Evolution and Human Behavior 35 (2014) 264–271

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

Women's faces and voices are cues to reproductive potential in industrial and forager societies☆

John R. Wheatley [a], Coren A. Apicella [b], Robert P. Burriss [c], Rodrigo A. Cárdenas [d], Drew H. Bailey [e], Lisa L.M. Welling [a,1], David A. Puts [a,f,*]

[a] Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA
[b] Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA
[c] Department of Psychology, Northumbria University, Newcastle upon Tyne, UK
[d] Department of Psychology, Pennsylvania State University, University Park, PA 16802, USA
[e] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[f] Center for Brain, Behavior, and Cognition, Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Initial receipt 24 July 2013
Final revision received 17 February 2014

Keywords:
Age
Attractiveness
BMI
Face
Fecundity
Testosterone
Voice

ABSTRACT

Women's faces and voices may be cues to their reproductive potential. If so, then individual differences in indices of female fecundity and residual reproductive value, such as hormonal profiles, body composition, and age, should be associated with women's facial and vocal attractiveness to men. However, previous research on these associations is sparse, has rendered mixed results, and is limited to Western samples. The current study therefore explored relationships between correlates of reproductive capability (testosterone levels, age, and body mass index [BMI]) and facial and vocal attractiveness in women from industrial and foraging societies. Women's facial and vocal attractiveness was associated with each of these indicators in at least one of the two samples. The patterns of these associations suggest that women's faces and voices provide cues to both common and unique components of reproductive potential and help explain the evolution of men's mating preferences.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Many of the physical characteristics that we find attractive in others may reflect aspects of underlying mate quality, such as fecundity (Buss, 1989; Symons, 1979; Thornhill & Grammer, 1999). There is generally high agreement among men in their ratings of women's vocal and facial attractiveness, and women's facial attractiveness and vocal attractiveness correlate, suggesting that these features indicate mate quality (Collins & Missing, 2003; Feinberg, 2008; Feinberg, DeBruine, Jones, & Perrett, 2008; Hume & Montgomerie, 2001; Jones, Feinberg, Debruine, Little, & Vukovic, 2008; Little, Jones, & DeBruine, 2011). Moreover, some studies suggest positive associations of fertility and fecundity with female attractiveness in both the face (Bobst & Lobmaier, 2012; Grammer, Fink, Møller, & Thornhill, 2003; Johnston & Franklin, 1993; Law Smith et al., 2006; Pflüger, Oberzaucher, Katina, Holzleitner, & Grammer, 2012; Puts et al., 2013; Roberts et al., 2004) and voice (Bryant & Haselton, 2009; Pipitone & Gallup, 2008; Puts et al., 2013). Much of this evidence concerns within-individual variation in fecundity and attractiveness. For example, women's faces and voices are more attractive during the phase of the menstrual cycle when their progesterone levels are relatively low, and their voices are more attractive when estradiol is high relative to their progesterone levels (Puts et al., 2013). These hormonal states correspond with peak fecundability (probability of conception) within the ovulatory cycle (Baird et al., 1999; Landgren, Undén, & Diczfalusy, 1980). If women's faces and voices also provide information useful for discriminating *between* mates based on reproductive potential, then facial and vocal attractiveness should predict individual differences in indicators of fecundity, such as trait-level hormonal profiles (Apter & Vihko, 1990; Baird et al., 1999; Carmina & Lobo, 1999; Landgren et al., 1980; Lawrence, McGarrigle, Radwanska, & Swyer, 1976; Puts et al., 2013; van Anders & Watson, 2006), age (Andersen, Wohlfahrt, Christens, Olsen, & Melbye, 2000; Menken, Trussell, & Larsen, 1986), and body composition (Ellison, 2003; Frisch, 1987; Grodstein, Goldman, & Cramer, 1994; Hill & Hurtado, 1996; Veleva et al., 2008).

Women's reproductive capacity has been negatively linked to their testosterone (T) levels. In a 13-year longitudinal study of nonclinical

☆ Hadza data collection was funded by Harvard University and the National Science Foundation. This material is based upon work supported by the National Science Foundation (Grant No. DGE1255832 awarded to JRW). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

* Corresponding author.
E-mail address: dap27@psu.edu (D.A. Puts).
[1] Current Address: Department of Psychology, Oakland University, Rochester, MI 48309.

http://dx.doi.org/10.1016/j.evolhumbehav.2014.02.006
1090-5138/© 2014 Elsevier Inc. All rights reserved.

females originally aged 7–17 years, age-adjusted serum T concentrations before any pregnancies were higher in women who subsequently had no pregnancies than in those who became pregnant (Apter & Vihko, 1990). High T concentrations are also associated with polycystic ovarian syndrome (PCOS) and anovulatory cycles (Carmina & Lobo, 1999). Among infertile women, ovulation induction resulted in a pregnancy rate of 75% in women with simple amenorrhea, but only 21% in women with PCOS and elevated androgens (Lawrence et al., 1976). The association between women's T concentrations and infertility appears to be due at least in part to menstrual cycle irregularity, which is negatively related to the likelihood of achieving pregnancy (Small et al., 2010). Higher T levels predicted menstrual cycle irregularity in large, nonclinical samples of both normally cycling and hormonally contracepting premenopausal women (van Anders & Watson, 2006; Wei, Schmidt, Dwyer, Norman, & Venn, 2009), as well as in adolescent girls (Fernandes et al., 2005).

Masculine facial and vocal characteristics are associated with high T in men (Dabbs & Mallinger, 1999; Evans, Neave, Wakelin, & Hamilton, 2008; Penton-Voak & Chen, 2004; Puts, Apicella, & Cárdenas, 2012). If the same is true of women, then we should see a negative relationship between women's T and facial and vocal attractiveness, since masculine faces and voices tend to be considered less attractive in women (Feinberg et al., 2008; Jones, Little, Watkins, Welling, & Debruine, 2011; Perrett et al., 1998; Puts, Barndt, Welling, Dawood, & Burriss, 2011). Surprisingly, associations between women's T levels and their attractiveness have yet to be investigated.

Age is another strong correlate of female reproductive potential. Women's fecundity is low in their teens, peaks in their mid-20s, and gradually declines to zero at menopause around age 45 (Menken et al., 1986). Likelihood of miscarriage mirrors this curvilinear pattern, with the greatest risks experienced by women under 20 or over 35 (Andersen et al., 2000). Men's mate preferences seemingly track these trends in female reproductive potential (Buss, 1989; Kenrick & Keefe, 1992). In contrast to fecundity, women's residual reproductive value (expected future offspring) peaks at reproductive maturity and declines thereafter. Correspondingly, researchers have consistently reported negative associations between female perceived or actual age and facial attractiveness (Fink, Grammer, & Matts, 2006; Jones & Hill, 1993; McLellan & McKelvie, 1993), including a study in which teenage female faces were rated along with young adult and older adult faces (Röder, Fink, & Jones, 2013). Few researchers have investigated age and vocal attractiveness, though Collins and Missing (2003) found that female voices perceived as more youthful were judged to be more attractive, and Röder et al. (2013) found that the voices of women aged 19–30 were rated as more attractive than those of girls aged 11–15 and women aged 50–65.

