on attractiveness (Table S3). The models for estradiol showed no main effect for estradiol, but there was a facial masculinity × estradiol interaction (Table S4). An examination of simple slopes indicated that women with high estradiol preferred the un-manipulated face versus the feminized face ($b = -0.10$, $t_{1204} = -2.95$, $p = 0.003$), whereas women with low estradiol showed no facial preference (p 0.871, see Fig. 5A). The estradiol

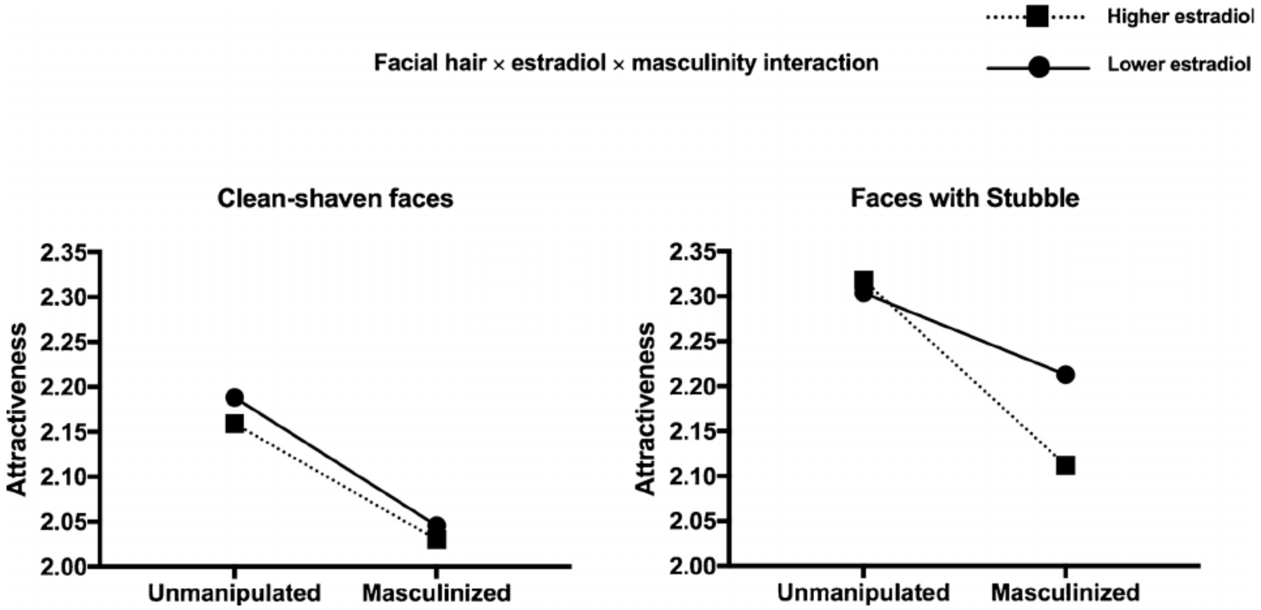

Fig. 6. The interaction between estradiol, masculinized facial shape, and stubble on attractiveness judgements. The figure plots the relationship between estradiol, feminized versus un-manipulated facial shape, and stubble versus clean-shaven faces on attractiveness judgements. Low and high estradiol values are 1-*SD* above and below the mean. Differences between slopes are not significant (*ps* ≥ 0.126).

models further showed a three-way interaction between stubble, masculinity, and estradiol (Table S4). Slope difference tests showed that no pair of slopes significantly differed ($ts_{1204} \leq -1.53$, $ps \geq 0.126$); in all conditions women preferred un-manipulated over masculinized faces (Fig. 6). There was no four-way estradiol interaction (Table S4).

For the progesterone models, there was a main effect of progesterone (Table S5), indicating that high progesterone decreased ratings of attractiveness. There was also a two-way interaction between progesterone and masculinity (Fig. 5B); an analysis of simple slopes showed that women with high and low progesterone both preferred the un-manipulated face over the masculinized face, but that the effect was larger at high values for progesterone (low progesterone: $b = -0.11$, $t_{1204} = -3.32$, $p = 0.001$; high progesterone: $b = -0.17$, $t_{1204} = -5.15$, $p < 0.001$).

5.2. Between-subject analysis

The models including estradiol revealed a significant estradiol x mating context interaction ($b = 0.77$, $SE = 0.29$, $p = 0.011$), which reflects that women with higher levels of estradiol gave higher ratings when judging short-term but not long-term relationships (Table S6). There was also an estradiol x masculinity x mating context interaction that was approaching significance at the 5% level ($b = -0.23$, $SE = 0.14$, $p = 0.091$; Table S6), which reflects that raw and feminized faces were rated as more attractive than masculinized faces for short-term but not long-term relationships (Fig. 7). There were no associations with progesterone (*ps* ≥ 0.131; Table S7).

6. Discussion

With respect to facial masculinity, across two studies we found that women's attractiveness ratings were strongest for the un-manipulated faces, followed by the feminized faces, and then the masculinized faces. This is consistent with the findings of Dixson et al. (2016), a study which used the same stimuli rated by a much larger sample of women. In both studies 1 and 2, where fertility was characterized categorically based on self-reported menstruation dates (between-subjects) and luteinising hormone peaks (within-subjects), respectively, we found no evidence for greater preferences for more masculine faces during the fertile window. Other recent studies have also reported no such increases (Peters et al., 2009; Marcinkowska et al., 2016). However, those studies included sample sizes that are potentially too small to reliably detect cycle shift effects (N = 27 in the within-subjects design of Peters et al., 2009 and N = 115 in the between-subjects design of Marcinkowska et al., 2018; see Gangestad et al., 2016, for detailed simulations and sample-size estimates).

