Abstract

For a set of color video clips that depict a 360° view of the bodies of 43 young Caucasian women who are within the normal range of percentage of body fat, we show that their attractiveness to both male and female observers depends strongly on their percentage of body fat and their level of skin tanning, but is not significantly related to their cardiovascular fitness (a key health measure). Although evolutionary psychology suggests that physical health should play a role in determining attractiveness judgments, it appears that cardiovascular fitness may be a weak cue, at least in bodies not undergoing cardiovascular exercise. Instead, it seems that more salient cues, such as body mass and skin tanning, are the primary determinants of attractiveness judgments.

Keywords: Female attractiveness; Female health; Cardiovascular fitness

1. Introduction

What drives attractiveness judgments? Some evolutionary psychologists have postulated that humans selected mates based on the display of certain physical cues that honestly signalled one mate to be more desirable (i.e., healthier and with a better reproductive potential) than another (Buss, 1987). Mate selection, in this way, would have enhanced their chances of successful reproduction, as mates with desirable cues are suggested to have been healthier and to have possessed greater reproductive potential. If attractiveness is an honest signal of a better-quality mate, it is, therefore, possible to suggest that an attractive body should also be a healthy body.

To explore this putative relationship in a contemporary population, we measured cardiovascular fitness (i.e., the capacity of the cardiovascular and respiratory systems to supply oxygen to working muscles during sustained physical activity). This is a key index of physical well-being and a strong predictor of longevity (Blair et al., 1996; Farrell, Braun, Cheung, Barlow, & Blair, 2002; Mora et al., 2003). For example, among women of reproductive age (15–44 years old) who are residing in England and Wales, cardiovascular diseases (such as heart disease, stroke, high blood pressure, etc.) are one of the highest causes of mortality (Barlow, Kohl, Gibbons, & Blair, 1995; Office of National Statistics, 2005), suggesting that cardiovascular fitness has a significant impact on health and mortality. Cardiovascular fitness has also been found to provide some protective benefits against cancer (e.g., Batty & Thune, 2000). Thus, although cardiovascular fitness only addresses a component of phenotypic quality, it is a component that should be related to judgments of attractiveness, if attractiveness is indeed related to actual health (Weeden & Sabini, 2005). To test this hypothesis, we took digital video clips of a set of volunteers’ bodies of known cardiovascular fitness, which were then rated for attractiveness.

Previous studies have used 2D photographs or simple 3D wire-frame images of real women as stimuli (e.g., Fan, Liu, Wu, & Dai, 2004; Tovee, Hancock, Mahmoodi, Singleton, & Cornelissen, 2002). However, such images fail to capture the full range of potential cues that an observer in real life might use to judge attractiveness. Therefore, in the current study, we used color video clips of bodies that were twice rotated through 360° (see Fig. 1). Within the context of a laboratory setting, this provided...
observers with many more cues than previously available with grey-level 2D images.

2. Methods

Forty-three Caucasian female volunteers, primarily university undergraduates, (age: mean=20.72 years; S.D.=2.19) consented to be filmed to provide stimuli for this study. Video clips were taken of our volunteers using a digital camcorder (with 625-horizontal-line resolution and an 800,000-pixel charge-coupled device chip) as they were twice rotated through 360° on a turntable (see Fig. 1). Clothing was standardized, with all women wearing an unsupportive flesh-colored vest and briefs. All volunteers adopted a standard pose, with feet slightly apart and arms on the sides. The volunteers’ faces were then blurred in video sequences to ensure anonymity.

To measure their cardiovascular fitness, we used a standard 6-min submaximal cycle ergometry test from which a reliable maximal oxygen consumption score (predicted VO₂ max) can be estimated using the Astrand nomogram (Astrand & Rhyming, 1954). A VO₂ max test measures the maximum rate at which oxygen can be absorbed and utilized by an individual during exercise and can quantify an individual’s aerobic work capacity. This test can be administered in either maximal or submaximal form. Maximal tests involve exercising individuals up to their maximal heart rate (HR) or until they reach a plateau of oxygen consumption; submaximal tests work up to only 80% of maximal HR. The former therefore provides an absolute measure of cardiovascular functioning, while the latter does not. However, using the Astrand nomogram and the norms from maximal testing, scores from submaximal tests can be used to reliably predict absolute cardiovascular functioning. A submaximal test was used here as it is less stressful on the individual and can be performed on anyone, regardless of fitness level.