Body mass index (BMI), defined as body mass (in kg) divided by height (in m) squared (Tovée, Reinhardt, Emery, & Cornelissen, 1998), may positively predict female fecundity among foragers (Ellison, 2003; Hill & Hurtado, 1996), where men tend to prefer women with higher BMIs (Sugiyama, 2004; Wetsman & Marlowe, 1999; Yu & Shepard, 1998). Men in such populations may utilize body fat as an indicator of a woman's energetic reserves available for bearing and nursing offspring. In industrial populations, where procuring sufficient food is less exigent, both low and high levels of BMI are associated with decreased female fertility (Grodstein et al., 1994; Rich-Edwards et al., 1994; Veleva et al., 2008). Men in these populations tend to prefer women of moderate BMI (Tovée, Hancock, Mahmoodi, Singleton, & Cornelissen, 2002; Tovée, Maisey, Emery, & Cornelissen, 1999). Given these context dependencies, BMI should positively predict facial and vocal attractiveness in foraging societies, and negatively predict facial and vocal attractiveness in industrial societies with a high mean BMI. Among Western women, a negative relationship between BMI and facial attractiveness has been reported by some studies (Collins & Missing, 2003; Davis, Shuster, Dionne, & Claridge, 2001; Hume & Montgomerie, 2001), yet others have reported no significant relationship (Hönekopp, Bartholomé, & Jansen, 2004; Thornhill & Grammer, 1999). Studies that examine facial adiposity (a correlate of BMI; Tinlin et al., 2013) provide more consistent results. Greater facial adiposity has been negatively associated with facial attractiveness (Coetzee et al., 2012), as well as indicators of physical and reproductive health (Coetzee, Perrett, & Stephen, 2009; Coetzee, Re, Perrett, Tiddeman, & Xiao, 2011; Tinlin et al., 2013). Associations between BMI and vocal attractiveness have received much less attention, and the few studies on this topic have produced mixed results (Collins & Missing, 2003; Hughes, Dispenza, & Gallup, 2004). No study of which we are aware has investigated these relationships among foragers.

In sum, associations between individual variation in women's reproductive potential and their facial and vocal attractiveness are not well established. Yet, if women's faces and voices are cues to fecundity and reproductive value, then T, age, and BMI should predict women's facial and vocal attractiveness. Some relevant associations have been studied insufficiently or not at all, and, where associations have been explored, results are inconsistent, which have hindered firm conclusions. Moreover, most samples have been drawn from Western universities. Because such populations are relatively buffered from environmental stressors, including pathogenic infection and nutritional deficit, the expression of mate quality indicators may be less variable than in populations leading more traditional lifestyles.

We examined relationships between the indicators of reproductive potential discussed above (testosterone, age, and BMI) and facial and vocal attractiveness in women 18 years and older from both industrial (Study 1) and foraging (Study 2) populations. Specifically, we predicted that women's (1) T levels would negatively predict facial and vocal attractiveness; (2) age would relate negatively linearly and/or negatively quadratically to attractiveness, with vocal and facial attractiveness peaking between sexual maturity in the mid-teens and peak fecundity in the mid-20s; and (3) BMI would relate to facial and vocal attractiveness negatively in the industrial population and positively in the foraging population.

2. Study 1: U.S. Sample

2.1. Methods

2.1.1. Participants

Three hundred forty-eight women (194 taking hormonal contraception) participated in this research as part of a larger study involving siblings. This study was approved by the ethics board at Michigan State University and is in accord with the Helsinki Declaration. The mean age ($\pm$ SD) of the participants was 20.0 $\pm$ 1.6 years (range 18–27). Self-reported ethnicities were 92.4 percent White, 3.0 percent Asian, 1.0 percent Hispanic or Latino, 1.0 percent Black or African American, 0.3 percent Native Hawaiian or Other Pacific Islander, 0.5 percent American Indian or Alaska Native, and 1.9 percent Other.

2.1.2. Procedures

Naturally-cycling participants were scheduled for two laboratory sessions according to self-reported menstrual cycle length and date of the beginning of the last menstrual bleeding. One session was scheduled to coincide with participants' late follicular phase, and the other was scheduled to occur during the mid-luteal phase, according to the methods of Puts (2006). Session order was counterbalanced across participants, and sessions occurred between 1300 h and 1600 h to minimize any influence of circadian hormonal fluctuations.

Because cyclic changes in E and P potentially confound relationships between T and attractiveness (Puts et al., 2013), we examined the influence of T in women taking oral contraception (OC), whose cyclic hormonal variation is suppressed. To capture diurnal T fluctuations, we scheduled participants taking OC for both a morning and an evening session, approximately one week apart, according to the methods of Puts et al. (2010). OC-using participants were randomly allocated to attend their first session during the morning or evening, with their second session taking place at the other time of day. Morning sessions began between 0820 h and 1000 h, and evening sessions began between 1720 h and 1900 h.

2.1.3. Saliva collection and hormonal analysis

Participants collected approximately 9 ml of saliva in sodium azide-treated polystyrene test tubes during both sessions. Contamination of saliva samples was minimized by having participants not eat, drink (except plain water), smoke, chew gum, or brush their teeth for 1 h before each session. Participants rinsed their mouths with water before chewing a piece of sugar-free Trident gum (inert in salivary hormone assays; Moffat & Hampson, 1996) to stimulate saliva flow. The tube was capped and left upright at room temperature for 18–24 h to allow mucins to settle. Tubes were then frozen at −20 °C until analysis by the Neuroendocrinology Assay Laboratory at the University of Western Ontario, Canada.

We obtained salivary unbound ("free") testosterone concentrations via radioimmunoassay (RIA). Following a double ether extraction, all samples were assayed in duplicate using a Coat-A-Count kit for total testosterone (Diagnostic Products, Los Angeles, CA), modified for use with saliva. RIAs were performed in two batches, sensitivity was 5–10 pg/ml, and the average intra-assay coefficient of variation was 6.3%. Duplicate assay concentrations were highly correlated (morning: $r[163] = .92$, evening: $r(155) = .84$) and were thus averaged. If a value was below detectable levels for one duplicate, then the other was used without averaging. This was the case for one session for two participants.

2.1.4. Data collection

Height was measured from a meter stick affixed to a wall, and weight was obtained using an electronic scale. Participants were provided wet wipes and instructed to remove any makeup, jewelry or glasses and to assume a neutral expression. Facial photographs were taken with a tripod-mounted Canon PowerShot S10 digital camera at a distance of approximately 1 m, a height adjusted to the participant, and standardized lighting conditions. All images were cropped beneath the chin, normalized on interpupillary distance, and rotated so that pupils lay on the same horizontal axis.

Participants were also recorded reading an excerpt from a standard voice passage, the Rainbow Passage (Fairbanks, 1960), using a Shure SM58 vocal cardioid microphone in an anechoic, soundproof booth. A curved wire kept the participant's mouth approximately 9.5 cm from the microphone. Goldwave software was used to record voices in mono at a sampling rate of 44,100 Hz and 16-bit quantization. Recordings were saved as uncompressed .WAV files.