In the present research, we exceeded the respective (between- and within-subject) sample sizes recently recommended by Gangestad et al. (2016) for adequate power to detect cycle phase shift effects. However, these sample size recommendations assumed an effect size of $d = 0.5$ (Gangestad et al., 2016) and two recent meta-analyses have estimated the shifts in preferences for facial masculinity to have an effect size closer to $d = 0.2$ (Gildersleeve et al., 2014; Wood et al., 2014). Even allowing for these meta-analyses to have underestimated the true effect size due to sub-optimal characterization of participants' cycle phase at the time of testing (Blake et al., 2016), an increase from $d = 0.2$ to $d = 0.5$ would mean a substantial underestimate. If we instead presume that the true effect size is closer to $d = 0.4$ (and it could realistically be even lower), the simulations reported by Gangestad et al. (2016) would then have recommended we use a minimum sample size of 1872, or 2398 (if allowing for error in self-report) to achieve 80% power in a between-subjects experiment. Even though our sample size in study 1 (N = 2161) was larger than many previous between-subject studies and we found non-significant associations when using a larger sample without restricting the range of ages (N = 5675; Table S1), we still cannot be certain that inadequate power did not contribute to the null result. This is a sobering observation for those relying on self-report methods to characterize participants' cycle phase.

In Study 2, however, statistical power is less likely to have been a problem. Gangestad et al. (2016) do not offer suggested sample sizes for estimated effect sizes less than $d = 0.5$ for within-subjects design, but with the high correlations of ratings across phases we observed in our data ($r > 0.86$ for both long- and short-term relationship context ratings), the required sample size to achieve adequate power for $d = 0.5$ actually drops to 30, lower than the *a priori* recommendation of Gangestad et al. (2016), which assumed an *r* across phases of only 0.5) and less than half the size of our sample. While our use of LH tests to pinpoint the LH surge that accompanies ovulation is a highly accurate

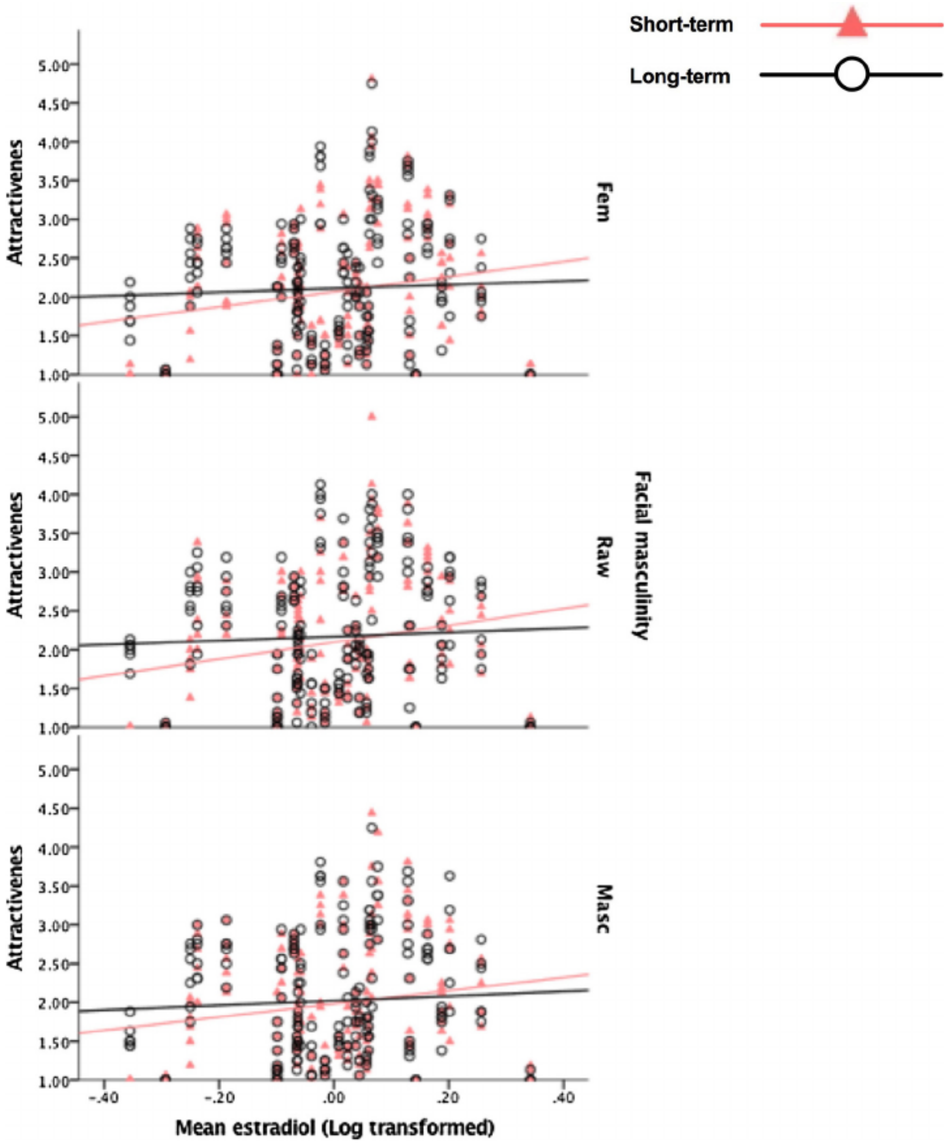

Fig. 7. The results from between-subject analyses. Data show an interaction between estradiol, masculinized facial shape, and mating context on attractiveness judgements. Women's attractiveness judgments for all faces were positively associated with estradiol when judging for short-term but not long-term relationships. This association was most pronounced for un-manipulated (i.e. raw) faces.