Predicted VO₂ max scores can also be reported in different absolute or relative terms, with the latter taking into consideration the body weight of the individual, allowing for easier comparison across participants. Here we chose to report relative predicted VO₂ max (r-predicted VO₂ max) as...

Fig. 1. Top: A subset of frames extracted from one video sequence, as the volunteer rotates through 360° counterclockwise. Lower left: A scatterplot showing the relationship between body attractiveness and the percentage of body fat. Lower right: A scatterplot showing the relationship between body attractiveness and r-predicted VO₂ max. In both plots, solid lines represent linear regressions between x- and y-variables; dotted lines represent their 95% confidence limits.
VO2 max) scores to ensure that body weight did not influence test scores. For our set of volunteers, r-predicted VO2 max ranged from 23.52 to 56.98 ml/min/kg (mean=38.43; S.D.=8.09). This is representative of the range of values found in the female population at large, and includes individuals from low to high cardiovascular fitness (e.g., Blair et al., 1989; DeVires & Housh, 1994). For American college students, the norm for women aged 20–29 years ranged from 28 ml/min/kg at low fitness to 49 ml/min/kg for high fitness; the 0 and 100 percentiles were 18 and 65 ml/min/kg, respectively (DeVires & Housh, 1994). We also calculated HR recovery time, as this has also been found to be independently predictive of mortality (Cole, Foody, Blackstone, & Lauer, 2000) and, therefore, may also provide a link between health and attractiveness. HR recovery time was defined as the change from peak HR during exercise to that measured after 2 min of recovery (Cole et al., 2000).

We measured volunteers’ weight and height, and then calculated their body mass index (BMI; kg/m²). The volunteers varied in BMI from 18.41 to 26.68 kg/m², with an average of 22.35 kg/m² (S.D.=2.30). Percentage of body fat was estimated from four skinfold thicknesses [biceps, triceps, subcapular (at a 45° angle), and suprailiac (at a 45° angle)] using the protocol and calculations of Durnin and Womersley (1974). We chose to follow the Durnin and Womersley (1974) procedure, as it has been extensively replicated (see Eston, 2003). Although the measurement of body fat obtained from skinfold measurements is usually within 3–5% of that obtained with hydrostatic weighing (McArdle, Katch, & Katch, 2000), this does leave a small margin for error. A comparison of the error rate with either three or seven skinfolds has been assessed by a number of studies (reviewed in Heyward & Stolarczyk, 1996), and it was concluded that both methods are equally good in accurately predicting body fatness. For this reason, we chose to measure fewer skinfolds for practical benefits. It has been reported that combined skinfolds from the biceps, triceps, subcapular, and suprailiac muscles are negatively correlated with body density (measured using hydrostatic weighing) at r = –.78 for British women (Durnin & Rahaman, 1967). Body fat is inversely correlated with body density (fat has a lower density than muscle), so these results indicate a strong correlation between body fat measured by skinfolds and body fat measured by hydrostatic weighing.

Skinfold measurements were taken with Harpenden calipers (calibrated to a constant spring pressure of 10 g/mm²). The women in the images had a percentage of body fat (%BF)=21.10–34.16 (average=27.70; S.D.=3.65). The relatively high correlation between BMI and the percentage of body fat (r = .67, p < .001) suggested that it would have been redundant to include both measures of body fat in our analysis. In previous studies (Tovée, Hancock, Mahmoodi, Singleton, & Cornelissen, 2002; Tovée, Maisey, Emery, & Cornelissen, 1999), BMI has been used in analyses, but the percentage of body fat is a more direct measure of body fat mass. We have therefore used the percentage of body fat, rather than BMI, in the subsequent analyses of our results.

We also obtained a number of biometric ratios from the women, including circumferential waist-to-hip ratio (WHR), circumferential waist-to-chest ratio (WCR), and torso-to-leg ratio (TLR) (calculated by torso length measured from the acromioclavicular joint to the perineum, and leg length as measured from the perineum to the ankle joint). The WHR range was 0.64–0.84, with an average of 0.74 (S.D.=0.04). The WCR range was 0.72–0.89, with an average of 0.80 (S.D.=0.04). The TLR range was 0.75–1.06, with an average of 0.88 (S.D.=0.06).