2.1.5. Face and voice ratings

Face photographs and voice recordings were rated by 568 men (mean age: 19.4 ± 1.8 years) at Pennsylvania State University. Each rater assessed one of 30 stimulus sets comprising approximately 25 voice recordings and 25 face photographs. Recordings and photographs were randomly allocated to a set, with the proviso that only one recording or photograph per participant be included in each set. Attractiveness for short- and long-term relationships were rated separately using 7-point Likert scales (7 = very attractive). The order in which participants completed the rating tasks (short- vs. long-term first, faces vs. voices first) was random across participants, as was the order in which stimuli were presented. Each stimulus set was rated by ≥15 raters (mean = 18.9). The first 15 ratings obtained of each voice and face stimulus were averaged to produce composite ratings of short- and long-term attractiveness for each photograph and recording. The remaining ratings were discarded.

2.1.6. Data treatment

Trait values correlated highly across sessions (weight: $r[243] > .99$, height: $r[243] = .98$) and were thus averaged (Table 1). Attractiveness ratings also correlated highly across mating contexts (short- vs. long-term; faces: $rs > 0.94$, voices: $rs > 0.93$) and sessions (faces: $rs > 0.71$, voices: $rs > 0.67$) and were therefore averaged. There was no significant difference in vocal or facial attractiveness between OC users and non-users (vocal $t[262] = -.399$, $p = .690$; facial $t[257] = -1.448$, $p = .149$), so contraceptive status was not included in regression models. Testosterone levels, which were positively skewed and thus log-transformed, exhibited the expected decrease from morning to evening (paired *t*-test: $t[92] = 9.03$, $p < .0001$) and were correlated between sessions $r[93] = .68$, $p < .0001$), indicating reliability in capturing between-subjects differences in T levels.

One-way ANOVAs were conducted to determine if ethnicity influenced our variables of interest. Tests indicated that the ethnicities differed significantly in weight ($F[7,258] = 2.271$, $p = .03$) and BMI ($F[7,258] = 2.61$, $p = .01$); the African American sample ($N = 2$) uniquely differed from other ethnicities on these variables. Therefore, ethnicity was included as a dichotomous variable (African American/non-African American) in regression models with weight or BMI as predictors.

2.1.7. Statistical analysis

We investigated the association of T with facial and vocal attractiveness among 194 OC-using women from 158 unique sibling groups (36 sister pairs and 122 singletons) using multilevel modeling. Multilevel modeling is preferable when observations are not completely independent of each other (nested structure). Ignoring such structure leads to underestimation of standard errors. Therefore, to account for the possibility that sessions within participants were correlated with each other, we nested sessions within participants. In addition, because these analyses included siblings recruited as part of a larger study, we nested participants within sibling pairs to control for any non-independence of siblings. Testosterone was treated as a varying (Level 1) predictor of women's facial and vocal attractiveness. Within- (Level 1) and between-participants (Level 2) variation in T levels were assessed separately as predictors of facial and vocal attractiveness, and time of day (Level 1) was added to these models to investigate whether T effects reflect attractiveness changes over the day, independent of T. We did not assess the influence of any sibling level variables (Level 3) on female attractiveness. Data were analyzed using random intercept multilevel models (using maximum likelihood estimation), using the *nlme* package in R (Pinheiro, Bates, DebRoy, Sarkar, & the R Development Core Team, 2012).

For analyses not involving within-subjects data (i.e., BMI and age; Table 1), Pearson correlation and multiple regression were used to facilitate interpretation of effect sizes and comparison across samples. One sister from each pair was randomly excluded from these analyses to control for any effects of relatedness. To explore quadratic effects of

Table 1
Summary statistics for university women (Study 1).

	N	Mean	S.D.	Range
Age	266	20.03	1.59	18–26
BMI	266	22.95	3.35	16.8–39.5
Height (cm)	266	165.34	6.16	148.0–185.3
Weight (kg)	266	62.90	10.97	39.0–123.1
AM testosterone (pg/ml)	119	20.49	11.92	4.5–67.5
PM testosterone (pg/ml)	113	14.52	7.80	3.0–34.5
Vocal attractiveness	264	3.66	.80	1.7–5.4
Facial attractiveness	259	2.78	.76	1.3–5.0

age, we entered both age and age^2 into our regression models. Age was centered to reduce collinearity between age and age^2. Variance inflation factors (VIFs) indicated that the regression models were unlikely to be confounded by multicollinearity.

2.2. Results

2.2.1. Intraclass correlations

Because participants were grouped together as sister pairs for the purposes of multilevel analyses, we calculated intraclass correlations to explore how much variance in facial and vocal attractiveness could be explained by between-subjects and between-sibling pair effects. Fifty-nine percent of the variance in facial attractiveness was explained by between-subjects and between-sibling pair effects. Of the between subjects variance, 32% was attributable to between-sibling pair effects. With regard to vocal attractiveness, 66% of the variance could be explained by between-subjects and between-sibling pair effects, and 0% of the between subjects variance was attributable to between-sibling pair effects. However, because most participants did not have a sibling in the samples (and were therefore considered to be in their own sibling group), the estimates of the proportion of the between subjects variance attributable to sibling pair should be interpreted with caution. There is some evidence that nesting within sibling pairs is useful in the case of the facial attractiveness models, as one would expect given that previous studies have shown facial attractiveness to be heritable (e.g., Mitchem et al., 2013). Regardless, nesting subjects within sibling pairs in multilevel modeling is a conservative approach; as the sibling variance approaches 0, the results will increasingly resemble those from linear regression.

2.2.2. Testosterone and facial attractiveness

We first entered testosterone (T) into a model to predict facial attractiveness. Its effect was significant ($t[114] = -3.42$, $p < .0001$; $n = 189$; regression weight $= -.51$; SE $= .15$). When session time (morning or afternoon) was entered into the model, session time was not statistically significant ($t[113] = -.73$, $p = .47$; $n = 189$; regression weight $= .10$; SE $= .14$), but the effect of T remained significant ($t[113] = -3.39$, $p = .001$; $n = 189$; regression weight $= -.56$; SE $= .17$).

Next, we partitioned T values into between-participants components (each participant's mean T value across sessions) and within-participants components (for each session, the difference between a participant's T for that session and her mean T across sessions) (Puts et al., 2010). Only participants with data for both sessions ($n = 126$) were used. Testosterone was related to facial attractiveness at both the within-participants ($t[114] = -2.15$, $p = .03$; $n = 125$) and between-participants ($t[20] = -2.55$, $p = .02$; $n = 125$) levels. When session time was entered into the model, it was not statistically significant ($t[113] = .12$, $p = .90$; $n = 125$; regression weight $= .02$; SE $= .17$), nor was the within-participants component of T ($t[113] = -1.57$, $p = .12$; $n = 125$; regression weight $= -.46$; SE $= .29$), but the between-participants component remained statistically significant ($t[20] = -2.54$, $p = .02$; $n = 125$; regression weight $= -.63$; SE $= .25$; Table 2).

2.2.3. Testosterone and vocal attractiveness

Testosterone was first entered into a model to predict vocal attractiveness. Its effect was not statistically significant ($t[121] = -1.05$, $p = .30$; $n = 192$; regression weight $= -.16$; SE $= .15$). Next, we partitioned T values into between- and within-participants components, as above. T was related to attractiveness only at the within-participants level ($t[121] = -1.99$, $p = .049$; $n = 125$). When session type (morning or afternoon) was entered into the model, it was statistically significant ($t[120] = 3.01$, $p = .003$; $n = 125$; regression weight $= .46$; SE $= .15$), but the within-participants component of T was not ($t[120] = .20$, $p = .85$; $n = 125$; regression weight $= .05$; SE $= .26$; Table 2).