method for reliably verifying the date of ovulation (Gangestad et al., 2016), the date ovulation occurs is in the latter half of the fertile window, after fertility has begun to decline from its peak (~24–48 h prior to ovulation). We tested women as soon after the LH surge was detected as was logistically possible and we managed to get participants into the lab within 24 h of detecting an LH surge. Thus, 22% participated on the day of the LH surge, 63% participant one day after and 15% two days after. The logistical difficulties associated with scheduling participant testing sessions on appropriate days are difficult to avoid. However, we cannot ignore the fact that for some women we may have captured the period closer to the end of the fertile window, which may have contributed error variance.

Central neuroendocrine effects of estradiol on female sexuality are well established in nonhuman primates (Dixson, 2012) and may underpin aspects of women's sexual desire, attractiveness and assertiveness (Roney and Simmons, 2013; Puts et al., 2013; Blake et al., 2017a,b; Jones et al., 2018a). Previous studies reported increased preferences for masculinity when estradiol is high (Roney and Simmons, 2008; Roney et al., 2011), and decreased preferences for masculinity when progesterone is high (Jones et al., 2005; Puts, 2006) and our findings are consistent with these reports. It was only when measuring preferences in relation to hormone levels that we found any evidence in favor of the ovulatory shift hypothesis, such that increasing estradiol predicted increases in preferences for relatively more masculine faces in both within and between-subject analyses; while high progesterone predicted higher attractiveness ratings for relatively more feminine faces in our within-subject but not our between subject analyses. Our findings differ from a very recent study (Jones et al., 2018b), using large sample of women (N = 584) and a within-subjects design, which reported no statistically significant associations between salivary hormone levels and women's preferences for facial masculinity.

These discrepancies may reflect differences in stimuli employed between different studies. In Study 2, we presented participants with stimuli varying in three levels of facial masculinity: un-manipulated male faces, those same faces 50% masculinized and 50% feminized. The overall pattern of preferences we observed was strongly quadratic with the un-manipulated faces (on average) much preferred over either the masculinized or feminized versions. Direct comparisons between the

50% masculinized and 50% feminized face ratings and hormone measures did not yield any support for the ovulatory shift hypothesis. Similarly, Jones et al. (2018b) presented participants with male faces that were 50% masculinized and 50% feminized, constructed using the same procedure with which we created our stimuli where for both cases 50% technically refers to a shift in morphology that is equivalent in magnitude to 50% of linear vector difference between an average male and an average female face. Attractiveness shifts for facial masculinity across the cycle (if indeed such shifts occur at all) may be strongest, or even contained entirely within, the relatively narrow range of facial masculinities that are perceived as attractive. In Study 1, we also included 25% masculinized and feminized faces and these less extreme manipulations were occasionally preferred to the un-manipulated faces. This suggests that the range of facial masculinity that females find attractive may be relatively narrow, from −25% to +25% either side of average masculinity − outside this range perceived attractiveness drops substantially and may not capture fertility-related preferences in facial masculinity. This possibility is supported by the fact that the comparisons in our study that provided evidence in favor of the ovulatory shift hypothesis were between the un-manipulated faces and the feminized faces (for estradiol) and between the un-manipulated faces and the masculinized faces (for progesterone). Recently, Holzleitner and Perrett (2017) used faces varying in increments of 50% in facial masculinity from −100% to 200%. Although they did not use 25% increments, they found curvilinear relationships for preferences within the first 50% increment in many cases. Data for our studies 1 and 2 were undertaken at the same time and we elected to follow the protocols of past laboratory-based studies of ovulatory shift effects in facial masculinity by restricting comparisons between +/−50% masculinity manipulations for study 2. We hope that our study draws attention towards how best to operationalize the degree of facial masculinity manipulations in future research seeking to test ovulatory shift effects in attractiveness judgments.