The volunteers in this image set were all Caucasian. However, even within this group, skin color varies due to factors such as the degree of sun tanning. As the video clips were in (24-bit) color, it was important to quantify this variation in skin tone, as it may have been a factor that influenced observers’ judgements of attractiveness. Using Corel Photo-Paint 10 (Corel, USA), the mean values for the red (R), green (G), and blue (B) channels (range, 0–255) were calculated within a standard size patch of skin on each volunteer’s back (50×60 pixels). The location of the patch was chosen as it was not affected by shadows, body curvature, clothing, or hair. The mean R, G, and B values of the skin patch for each volunteer’s image were then factor-analyzed to compress them to a single value related to skin tone. The optimum factor analysis was without rotation; the first factor held a communality of 2.84 and explained 94.5% of the variance. The individual R, G, and B values correlated with the resultant factor at r ≥ −.95 and p < .001. A positive number on this factor indicated darker skin, and a negative number indicated lighter skin. We called this factor the skin tone index. It is important to note that this factor is simply a covariate related to stimulus coloration and should not be treated as a direct measure of skin color.

The stimulus images were then presented to 20 male (age: average=21.45 years; S.D.=1.50) and 20 female (age: average=20.70 years; S.D.=1.34) Caucasian observers to be rated. For rating, the films were shown in a randomized order on a 17-in. LCD monitor (1280×1024-pixel resolution); each video clip measured 280×576 pixels and took approximately 15 s to rotate twice through 360°. Each observer watched the entire set of video clips to become aware of the range of shapes, sizes, and features of our volunteers. On second viewing, observers were asked to rate the images for attractiveness on a scale of 0–9 using a numeric keypad.

3. Results

3.1. Reliability

Winer’s intraclass reliability for k means showed a high degree of interrater agreement for both male and female
relationships between attractiveness and the percentage of body fat, WCR, and the skin tone index.

To further quantify these results, we ran a multiple regression analysis using PROC REG in SAS (SAS Institute, North Carolina, USA). All explanatory variables were centered by converting them to z-scores (Altman, 1991), and all were entered into the starting model. Next, we used PROC REG to estimate a best-fit model by simultaneously minimizing Mallow’s Cp statistic, maximizing $R^2$, and permitting only explanatory variables that are significant at $p<.05$. According to these criteria, the best-fit model, using centered variables, was $y=0.35x_1-0.60x_2$ (where $y$=attractiveness, $x_1=$skin tone, and $x_2=$percentage of body fat). This simple model explained ~63% of the total variance in ratings; individually, the percentage of body fat could explain ~51% and skin tone could explain ~$32\%$. Given that skin tone and the percentage of body fat were moderately correlated with each other (see Table 1), we also calculated the variance inflation factor for the explanatory variables in the best-fit model; neither exceeded 1.5, which was well within the acceptable limit of 10 (Birkes & Dodge, 1993).

### 4. Discussion

For a set of color video clips representing a 360° view of female bodies, we found that judgements of attractiveness were best explained by the percentage of body fat. This presentation format provides as many visual cues that an observer would see in the real world but in a controlled laboratory setting as possible. Only one previous study has used images that could be rotated through 360°. However, these images were shown as wire frames, rather than as 3D solid shapes with color, texture, surface curvature, shadowing, and shading (Fan et al., 2004). Our stimuli included these features and thus provided a more realistic representation of the human body. Consistent with our results, previous studies on attractiveness judgements that used relatively impoverished stimuli showed similar effects for BMI (which was closely correlated with the percentage of body fat in this sample) and a close volumetric equivalent, the volume height index (e.g., Fan et al., 2004; Tovée et al., 2002, 1999).
The present results also show a positive link between attractiveness and skin tone, with those possessing darker-toned skin receiving higher ratings of attractiveness. Although many cross-cultural studies have suggested that paler skin is generally regarded as more attractive by human populations (Aoki, 2002; Barber, 1995; Bond & Cash, 1992; Hill, 2002; Sahay & Piran, 1997), several studies have suggested that, for Caucasian faces in the Europe and America, moderate levels of tanning are regarded as more attractive than no tanning (e.g., Fink, Grammer, & Thornhill, 2001; Sahay & Piran, 1997). A paler coloring has been linked to both youth and fertility levels during the menstrual cycle (e.g., Frost, 1988; Van den Berghe & Frost, 1986); therefore, one might have argued that paler skin on the female body should be regarded as more attractive and healthy. However, skin color may be a culturally based status symbol indicating that someone with a tan can afford the free time necessary to acquire one, just as, in the 18th century, a paler complexion indicated higher status as it showed that an individual did not have to undertake manual outdoor labor (Etcoff, 1999). If this were the case, then one might expect preference for a darker skin tone to extend to all ethnic groups within a society. However, the preference for a more tanned appearance seems to be largely specific to Caucasians in Western cultures (Bond & Cash, 1992; Hill, 2002; Sahay & Piran, 1997).