Table 2
Summary of multilevel modeling analyses of associations between testosterone (T) and facial and vocal attractiveness (Att) in university women (Study 1).

Model	Outcome	Predictor	Controls	Regression Weight	SE	t	p
1	Facial Att	T		−.51	.15	−3.42	<.0001
2		T	Time	−.56	.17	−3.39	.001
3	Facial Att	T (within)		−.48	.23	−2.15	.03
4		T (between)		−.63	.25	−2.55	.02
		T (within)	Time	−.46	.29	−1.57	.12
		T (between)	Time	−.63	.25	−2.54	.02
5	Vocal Att	T		−.16	.15	−1.05	.30
		T	Time	.08	.17	.49	.63
6	Vocal Att	T (within)		−.41	.21	−1.99	.049
7		T (between)		.05	.26	.19	.85
		T (within)	Time	.09	.26	.33	.74
		T (between)	Time	.05	.26	.20	.85

2.2.4. Facial attractiveness, age, and BMI

Facial attractiveness correlated negatively with age ($r[259] = -.16$, $p = .01$) and BMI ($r[259] = -.43$, $p < .0001$; Table 3). In a multiple regression model with age and age^2 as predictors ($R^2 = .03$, $F[2,256] = 3.32$, $p = .04$), age significantly predicted facial attractiveness ($\beta = -.14$, $t = -2.18$, $p = .03$) while age^2 did not ($\beta = -.03$, $t = -.48$, $p = .63$). A model with ethnicity, age, BMI, and BMI^2 was significant overall ($R^2 = .19$, $F[4,254] = 15.11$, $p < .0001$), though no individual predictor reached statistical significance (all $ps > .2$). In a model with ethnicity, age, and BMI as predictors ($R^2 = .19$, $F[3,255] = 20.19$, $p < .0001$), BMI significantly predicted facial attractiveness ($\beta = -.41$, $t = -7.04$, $p < .0001$) while ethnicity ($\beta = .07$, $t = 1.25$, $p = .21$) and age ($\beta = -.08$, $t = -1.33$, $p = .19$) did not. Because women with particularly high BMIs might have disproportionately influenced the association between BMI and facial attractiveness, we checked for the presence of influential observations by measuring the leverage and Cook's distance of each observation (values > 1 are considered influential; Cook & Weisberg, 1982). No outliers were detected.

Because the use of ratio variables in regression equations can lead to a loss of information regarding the individual components of the ratio (Kronmal, 1993), the components of BMI (height and weight) were included separately in a regression model to assess their independent associations with facial attractiveness. A model with ethnicity, age, height, and weight as predictors of facial attractiveness ($R^2 = .19$, $F[4,254] = 15.16$, $p < .0001$) showed weight ($\beta = -.46$, $t = -6.89$, $p < .0001$) to be the main driver of the association, though height was a significant predictor as well ($\beta = .15$, $t = 2.29$, $p = .02$; Table 4).

2.2.5. Vocal attractiveness, age, and BMI

Vocal attractiveness correlated negatively with BMI ($r[264] = -.15$, $p = .02$), but not age ($r[264] = .05$, $p = .45$). A multiple regression model with ethnicity, age, age^2, BMI, and BMI^2 was significant overall

Table 3
Correlations between age, BMI, facial attractiveness and vocal attractiveness among university women (Study 1).

	Age	BMI	Facial Attractiveness
BMI	.20**		
Facial Attractiveness	−.16*	−.43***	
Vocal Attractiveness	.05	−.15*	.17**

Note. *p < .05, **p < .01, ***p < .001.

Table 4
Summary of key multiple linear regression models of facial and vocal attractiveness for university women (Study 1).

Outcome	Predictors	β	t	p	R^2
Facial Attractiveness*	age	−.14	−2.18	.03	.03
	age^2	−.03	−.48	.63	
Facial Attractiveness****	ethnicity	.07	1.25	.21	.19
	age	−.08	−1.33	.19	
	BMI	−.41	7.04	.0001	
Facial Attractiveness****	ethnicity	.07	1.29	.20	.19
	age	−.07	−1.28	.20	
	height	.15	2.29	.02	
	weight	−.46	−6.89	.0001	
Vocal Attractiveness**	ethnicity	.17	2.80	.006	.06
	age	.07	1.20	.23	
	BMI	−.14	−2.22	.03	
Vocal Attractiveness***	ethnicity	.16	2.70	.007	.08
	age	.08	1.24	.22	
	height	−.09	−1.31	.19	
	weight	−.14	−1.92	.056	

Note. Overall model *p < .05, **p < .01, ***p < .001, ****p < .0001.

(R^2 = .06, F[5,258] = 3.31, p = .006), though ethnicity was the only significant predictor (β = .173, t = 2.83, p = .005; all other ps > .4). In a model with ethnicity, age, and BMI as predictors of vocal attractiveness (R^2 = .06, F[3,260] = 5.21, p =.002), ethnicity (β = .17, t = 2.80, p = .006) and BMI (β = −.14, t = −2.22, p = .03) were significant predictors, but age was not (β = .07, t = 1.20, p = .23). As above, we checked for the presence of influential observations in our model by measuring the leverage and Cook's distance of each observation. Again, no outliers were detected. We then included the components of BMI separately in a regression model to assess their individual associations with vocal attractiveness. In a model with ethnicity, age, height, and weight as predictors (R^2 = .08, F[4,259] = 5.48, p < .001), ethnicity (β = .16, t = 2.70, p = .007) was a significant predictor, weight (β = −.14, t = −1.92, p = .056) was a marginally significant predictor, and age (β = .08, t = 1.24, p = .22) and height (β = −.09, t = −1.31, p =.22) were nonsignificant. Table 4 summarizes these models.

3. Study 2: Hadza sample

3.1. Methods

3.1.1. Participants

The Hadza, a fulltime hunter–gatherer group, live in remote savannah-woodland areas around Lake Eyasi in northern Tanzania. They subsist on game hunted with bow and arrow and foraged foods such as berries, baobab fruit, honey, and tubers. They are approximately 1000 in number and live in camps of about 30 individuals. Fifty-three women from eight camps participated. This study was approved by the ethics board at Harvard University and is in accord with the Helsinki Declaration.

3.1.2. Data collection

Height was measured with a portable stadiometer, and weight was obtained using an electronic scale (Tanita Ultimate Scale 2000). Full body photographs were taken of participants dressed in black leotards, and photos were printed and cropped at the level of the chest so that shoulders and neck were visible as well as the face. Participants were also audio recorded inside a Land Rover speaking the word "hujambo," which roughly translates as "hello" in English, with a Seinnheiser MKH-60 microphone. Recordings were encoded directly onto computer hard disk in mono using Sonic Foundry's SOUND FORGE at 44,100 Hz sampling rate and 16-bit quantization and saved as uncompressed .WAV files. The Tanzanian government did not grant permission for exporting saliva samples for T assays.

Table 5
Summary statistics for Hadza women (Study 2).