Alternatively, the differences in findings across studies could be due to reduced ecological validity of computer-morphed stimuli compared to natural stimuli (Dixson et al., 2017c), so that ratings of attractiveness are due to by-products of the morphing procedures rather than the actual masculinity manipulations. For example, past research on female physical attractiveness has shown that while artificially manipulating traits like the waist line singularly is effective in altering shape, it confounds inter-correlated measures that relate to size and weight (Brooks et al., 2015; Dixson, 2018). While the same issue of collinearity in facial attributes could confound facial stimuli morphed along dimensions of masculinity using computer graphic programs, there are several reasons to doubt this to be the case. Thus, the morphing techniques use natural variation in androgen dependent processes quantified using multivariate statistics rather than single indices to transform masculinity (DeBruine et al., 2006; Scott and Penton-Voak, 2011). These approaches have included morphing between average composites of male and female faces, morphing on the basis of rated masculinity, and changes in testosterone that occur during male adolescence (DeBruine et al., 2006). Comparisons among these different techniques revealed that all three methods enhanced judgments of masculinity (all $t > 65.4$, all $p < 0.001$), dominance (all $t > 14.6$, all $p < 0.001$) and attractiveness (all $t > 10.8$, all $p < 0.001$; DeBruine et al., 2006). Moreover, there was strong concordance in attractiveness judgments among these methods (all $rs > 0.53$; all $ps < 0.001$), which predicted ideal preferences for facial masculinity and the degree of facial masculinity in women's actual partners (DeBruine et al., 2006). Therefore, the approach we employed in the current study wherein facial masculinity was altered in natural photographs by morphing between an average male and female face is high in ecological validity (DeBruine et al., 2006). Nevertheless, while researchers have taken great care to outline how stimulus preparation can introduce confounds into tests of facial masculinity in men's facial attractiveness (DeBruine et al., 2010b; Scott and Penton-Voak, 2011), we suggest future research on ovulatory shifts in facial masculinity preferences may benefit from using smaller degrees of manipulation in facial masculinity.

In concordance with past studies (Janif et al., 2014; Neave and Shields, 2008; Dixson et al., 2016), we found that women's preferences were strongest for heavy stubble, with mixed preferences for clean-shaven faces and full beards. However, women's preferences for facial hair did not vary between the fertile and non-fertile phases in either Study 1 or Study 2, supporting previous null results using indirect measures to characterize fecundability (Dixson and Brooks, 2013; Dixson et al., 2013; Dixson and Rantala, 2016, 2017). With respect to hormonal influences on women's attractiveness judgments of facial hair, within-subject preferences for un-manipulated over masculinised faces were higher among women with higher estradiol for faces with stubble, but this same effect was reduced for clean-shaven faces. Past research has shown women's attractiveness judgments converge on stubble, which also received intermediate ratings for masculinity and dominance (Dixson and Brooks, 2013; Neave and Shields, 2008). It was hypothesized that there is a threshold at which facial hair enhances attractiveness, such that less attractive feminized facial shape may be more attractive when presented in concert with stubble, while greater masculinity in concert with beardedness may be more attractive at the higher fertility phase of the menstrual cycle (Dixson and Brooks, 2013; Neave and Shields, 2008). Our results provide some support for this hypothesis, as reduced facial masculinity was judged as most attractive when presented in concert with stubble, with attractiveness ratings declining as facial hair and facial masculinity became more pronounced. The current results also suggest variation in estradiol underpins a subtle relationship between facial hair and facial masculinity in women's attractiveness judgments of male faces. However, we note that past research quantifying the hormonal correlates of face preferences among women are mixed. Thus, women with higher estradiol at mid-cycle stated the highest preferences for facial masculinity in between-subjects (Roney and Simmons, 2008) and within-subject preferences for facial masculinity (Roney et al., 2011; Ditzen et al., 2017). However, other studies did not report women's preferences were associated with estradiol in either between-subject (Marcinkowska et al., 2016; Escasa-Dorne et al., 2017) or within-subject studies (Bobst et al., 2014; Dixson et al., 2018b; Jones et al., 2018b; Marcinkowska et al., 2018; Welling et al., 2007). Taken together, the evidence that mid-cycle and peri-ovulatory increases in estradiol are associated with elevated sexual desire for masculine partners is mixed and additional studies testing associations between women's hormone levels over the menstrual cycle and their mate preferences would be valuable.

In past studies, women's preferences for facial hair were positively associated with that of their partners (Dixson et al., 2013; Janif et al., 2014; Valentova et al., 2017) and facial hair interacts subtly with facial masculinity to determine the attractiveness of beardedness (Dixson et al., 2016, 2017a) and strongly enhances facial masculinity, particularly jaw size, to augment ratings of male dominance (Dixson et al., 2017a; Sherlock et al., 2017). Beards may also enhance male attractiveness under conditions of greater intra-sexual competition (Dixson et al., 2005; Dixson et al., 2018a; Grueter et al., 2015). Thus, Barber (2001) showed using cross-sectional data from London spanning 1842–1972 that beards became more popular among men getting married at times when the adult sex ratio was more male-biased. This suggests that men unconsciously choose to be bearded when intra-sexual competition is higher or that female choice underpins frequency-dependence in facial hair styles. Cross-cultural research quantifying the frequency of beards among men reported that men were more bearded in larger cities, where income was less evenly distributed, and women's preferences for beards were strongest (Dixson et al., 2017b). This suggests that between-population differences in men's grooming decisions are associated with both intra-sexual competition and female choice. Experimental research has also shown that women state stronger preferences for facial hair when beards are rare than when they are common (Janif et al., 2014), so that female preferences within populations and men's grooming patterns may be negative frequency-dependent. Alternatively, full untrimmed beards may reduce male facial attractiveness by appearing untidy and of low socioeconomic status rather than due to enhancing masculinity and social dominance. There is some evidence that beards render male faces as looking dirtier than clean-shaven faces (Roll and Verinis, 1971). While self-reported pathogen disgust is positively associated with women's preferences for beards (McIntosh et al., 2017), a role of untidiness and perceived lower socioeconomic status may have impacted on women's attractiveness ratings for beards in our study and further research on how grooming facial hair impacts on judgments of men's facial attractiveness would be valuable.