	N	Mean	S.D.	Range
Age	53	32.21	9.09	18–53
BMI	53	20.16	2.01	16.8–25.8
Height (cm)	53	150.62	6.15	140.0–174.5
Weight (kg)	53	45.86	6.41	34.6–64.3
Vocal attractiveness	53	3.50	.57	2.2–4.8
Facial attractiveness	48	4.64	1.37	2.2–7.2

3.1.3. Face and voice ratings

Each facial photograph was judged by an average of 15 Hadza men unfamiliar with the participants. Printed photos of participants were shuffled and placed on the ground in groups of ten. Men were then asked to rank the photographs in order of attractiveness (least attractive = 1). Once the first set of photographs was ranked, another set was presented, and the process continued until every participant had been ranked. Each participant's scores were averaged to give an overall measure of attractiveness.

Each voice recording was rated by 29 men (mean age: 19.4 ± 1.4 years) at Pennsylvania State University. Using 7-point Likert scales, men rated each voice on attractiveness for short- and long-term relationships. The order in which participants completed these two rating tasks was random across participants, as was the order in which stimuli were presented. Ratings were averaged across raters to produce a measure of short- and long-term vocal attractiveness for each participant.

3.1.4. Data treatment and analysis

Because short-term attractiveness and long-term vocal attractiveness were highly correlated (r[51] = .78), they were averaged to create a composite measure of vocal attractiveness (Table 5). Age (Table 5) was log-transformed to correct skew and centered to reduce collinearity between age and age^2. As in Study 1, Pearson correlation and multiple regression were used to explore relationships between facial and vocal attractiveness, age and BMI. VIFs indicated that multicollinearity did not confound regression models.

3.2. Results

3.2.1. Facial attractiveness, age, and BMI

Facial attractiveness (Table 5) significantly negatively correlated with age (r[48] = −.31, p = .031), but not BMI (r[48] = −.06, p = .713; Table 6). In a regression model with age and age^2 as predictors (R^2 = .14, F[2,45] = 3.76, p = .03), age was a significant predictor (β = −.36, t = −2.59, p = .01), but age^2 was not (β = −.18, t = − 1.28, p = .21). A model with BMI (Table 5), age, and age^2 as predictors failed to achieve significance (R^2 = .14, F[3,44] = 2.46, p = .08), as did a model that included BMI^2 as a fourth variable (R^2 = .15, F[4,43] =1.95, p = .12). These models are summarized in Table 7.

3.2.2. Vocal attractiveness, age, and BMI

Age was significantly negatively correlated with vocal attractiveness (r[53] = −.43, p = .001, Table 6), but BMI was not (r[53] = .06, p = .68; Table 6). In a model with age and age^2 as predictors (R^2 = .30, F[2,50] = 10.75, p = .0001), both age (β = −.32, t = −2.53, p = .02)

Table 6
Correlations between age, BMI, facial attractiveness and vocal attractiveness among Hadza women (Study 2).

	Age	BMI	Facial Attractiveness
BMI	.07		
Facial Attractiveness	−.31*	−.06	
Vocal Attractiveness	−.38**	.06	.31*

Note. *p < .05, **p < .01.

Table 7
Summary of key multiple linear regression models of facial and vocal attractiveness for Hadza women (Study 2).

Outcome	Predictors	β	t	p	R^2
Facial Attractiveness*	age	−.36	−2.59	.01	.14
	age^2	−.18	−1.28	.21	
Facial Attractiveness	age	−.34	−2.33	.03	.15
	age^2	−.18	−1.28	.21	
	BMI	1.55	.69	.49	
	BMI^2	−1.58	−.70	.49	
Vocal Attractiveness***	age	−.32	−2.53	.02	.30
	age^2	−.36	−2.88	.006	
Vocal Attractiveness**	age	−.33	−2.54	.01	.31
	age^2	−.35	−2.72	.009	
	BMI	−.56	−.29	.78	
	BMI^2	.63	.32	.75	

Note. Overall model *p < .05, **p = .001, ***p = .0001.

and age^2 ($\beta = -.36$, t $= -2.88$, $p = .006$) were significant predictors (Fig. 1). In a model with age, age^2 and BMI as predictors ($R^2 = .31$, $F[3,49] = 7.184$, $p < .001$), age ($\beta = -.32$, t $= -2.55$, $p = .01$) and age^2 ($\beta = -.36$, t $= -2.83$, $p = .007$) significantly predicted vocal attractiveness, but BMI did not ($\beta = .07$, $t = .59$, $p = .56$). A model that included BMI^2 as a fourth predictor produced similar results ($R^2 = .31$, $F[4,48] = 5.32$, $p = .001$), with age ($\beta = -.33$, $t = -2.54$, $p = .01$) and age^2 ($\beta = -.35$, $t = -2.72$, $p = .009$) significantly predicting vocal attractiveness, though BMI ($\beta = -.56$, $t = -.29$, $p = .78$) and BMI^2 ($\beta = .63$, $t = .32$, $p = .75$) did not (Table 7).

4. Discussion

Prior research suggests that men prefer the physical characteristics of women with high reproductive potential; however, this research is sparse, and results are mixed and limited to Western samples. We therefore investigated the associations of facial and vocal attractiveness with three putative indicators of reproductive potential: testosterone (T), body mass index (BMI), and age. This study is the first to utilize data from both industrial and foraging populations to explore such relationships.

Consistent with results of previous findings (Collins & Missing, 2003), facial attractiveness and vocal attractiveness were significantly positively correlated within a sample of U.S. university students and a sample of Tanzanian foragers. This provides further evidence that facial attractiveness and vocal attractiveness are influenced by common underlying aspects of mate quality and may therefore provide convergent cues to female reproductive potential (Feinberg et al., 2005). However, women's faces and voices also appear to convey some unique information about mate quality: Correlations between facial and vocal attractiveness were modest ($r = .2$-$.3$), and facial and vocal attractiveness differed in their associations with measures of mate quality.

For example, individual differences in T levels predicted facial attractiveness, but not vocal attractiveness. The negative associations of T with facial attractiveness on the one hand, and fecundity (Apter & Vihko, 1990; Carmina & Lobo, 1999; Lawrence et al., 1976) and regular menstrual cycles (Fernandes et al., 2005; van Anders & Watson, 2006; Wei et al., 2009) on the other, may partly explain the importance of female facial attractiveness in men's mate choices. We note here that T concentrations were previously associated with menstrual cycle irregularity in a sample of women using hormonal contraception with T levels of comparable mean (13.6 vs. 14.5 pg/ml) and variance (S.D. = 6.1 vs. 7.8 pg/ml; van Anders & Watson, 2006) to our sample.

Testosterone concentrations have also been positively related to sexual motivation in lactating women (Alder, Cook, Davidson, West, & Bancroft, 1986), and T treatment improved sexual function in middle-aged women (mean age: 40 years) with low libido (Goldstat, Briganti, Tran, Wolfe, & Davis, 2003), and in women aged 31–56 years with impaired sexual function following removal of the ovaries (Shifren et al., 2000). However, in nonclinical, non-lactating young women, T concentrations appear to be unrelated to sexual interest (Edelstein, Chopik, & Kean, 2011; Roney & Simmons, 2013) or sexual behavior (Apter & Vihko, 1990). Moreover, if a relationship between T and sexual motivation exists, it might suggest that T would be positively related to attractiveness—the opposite of what we observed. The negative relationship between T and facial attractiveness is thus more likely to reflect male adaptations for recruiting highly fecund mates than for targeting sexually interested women.

Both facial attractiveness and vocal attractiveness were related to state (within-participant) T levels. Although these relationships became non-significant with session type included in the statistical models, this does not indicate that within-individual changes in attractiveness were uninfluenced by T. Given that T levels covaried with session type, T may have been the causal mechanism driving diurnal changes in attractiveness. The evidence for this was stronger in the case of facial attractiveness, as the effect of time of day also became non-significant with T included in the model, suggesting that T mediated the effect of time of day. Experimental manipulation of T levels may allow the effects of T and time of day to be disentangled in future research. Because of the observational nature of the current study, we are unable to determine whether T, time of day, or a third variable related to both (e.g., cortisol; Aedo, Landgren, & Diczfalusy, 1981) most strongly influenced facial or vocal attractiveness within individuals.

The associations that we found between attractiveness and age also suggest that women's faces and voices are at least partly non-redundant cues to reproductive potential. Younger women had more attractive faces in both the U.S. and Hadza samples, and more attractive voices in the Hadza sample. Age was also curvilinearly related to vocal attractiveness in Hadza women, with peak attractiveness in the mid-twenties. (The narrow age range of participants likely hindered our ability to detect any curvilinear relationship in the U.S. sample.) These results closely parallel those of Röder et al. (2013), who found in a German sample that facial attractiveness decreased with age from around puberty. In contrast, vocal attractiveness in the Röder et al. study peaked in early adulthood (19–30), was lower around puberty, and was still lower around menopause. Although faces and voices likely provide information about both residual reproductive value and current fecundity, these age trends suggest that women's voices provide relatively greater information about current fecundity, which peaks in the mid-twenties (Andersen et al., 2000; Menken et al., 1986), whereas women's faces provide relatively greater information about residual reproductive value, which peaks at

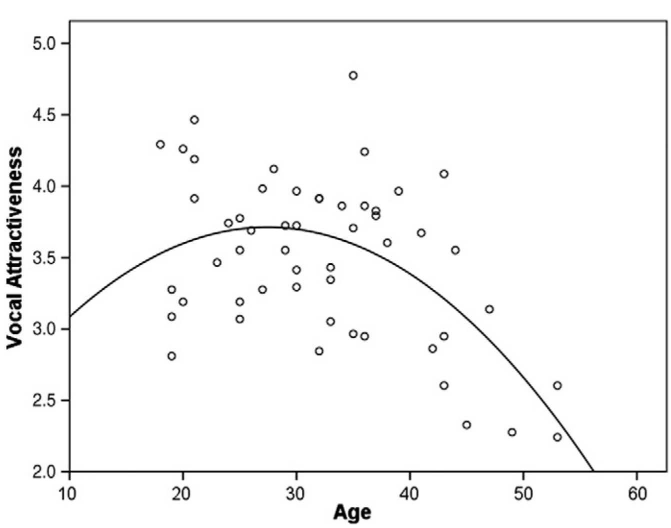

Fig. 1. Vocal attractiveness peaks at approximately 26 years of age among Hadza women.

sexual maturity (see also Puts, Jones, & DeBruine, 2012). In support of the latter hypothesis, Confer, Perilloux, and Buss (2010) found that men more often chose to view women's faces than their bodies when asked to evaluate the women's attractiveness for a long-term relationship vs. a short-term relationship.

Of course, age can influence mate choice at the proximate level only because it has perceptible phenotypic correlates. In our university sample, age positively predicted BMI, and with BMI statistically controlled, age no longer significantly predicted facial attractiveness. Though BMI also has its own independent influence on reproductive capability and fecundity, (Grodstein et al., 1994; Lake, Power, & Cole, 1997; Veleva et al., 2008; Zaadstra et al., 1993), this result supports the interpretation that facial adiposity provides important cues to age-related changes in reproductive capability (Coetzee et al., 2009; Coetzee et al., 2011; Tinlin et al., 2013). In fact, a lower BMI predicted greater facial and vocal attractiveness in this sample. When BMI was deconstructed into its constituent variables, weight was the main driver of these associations, a result consistent with our predictions and previous research (Coetzee, Chen, Perrett, & Stephen, 2010; Davis et al., 2001; Hume & Montgomerie, 2001; Tovée et al., 1999; Tovée et al., 2002).

However, BMI and weight were uncorrelated with facial and vocal attractiveness among Hadza women. The lack of a significant association may reflect the smaller sample and narrower range of BMI values (16.8–25.8) compared to our university sample (16.8–39.5; Levine's test for homogeneity: $F[1,317] = 4.24$, $p = .04$). Indeed, when we reduced the range of BMI values in the university sample to that of the Hadza sample and randomly selected a sample of equal size, the correlations of BMI with facial attractiveness ($r[53] = -.17$, $p = .235$) and vocal attractiveness ($r[53] = .04$, $p = .765$) were no longer statistically significant. It is also possible that BMI has a stronger influence on attractiveness in industrial populations, where overweight individuals are stigmatized (Puhl & Heuer, 2009).

One potential concern is that we measured T only in participants who were taking OC, and that the observed relationships with T reflect effects of OC. We believe this is unlikely. While OC use lowers free (bioavailable) T, it does so to a similar degree across OC types (van der Vange, Blankenstein, Kloosterboer, Haspels, & Thijssen, 1990), particularly after the first cycle of use (Wiegratz et al., 2003). Furthermore, the restricted range of T levels in women taking OC should, if anything, weaken associations between T and attractiveness rather than artificially inflating them. Likewise, any direct effects of differences in estradiol and progesterone dosage across OC types would represent statistical noise, which should also weaken relationships with T. Although the observed relationships between T and attractiveness are not likely to reflect effects of OC, we reasoned that it may be easier to detect relationships with T in OC-using women due to suppression of cyclic fluctuations in progesterone and estradiol.

Another potential concern is that, while Hadza women's facial attractiveness was rated by Hadza men, their vocal attractiveness was rated by U.S. university men. Although this limitation should be acknowledged, and future research should explore male voice preferences among the Hadza and other traditional societies, previous research indicates general agreement in men's vocal preferences across societies. For example, both U.S. university men (Puts et al., 2011) and Hadza men (Apicella & Feinberg, 2009) prefer higher-pitched voices in women.

4.1. Summary

Our results suggest that variables affecting women's reproductive potential are associated with facial and vocal attractiveness, and thus that women's faces and voices are cues to underlying mate quality (Feinberg, 2008; Feinberg et al., 2005). Trait T levels, a direct determinant of reproductive physiology, strongly predicted perceptions of facial attractiveness, and there was some evidence of a within-subjects effect of changes in T on both facial and vocal attractiveness. To our knowledge, this is the first study to report relationships between women's T levels and their facial or vocal attractiveness. Age, another variable with global and well-specified effects on reproductive physiology, negatively predicted facial attractiveness in both samples and curvilinearly predicted vocal attractiveness among Hadza women, with vocal attractiveness peaking in the mid-twenties. These relationships with age suggest that women's faces provide relatively more information about residual reproductive value, whereas women's voices provide relatively greater information about fecundity. BMI significantly predicted perceptions of both facial and vocal attractiveness in our university sample, but not in our Hadza sample, probably due to reduced sample size and BMI range in the traditional sample. Taken together, these data suggest that women's faces and voices convey information about both common and unique components of reproductive potential.

Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2014.02.006.