In conclusion, we did not find evidence in support of the ovulatory shift hypothesis for comparisons between high and low fertile phase attractiveness ratings in either a large sample between-subjects design, or a smaller sample within-subjects design that used LH surge detection to target the high fertile phase. We do report evidence in favor of the ovulatory shift hypothesis for facial masculinity among within-participant and to some extent between-participant fluctuations in salivary hormone levels. However, we acknowledge that these observations are not consistent with a recent study reporting no relationships between salivary hormones and women's attractiveness judgments of facial masculinity using a sample much larger than ours (Jones et al., 2018b) and further replication would therefore be beneficial.

Conflicts of interest

None.

Ethics statement

This research was carried out in accordance with the Declaration of Helsinki and was pre-approved by the Human Ethics Committee at the University of New South Wales.

Funding statement

This study was supported by an ARC Discovery Grant to RCB and BJWD, a University of Queensland Postdoctoral Fellowship to BJWD and an Australian Research Council Future Fellowship to TFD (FT140100291). These funding sources had no further role in the study design, in data collection, analysis and interpretation of the data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.psyneuen.2018.04.007.

References

Aiken, L.S., West, S.G., Reno, R.R., 1991. Multiple Regression: Testing and Interpreting Interactions. Sage.

Barber, N., 2001. Mustache fashion covaries with a good marriage market for women. J. Nonverbal. Behav. 25, 261–272.

Blake, K.R., Bastian, B., O'Dean, S.M., Denson, T.F., 2017a. High estradiol and low progesterone are associated with high assertiveness in women. Psychoneuroendocrinology 75, 91–99.

Blake, K., Dixson, B.J., O'Dean, S.M., Denson, T.F., 2016. Standardized methodological protocols for measuring the effects of fertility on women's behavior: a data-driven approach contrasting counting and hormonal methods. Horm. Behav. 81, 74–83.

Blake, K.R., Dixson, B.J., O'Dean, S.M., Denson, T.F., 2017b. No compelling positive association between ovarian hormones and wearing red clothing when using multinomial analyses. Horm. Behav. 90, 129–135.

Bobst, C., Sauter, S., Foppa, A., Lobmaier, J.S., 2014. Early follicular testosterone level predicts preference for masculinity in male faces – but not for women taking hormonal contraception. Psychoneuroendocrinology 41, 142–150.

Brooks, R.C., Shelly, J.P., Jordan, L.A., Dixson, B.J., 2015. The multivariate evolution of female body shape in an artificial digital ecosystem. Evol. Hum. Behav. 36, 351–358.

DeBruine, L., Jones, B.C., Frederick, D.A., Haselton, M.G., Penton-Voak, I.S., Perrett, D.I., 2010a. Evidence for menstrual cycle shifts in women's preferences for masculinity: a response to Harris (in press) Menstrual cycle and facial preferences reconsidered. Evolut. Psychol. 8, 768–775 (147470491000800416).

DeBruine, L.M., Jones, B.C., Little, A.C., Boothroyd, L.G., Perrett, D.I., Penton-Voak, I.S., et al., 2006. Correlated preferences for facial masculinity and ideal or actual partner's masculinity. Proc. R. Soc. Lond. B 273, 1355–1360.

DeBruine, L.M., Jones, B.C., Smith, F.G., Little, A.C., 2010b. Are attractive men's faces masculine or feminine? The importance of controlling confounds in face stimuli. J. Exp. Psychol. Hum. Percept. Perform 36, 751.

Ditzen, B., Palm-Fischbacher, S., Gossweiler, L., Stucky, L., Ehlert, U., 2017. Effects of stress on women's preference for male facial masculinity and their endocrine correlates. Psychoneuroendocrinology 82, 67–674.

Dixson, A.F., 2012. Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Humans, 2nd edition. .

Dixson, A., Dixson, B., Anderson, M., 2005. Sexual selection and the evolution of visually conspicuous sexually dimorphic traits in male monkeys, apes, and human beings. Ann. Rev. Sex. Res. 16, 1–19.

Dixson, B.J.W., 2018. Is male facial width-to-height ratio the target of sexual selection? Arch. Sex. Behav. 47, 827–828. http://dx.doi.org/10.1007/s10508-018-1184-9.

Dixson, B.J., Brooks, R.C., 2013. The role of facial hair in women's perceptions of men's attractiveness, health, masculinity and parenting abilities. Evol. Hum. Behav. 34, 236–241.

Dixson, B.J., Lee, A.J., Blake, K.R., Jasienska, G., Marcinkowska, U.M., 2018b. Women's preferences for men's beards show no relation to their ovarian cycle phase and sex hormone levels. Horm. Behav. 97, 137–144.

Dixson, B.J., Lee, A.J., Sherlock, J.M., Talamas, S.N., 2017a. Beneath the beard: do facial morphometrics influence the strength of judgments of men's beardedness? Evol. Hum. Behav. 38, 164–174.

Dixson, B.J., Little, A.C., Dixson, H.G.W., Brooks, R.C., 2017c. Do prevailing environmental factors influence human preferences for facial morphology? Behav. Ecol. 28, 1217–1227.