Acknowledgments

We thank Marc Breedlove and Cynthia Jordan for research support, Bradly Alicea, Jill Armington, Emily Barben, Julia Barndt, Michael Burla, Kathryn Cheney, Michael DiRaimo, Melina Durhal, Rebecca Frysinger, Sana Khan, Mallory Leinenger, Heather Malinowski, Samantha Melonas, Ernestine Mitchell, Joe Morehouse, Diana Rosaleyra, Charlene Scheld, Kevin Singh, Sara Sutherland, Lisa Vroman, Tyesha Washington, and Molly Zolianbawi for assistance with study preparation and data collection, and Elizabeth Hampson and Bavani Rajakumar for assistance with hormonal assays. CAA would like to thank the Tanzanian Commission on Science and Technology, Audax Mabulla, and Deus Haraja.

References

Aedo, A. R., Landgren, B. M., & Diczfalusy, E. (1981). Studies on ovarian and adrenal steroids at different phases of the menstrual cycle: II. A comparative assessment of the circadian variation in steroid and lutropin levels during the follicular, periovulatory and luteal phases. *Contraception*, *23*, 407–424.

Alder, E. M., Cook, A., Davidson, D., West, C., & Bancroft, J. (1986). Hormones, mood and sexuality in lactating women. *British Journal of Psychiatry*, *148*, 74–79.

Andersen, A. N., Wohlfahrt, J., Christens, P., Olsen, J., & Melbye, M. (2000). Maternal age and fetal loss: Population based register linkage study. *British Medical Journal*, *320*, 1708–1712.

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter–gatherers. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1077–1082.

Apter, D., & Vihko, R. (1990). Endocrine determinants of fertility: Serum androgen concentrations during follow-up of adolescents into the third decade of life. *Journal of Clinical Endocrinology and Metabolism*, *71*, 970–974.

Baird, D. D., Weinberg, C. R., Zhou, H., Kamel, F., McConnaughey, D. R., Kesner, J. S., et al. (1999). Preimplantation urinary hormone profiles and the probability of conception in healthy women. *Fertility and Sterility*, *71*, 40–49.

Bobst, C., & Lobmaier, J. S. (2012). Men's preference for the ovulating female is triggered by subtle face shape differences. *Hormones and Behavior*, *62*, 413–417.

Bryant, G. A., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters*, *5*, 12–15.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1–49.

Carmina, E., & Lobo, R. A. (1999). Do hyperandrogenic women with normal menses have polycystic ovary syndrome? *Fertility and Sterility*, *71*, 319–322.

Coetzee, V., Chen, J., Perrett, D. I., & Stephen, I. D. (2010). Deciphering faces: Quantifiable visual cues to weight. *Perception*, *39*, 51–61.

Coetzee, V., Faerber, S., Greeff, J., Lefevre, C., Re, D., & Perrett, D. (2012). African perceptions of female attractiveness. *PLoS ONE*, *7*, e48116.

Coetzee, V., Perrett, D. I., & Stephen, I. D. (2009). Facial adiposity: A cue to health? *Perception*, *38*, 1700–1711.

Coetzee, V., Re, D., Perrett, D. I., Tiddeman, B. P., & Xiao, D. (2011). Judging the health and attractiveness of female faces: Is the most attractive level of facial adiposity also considered the healthiest? *Body Image*, *8*, 190–193.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, *65*, 997–1004.

Confer, J. C., Perilloux, C., & Buss, D. M. (2010). More than just a pretty face: Men's priority shifts toward bodily attractiveness in short-term versus long-term mating contexts. *Evolution and Human Behavior*, *31*, 348–353.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*, *Vol. 5*, New York: Chapman and Hall.

Dabbs, J. M., Jr., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences*, *27*, 801–804.

Davis, C., Shuster, B., Dionne, M., & Claridge, G. (2001). Do you see what I see?: Facial attractiveness and weight preoccupation in college women. *Journal of Social and Clinical Psychology*, *20*, 147–160.

Edelstein, R. S., Chopik, W. J., & Kean, E. L. (2011). Sociosexuality moderates the association between testosterone and relationship status in men and women. *Hormones and Behavior*, *60*, 248–255.

Ellison, P. T. (2003). Energetics and reproductive effort. *American Journal of Human Biology*, *15*, 342–351.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology and Behavior*, *93*, 783–788.

Fairbanks, G. (1960). *Voice and articulation drillbook.* New York: Harper.

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology*, *17*, 112–118.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, *37*, 615–623.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., et al. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, *26*, 398–408.

Fernandes, A. R., de Sa Rosa e Silva, A. C., Romao, G. S., Pata, M. C., & dos Reis,, R. M. (2005). Insulin resistance in adolescents with menstrual irregularities. *Journal of Pediatric and Adolescent Gynecology*, *18*, 269–274.

Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior*, *27*, 433–442.

Frisch, R. E. (1987). Body fat, menarche, fitness and fertility. *Human Reproduction*, *2*, 521–533.

Goldstat, R., Briganti, E., Tran, J., Wolfe, R., & Davis, S. R. (2003). Transdermal testosterone therapy improves well-being, mood, and sexual function in premenopausal women. *Menopause*, *10*, 390–398.

Grammer, K., Fink, B., Møller, A. P., & Thornhill, R. (2003). Darwinian aesthetics: Sexual selection and the biology of beauty. *Biological Reviews of the Cambridge Philosophical Society*, *78*, 385–407.

Grodstein, F., Goldman, M. B., & Cramer, D. W. (1994). Body mass index and ovulatory infertility. *Epidemiology*, *5*, 247–250.

Hill, K., & Hurtado, M. A. (1996). *Ache life history: The ecology and demography of a foraging people.* New York: Aldine de Gruyter.

Hönekopp, J., Bartholomé, T., & Jansen, G. (2004). Facial attractiveness, symmetry, and physical fitness in young women. *Human Nature*, *15*, 147–167.

Hughes, S. M., Dispenza, F., & Gallup, G. G., Jr. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, *25*, 295–304.

Hume, D. K., & Montgomerie, R. (2001). Facial attractiveness signals different aspects of "quality" in women and men. *Evolution and Human Behavior*, *22*, 93–112.

Johnston, V. S., & Franklin, M. (1993). Is beauty in the eye of the beholder? *Ethology and Sociobiology*, *14*, 183–199.

Jones, B. C., Feinberg, D. R., Debruine, L. M., Little, A. C., & Vukovic, J. (2008). Integrating cues of social interest and voice pitch in men's preferences for women's voices. *Biology Letters*, *4*, 192–194.

Jones, D., & Hill, K. (1993). Criteria of facial attractiveness in five populations. *Human Nature*, *4*, 271–296.

Jones, B. C., Little, A. C., Watkins, C. D., Welling, L. L., & DeBruine, L. M. (2011). Reported sexual desire predicts men's preferences for sexually dimorphic cues in women's faces. *Archives of Sexual Behavior*, *40*, 1281–1285.

Kenrick, D. T., & Keefe, R. C. (1992). Age preferences in mates reflect sex differences in human reproductive strategies. *Behavioral and Brain Sciences*, *15*, 75–91.

Kronmal, R. A. (1993). Spurious correlation and the fallacy of the ratio standard revisited. *Journal of the Royal Statistical Society A: Statistics in Society*, *156*, 379–392.