Dixson, B.J., Rantala, M.J., 2016. The role of facial and body hair distribution in women's judgments of men's sexual attractiveness. Arch. Sex. Behav. 45, 877–889.

Dixson, B.J., Rantala, M.J., 2017. Further evidence using a continuous measure of conception probability that women's preferences for male facial and body hair may not change with fecundability. Arch. Sex. Behav. 46, 1159–1160. http://dx.doi.org/10.1007/s10508-017-0973-x.

Dixson, B.J., Rantala, M.J., Melo, E.F., Brooks, R.C., 2017b. Beards and the big city: displays of masculinity may be amplified under crowded conditions. Evol. Hum. Behav. 38, 259–264.

Dixson, B.J., Sherlock, J.M., Cornwell, W.K., Kasumovic, M.M., 2018a. Contest competition and men's facial hair: beards may not provide advantages in combat. Evol. Hum. Behav. 97, 137–1444.

Dixson, B.J., Sullikowski, D., Gouda-Vossos, A., Rantala, M.J., Brooks, R.C., 2016. The masculinity paradox: facial masculinity and beardedness interact to determine women's ratings of men's facial attractiveness. J. Evol. Biol. 29, 2311–2320.

Dixson, B.J., Tam, J.C., Awasthy, M., 2013. Do women's preferences for men's facial hair change with reproductive status? Behav. Ecol. 24, 708–716.

Dixson, B.J., Vasey, P.L., 2012. Beards augment perceptions of men's age, social status, and aggressiveness but not attractiveness. Behav. Ecol. 23, 481–490.

Escasa-Dorne, M.J., Manlove, H., Gray, P.B., 2017. Women express a preference for feminized male faces after giving birth. Adap. Hum. Behav. Physiol. 1, 30–42.

Farthing, M.J.G., Mattei, A.M., Edwards, C.R.W., Dawson, A.M., 1982. Relationship between plasma testosterone and dihydratestosterone concentrations and male facial hair growth. Br. J. Dermatol. 107, 559–564.

Gangestad, S.W., Haselton, M.G., 2015. Human estrus: implications for relationship science. Curr. Opin. Psychol. 1, 45–51.

Gangestad, S.W., Haselton, M.G., Welling, L.L., Gildersleeve, K., Pillsworth, E.G., Burriss, R.P., et al., 2016. How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. Evol. Hum. Behav. 37, 85–96.

Gangestad, S.W., Thornhill, R., 2008. Human oestrus. Proc. R. Soc. Lond. B: Biol. Sci. 275, 991–1000.

Geniole, S.N., Denson, T.F., Dixson, B.J., Carré, J.M., McCormick, C.M., 2015. Evidence from meta analyses of the facial width-to-height ratio as an evolved cue of threat. PLoS One 10 (7), e0132726. http://dx.doi.org/10.1371/journal.pone.0132726.

Geniole, S.N., McCormick, C.M., 2015. Facing our ancestors: judgements of aggression are consistent and related to the facial width-to-height ratio in men irrespective of beards. Evol. Hum. Behav. 36, 279–285.

Gildersleeve, K., Haselton, M.G., Fales, M.R., 2014. Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. Psych. Bull. 140, 1205–1259.

Grueter, C.C., Isler, K., Dixson, B.J., 2015. Are badges of status adaptive in large complex primate groups? Evol. Hum. Behav. 36, 398–406.

Harris, C.R., 2011. Menstrual cycle and facial preferences reconsidered. Sex Roles 64, 669–681.

Harris, C.R., 2013. Shifts in masculinity preferences across the menstrual cycle: still not there. Sex Roles 69, 507–515.

Harris, C.R., Pashler, H., Mickes, L., 2014. 2014. Elastic analysis procedures: an incurable (but preventable) problem in the fertility effect literature. Comment on Gildersleeve, Haselton, and Fales. Psychol. Bull. 140, 1260–1264.

Havlíček, J., Cobey, K.D., Barrett, L., Klapilová, K., Roberts, S.C., 2015. The spandrels of Santa Barbara? A new perspective on the peri-ovulation paradigm. Behav. Ecol. 26, 1249–1260.

Hill, A.K., Hunt, J., Welling, L.L., Cárdenas, R.A., Rotella, M.A., Wheatley, J.R., et al., 2013. Quantifying the strength and form of sexual selection on men's traits. Evol. Hum. Behav. 34, 334–341.

Holzleitner, I.J., Perrett, D.I., 2017. Women's preferences for men's facial masculinity:

trade-off accounts revisited. Adapt. Hum. Behav. Physiol. 3, 304–320.

Janif, Z.J., Brooks, R.C., Dixson, B.J., 2014. Negative frequency-dependent preferences and variation in male facial hair. Biol. Lett. 10 (4), 20130958.

Jasienska, G., 2003. Energy metabolism and the evolution of reproductive suppression in the human female. Acta. Biotheor. 51, 1–18.

Jasienska, G., Kapiszewska, M., Ellison, P.T., et al., 2006a. CYP17 genotypes differ in salivary 17-beta estradiol levels: a study based on hormonal profiles from entire menstrual cycles. Cancer. Epidemiol. Biomarkers. Prevent. 15, 2131–2135.

Jasienska, G., Ziomkiewicz, A., Lipson, S.F., Thune, I., Ellison, P.T., 2006b. High ponderal index at birth predicts high estradiol levels in adult women. Am. J. Hum. Biol. 18, 133–140.

Jasienska, G., Jasienski, M., 2008. Interpopulation, interindividual, intercycle, and intracycle natural variation in progesterone levels: a quantitative assessment and implications for population studies. Am. J. Hum. Biol. 20, 35–42.

Jones, B.C., Little, A.C., Boothroyd, L., DeBruine, L.M., Feinberg, D.R., Smith, M.L., et al., 2005. Commitment to relationships and preferences for femininity and apparent health in faces are strongest on days of the menstrual cycle when progesterone level is high. Horm. Behav. 48, 283–290.

Jones, B.C., Hahn, A.C., Fisher, C.I., Wang, H., Kandrik, M., DeBruine, L.M., 2018a. General sexual desire, but not desire for uncommitted sexual relationships, tracks changes in women's hormonal status. Psychoneuroendocrinology 88, 153–157.

Jones, B.C., Hahn, A.C., Fisher, C.I., Wang, H., Kandrik, M., Han, C., Fasolt, V., Morrison, D., Lee, A., Holzleitner, I., Roberts, S.C., Little, A., DeBruine, L., 2018b. Women's preferences for facial masculinity are not related to their hormonal status. Psychol. Sci. http://dx.doi.org/10.1101/136549.

Jukic, A.M.Z., Weinberg, C.R., Wilcox, A.J., McConnaughey, D.R., Hornsby, P., Baird, D.D., 2008. Accuracy of reporting of menstrual cycle length. Am. J. Epidemiol. 167, 25–33.

Knussman, R., Christiansen, K., 1988. Attributes of masculinity and androgen level. Homo 39, 45–50.

Kruger, D.J., 2006. Male facial masculinity influences attributions of personality and reproductive strategy. Pers. Relat. 13, 451–463.

Lee, K.J., Perrett, D.I., 2000. Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments. Perception 29, 1291–1312.

Lipson, S.F., Ellison, P.T., 1992. Normative study of age variation in salivary progesterone profiles. J. Biosoc. Sci. 24, 233–234.

Little, A.C., Jones, B.C., 2012. Variation in facial masculinity and symmetry preferences across the menstrual cycle is moderated by relationship context. Psychoneuroendocrinology 37, 999–1008.

Little, A.C., Jones, B.C., DeBruine, L.M., 2008. Preferences for variation in masculinity in real male faces change across the menstrual cycle: women prefer more masculine faces when they are more fertile. Pers. Indiv. Diff. 45, 478–482.

Marcinkowska, U.M., Ellison, P.T., Galbarczyk, A., Milkowska, K., Pawlowski, B., Thune Jasienska, G., 2016. Lack of support for relation between woman's masculinity preference, estradiol level and mating context. Horm. Behav. 78, 1–7.

Marcinkowska, U.M., Galbarczyk, A., Jasienska, G., 2018. La donna è mobile? Lack of cyclical shifts in facial symmetry, and facial and body masculinity preferences—a hormone based study. Psychoneuroendocrinology 88, 47–53.

Marečková, K., Weinbrand, Z., Chakravarty, M.M., Lawrence, C., Aleong, R., Leonard, G., et al., 2011. Testosterone-mediated sex differences in the face shape during adolescence: subjective impressions and objective features. Horm. Behav. 60, 681–690.

McIntosh, T.L., Lee, A.J., Sidari, M.J., Stower, R.E., Sherlock, J.M., Dixson, B.J., 2017. Microbes and masculinity: does exposure to pathogenic cues alter women's preferences for male facial masculinity and beardedness? PLoS One 12 (6), e0178206.

Muller, M.N., 2017. Testosterone and reproductive effort in male primates. Horm. Behav. 91, 36–51.

Muscarella, F., Cunningham, M.R., 1996. The evolutionary significance and social perception of male pattern baldness and facial hair. Ethol. Sociobiol. 17, 99–117.

Neave, N., Shields, K., 2008. The effects of facial hair manipulation on female perceptions of attractiveness, masculinity, and dominance in male faces. Pers. Indiv. Diff. 45, 373–377.

Oldmeadow, J.A., Dixson, B.J., 2016. The association between men's sexist attitudes and facial hair. Arch. Sex. Behav. 45, 891–899.

Penton-Voak, I.S., Perrett, D.I., 2000. Female preference for male faces changes cyclically. Further evidence. Evol. Hum. Behav. 21, 39–48.

Penton-Voak, I.S., Chen, J.Y., 2004. High salivary testosterone is linked to masculine male facial appearance in humans. Evol. Hum. Behav. 25 (4), 229–241.

Penton-Voak, I.S., Perrett, D.I., Castles, D.L., Kobayashi, T., Burt, D.M., Murray, L.K., Minamisawa, R., 1999. Menstrual cycle alters face preference. Nature 399, 741–742.

Perrett, D.I., Lee, K.J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D.M., et al., 1998. Effects of sexual dimorphism on facial attractiveness. Nature 394, 884–887.

Peters, M., Simmons, L.W., Rhodes, G., 2009. Preferences across the menstrual cycle for masculinity and symmetry in photographs of male faces and bodies. PLoS One 4 (1), e4138.

Puts, D.A., 2006. Cyclic variation in women's preferences for masculine traits: potential hormonal causes. Hum. Nat. 17, 114–127.