Lake, J. K., Power, C., & Cole, T. J. (1997). Women's reproductive health: The role of body mass index in early and adult life. *International Journal of Obesity*, *21*, 432–438.

Landgren, B. M., Undén, A. L., & Diczfalusy, E. (1980). Hormonal profile of the cycle in 68 normally menstruating women. *Acta Endocrinologica*, *94*, 89–98.

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., et al. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 135–140.

Lawrence, D. M., McGarrigle, H. H., Radwanska, E., & Swyer, G. I. (1976). Plasma testosterone and androstenedione levels during monitored induction of ovulation in infertile women with 'simple' amenorrhoea and with the polycystic ovary syndrome. *Clinical Endocrinology*, *5*, 609–618.

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: Evolutionary based research. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *366*, 1638–1659.

McLellan, B., & McKelvie, S. J. (1993). Effects of age and gender on perceived facial attractiveness. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, *25*, 135.

Menken, J., Trussell, J., & Larsen, U. (1986). Age and infertility. *Science*, *233*, 1389–1394.

Mitchem, D., Purkey, A., Grebe, N., Carey, G., Garver-Apgar, C., Bates, T., et al. (2013). Estimating the sex-specific effects of genes on facial attractiveness and sexual dimorphism. *Behavior Genetics*, 1–12.

Moffat, S. D., & Hampson, E. (1996). Salivary testosterone levels in left- and right-handed adults. *Neuropsychologia*, *34*, 225–233.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, *25*, 229–241.

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.

Pflüger, L. S., Oberzaucher, E., Katina, S., Holzleitner, I. J., & Grammer, K. (2012). Cues to fertility: Perceived attractiveness and facial shape predict reproductive success. *Evolution and Human Behavior*, *33*, 708–714.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & the R Development Core Team. (2012). nlme: Linear and nonlinear mixed effects models (version 3.1-103): R package.

Pipitone, R. N., & Gallup, G. G., Jr. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*, *29*, 268–274.

Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity*, *17*, 941–964.

Puts, D. (2006). Cyclic variation in women's preferences for masculine traits. *Human Nature*, *17*, 114–127.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 601–609.

Puts, D. A., Bailey, D. H., Cardenas, R. A., Burriss, R. P., Welling, L. L., Wheatley, J. R., et al. (2013). Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle. *Hormones and Behavior*, *63*, 13–19.

Puts, D. A., Barndt, J. L., Welling, L. L. M., Dawood, K., & Burriss, R. P. (2011). Intrasexual competition among women: Vocal femininity affects perceptions of attractiveness and flirtatiousness. *Personality and Individual Differences*, *50*, 111–115.

Puts, D. A., Cárdenas, R. A., Bailey, D. H., Burriss, R. P., Jordan, C. L., & Breedlove, S. M. (2010). Salivary testosterone does not predict mental rotation performance in men or women. *Hormones and Behavior*, *58*, 282–289.

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, *49*, 227–243.

Rich-Edwards, J. W., Goldman, M. B., Willett, W. C., Hunter, D. J., Stampfer, M. J., Colditz, G. A., et al. (1994). Adolescent body mass index and infertility caused by ovulatory disorder. *American Journal of Obstetrics and Gynecology*, *171*, 171–177.

Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C., et al. (2004). Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society B: Biological Sciences*, *271*, S270–S272.

Röder, S., Fink, B., & Jones, B. C. (2013). Facial, olfactory, and vocal cues to female reproductive value. *Evolutionary Psychology*, *11*, 392–404.

Roney, J. R., & Simmons, Z. L. (2013). Hormonal predictors of sexual motivation in natural menstrual cycles. *Hormones and Behavior*, *63*, 636–645.

Shifren, J. L., Braunstein, G. D., Simon, J. A., Casson, P. R., Buster, J. E., Redmond, G. P., et al. (2000). Transdermal testosterone treatment in women with impaired sexual function after oophorectomy. *New England Journal of Medicine*, *343*, 682–688.

Small, C. M., Manatunga, A. K., Klein, M., Dominguez, C. E., Feigelson, H. S., McChesney, R., et al. (2010). Menstrual cycle variability and the likelihood of achieving pregnancy. *Reviews on Environmental Health*, *25*, 369.

Sugiyama, L. S. (2004). Is beauty in the context-sensitive adaptations of the beholder?: Shiwiar use of waist-to-hip ratio in assessments of female mate value. *Evolution and Human Behavior*, *25*, 51–62.

Symons, D. (1979). *The evolution of human sexuality.* New York: Oxford University Press.

Thornhill, R., & Grammer, K. (1999). The body and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, *20*, 105–120.

Tinlin, R. M., Watkins, C. D., Welling, L. L. M., DeBruine, L. M., Al-Dujaili, E. A. S., & Jones, B. C. (2013). Perceived facial adiposity conveys information about women's health. *British Journal of Psychology*, *104*, 235–248.

Tovée, M. J., Hancock, P. J. B., Mahmoodi, S., Singleton, B. R. R., & Cornelissen, P. L. (2002). Human female attractiveness: Waveform analysis of body shape. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 2205–2213.

Tovée, M. J., Maisey, D. S., Emery, J. L., & Cornelissen, P. L. (1999). Visual cues to female physical attractiveness. *Proceedings of the Royal Society B: Biological Sciences*, *266*, 211–218.

Tovée, M. J., Reinhardt, S., Emery, J. L., & Cornelissen, P. L. (1998). Optimum body-mass index and maximum sexual attractiveness. *The Lancet*, *352*, 548.

van Anders, S. M., & Watson, N. V. (2006). Menstrual cycle irregularities are associated with testosterone levels in healthy premenopausal women. *American Journal of Human Biology*, *18*, 841–844.

van der Vange, N., Blankenstein, M. A., Kloosterboer, H. J., Haspels, A. A., & Thijssen, J. H. (1990). Effects of seven low-dose combined oral contraceptives on sex hormone binding globulin, corticosteroid binding globulin, total and free testosterone. *Contraception*, *41*, 345–352.

Veleva, Z., Tiitinen, A., Vilska, S., Hydén-Granskog, C., Tomás, C., Martikainen, H., et al. (2008). High and low BMI increase the risk of miscarriage after IVF/ICSI and FET. *Human Reproduction*, *23*, 878–884.

Wei, S., Schmidt, M. D., Dwyer, T., Norman, R. J., & Venn, A. J. (2009). Obesity and menstrual irregularity: Associations with SHBG, testosterone, and insulin. *Obesity*, *17*, 1070–1076.

Wetsman, A., & Marlowe, F. (1999). How universal are preferences for female waist-to-hip ratios? Evidence from the Hadza of Tanzania. *Evolution and Human Behavior*, *20*, 219–228.

Wiegratz, I., Kutschera, E., Lee, J. H., Moore, C., Mellinger, U., Winkler, U. H., et al. (2003). Effect of four different oral contraceptives on various sex hormones and serum-binding globulins. *Contraception*, *67*, 25–32.

Yu, D. W., & Shepard, G. H. (1998). Is beauty in the eye of the beholder? *Nature*, *396*, 321–322.

Zaadstra, B. M., Seidell, J. C., Van Noord, P. A., te Velde, E. R., Habbema, J. D., Vrieswijk, B., et al. (1993). Fat and female fecundity: Prospective study of effect of body fat distribution on conception rates. *BMJ*, *306*, 484–487.