Puts, D.A., Bailey, D.H., Cárdenas, R.A., Burriss, R.P., Welling, L.L., Wheatley, J.R., Dawood, K., 2013. Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle. Horm. Behav. 63, 13–19.

Randall, V.A., 2008. Androgens and hair growth. Dermatol. Ther. 21, 314–328.

Rantala, M.J., Moore, F.R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., Krams, I., 2012. Evidence for the stress-linked immunocompetence handicap hypothesis in humans. Nat. Comm. 3, 694.

Rhodes, G., Chan, J., Zebrowitz, L.A., Simmons, L.W., 2003. Does sexual dimorphism in human faces signal health? Proc. R. Soc. Lond. B: Biol. Sci. 270, S93–S95.

Rhodes, G., Morley, G., Simmons, L.W., 2013. Women can judge sexual unfaithfulness from unfamiliar men's faces. Biol. Lett. 9 (1), 20120908.

Rhodes, G., Simmons, L.W., Peters, M., 2005. Attractiveness and sexual behavior. Does attractiveness enhance mating success? Evol. Hum. Behav. 26, 186–201.

Roll, S., Verinis, J.S., 1971. Stereotypes of scalp and facial hair as measured by the semantic differential. Psychol. Rep. 28, 975–980.

Roney, J.R., Simmons, Z.L., 2008. Women's estradiol predicts preference for facial cues of men's testosterone. Horm. Behav. 53, 14–19.

Roney, J.R., Simmons, Z.L., Gray, P.B., 2011. Changes in estradiol predict within women shifts in attraction to facial cues of men's testosterone. Psychoneuroendocrinology 36, 742–749.

Roney, J.R., Simmons, Z.L., 2013. Hormonal predictors of sexual motivation in natural menstrual cycles. Horm. Behav. 63, 636–645.

Scott, I.M., Penton-Voak, I.S., 2011. The validity of composite photographs for assessing masculinity preferences. Perception 40, 323–331.

Sherlock, J.M., Tegg, B., Sulikowski, D., Dixson, B.J., 2017. Facial masculinity and beardedness determine men's explicit, but not their implicit, responses to male dominance. Adapt. Hum. Behav. Physiol. 3, 14–29.

Small, C.M., Manatunga, A.K., Marcus, M., 2007. Validity of self-reported menstrual cycle length. Ann. Epidemiol. 17, 163–170.

Thornhill, R., Gangestad, S.W., 2006. Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. Evol. Hum. Behav. 27, 131–144.

Tiddeman, B., Burt, M., Perrett, D.I., 2001. Prototyping and transforming facial textures for perception research. IEEE Comput. Graph. Appl. 21, 42–50.

Trotter, M., 1922. A study of facial hair in the White and Negro races. Wash. Univ. Stud. 9, 973–989.

Twisk, J.W.R., 2006. Applied Multilevel Analysis: A Practical Guide For Medical Researchers, 6th edition. Cambridge University Press, Cambridge, UK.

Valentova, J.V., Varella, M., Bártová, K., Štěrbová, Z., Dixson, B.J.W., 2017. Mate preferences and choices for facial and body hair in heterosexual women and homosexual men: effects of sex, population, homogamy, and imprinting-like effects. Evol. Hum. Behav. 38, 241–248.

Welling, L.L., Jones, B.C., DeBruine, L.M., Conway, C.A., Law Smith, M.J., Little, A.C., Feinberg, D.R., Sharp, M.A., Al-Dujaili, E.A., 2007. Raised salivary testosterone in women is associated with increased attraction to masculine faces. Horm. Behav. 52, 156–161.

Whitehouse, A.J.O., Gilani, S.Z., Shafait, F., Mian, A., Tan, D.W., Maybery, M.T., Keelan, J.A., Hart, R., Handelsman, D.J., Goonawardene, M., Eastwood, P., 2015. Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. Proc. R. Soc. Lond. B 282, 20151351.

Windhager, S., Schaefer, K., Fink, B., 2011. Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness dominance, and masculinity. Am. J. Hum. Biol. 23, 805–814.

Wilcox, A.J., Dunson, D.B., Weinberg, C.R., Trussell, J., Baird, D.D., 2001. Likelihood of conception with a single act of intercourse: providing benchmark rates for assessment of post-coital contraceptives. Contraception 63 (4), 211–215.

Wood, D.R., 1986. Self-perceived masculinity between bearded and non-bearded males. Percept. Mot. Skills 62, 769–770.

Wood, W., Carden, L., 2014. Elusiveness of menstrual cycle effects on mate preferences comment on Gildersleeve, Haselton, and Fales (2014). Psychol. Bull. 140, 1265–1271.

Wood, W., Kressel, L., Joshi, P.D., Louie, B., 2014. Meta-analysis of menstrual cycle effects on women's mate preferences. Emot. Rev. 6, 229–249.

Zietsch, B.P., Lee, A.J., Sherlock, J.M., Jern, P., 2015. Variation in women's facial masculinity preference is better explained by genetic differences than by previously identified context-dependent effects. Psychol. Sci. 28, 144–1440.

Ziomkiewicz, A., Ellison, P.T., Lipson, S.F., Thune, I., Jasienska, G., 2008. Body fat, energy balance and estradiol levels: a study based on hormonal profiles from complete menstrual cycles. Hum. Reprod. 23, 2555–2563.