In the current study, the correlation matrix showed that WCR was linked to attractiveness, although WHR was not. It has been suggested that women with a large bust relative to waist width (i.e., a low WCR) have higher levels of estrogen, which may in turn be positively associated with a higher probability of conception (Jasienka, Ziomkiewicz, Ellison, Lipson, & Thune, 2004). Thus, WCR might be used as a proxy for estrogen levels, and the prominent positioning of the breasts on the front of the body makes WCR a comparatively easy judgement to make and thus allows it to influence attractiveness ratings. A similar role has also been suggested for WHR (Singh, 1993). However, neither torso ratio survived the multiple regression analysis in this study, most likely because both WCR and WHR were correlated with the percentage of body fat in this sample. [Note: A correlation between the percentage of body fat and WHR is to be expected since large-scale epidemiological samples repeatedly demonstrate correlations between WHR and BMI (e.g., UK Department of Health, 2003).] This suggests that the major effects of the percentage of body fat on this sample of images may have been related more to dietary habits, and any additional influence of estrogen levels on body shape was either missed through sampling effects or simply outweighed by other factors.

Some previous studies using line-drawn figures have reported a strong relationship between WHR and attractiveness judgements (e.g., Hess, 1995; Singh, 1993). However, these line-drawn stimuli covaried BMI and WHR (i.e., as the WHR increases, so does the apparent BMI of the figures); thus, the change in attractiveness rating could be due to changing WHR, BMI, or both (Tovée & Cornelissen, 1999). Studies that have used digital photographs of real women have found a strong effect of changing BMI and a much weaker effect of WHR (e.g., Thornhill & Grammar, 1999; Tovée et al., 1998, 2002). Consistent with this result, other studies that have attempted to independently manipulate shape changes found that, although WHR can be significantly correlated with attractiveness judgements, it is a much weaker cue than BMI (Tovée, Benson, Emery, Mason, & Cohen-Tovée, 2003; Smith, Tovée, Hancock, Cox, & Cornelissen, 2006; Streeter & McBurney, 2003). These findings are consistent with our own results. However, Smith et al. have also shown that there is considerable variability in the shape of torsos (synthesized from four independent features defined by principal components analysis) that are treated as equally attractive, suggesting that observers may make tradeoffs among different biometric attributes when judging attractiveness (Smith, Tovée, Hancock, Bateson, Cox, & Cornelissen, 2006). Ultimately, to elucidate this complex problem, further research is required to identify exactly what image features drive perceptual judgments of attractiveness. But we also need to understand how genetic, environmental, and physiological factors interact to modify phenotypic appearance in the first place. Only then will it be possible to map the relationships between desired quality and biometric proxies such as the percentage of body fat, BMI, and WHR.

Our study shows a comparatively weak and nonsignificant correlation of .19 between attractiveness and VO2 max, which is of a similar order of magnitude to the correlations found in previous studies that explored the relationship between physical health and female facial attractiveness (e.g., Hönkopp, Bartholomé, & Jansen, 2004; Shackelford & Larsen, 1999). Given the strength of this relationship, how many image clips would we need to reach statistical significance? Assuming an effect size of 0.19, reaching statistical significance (α = .05) using a two-tailed test with an acceptable power of 0.80 would require a sample size of 214 image clips. This suggests that information from a very large number of bodies would have to be assimilated in our decision-making process before we could reliably judge the relationship between CRF and attractiveness. Even if we assume that we had obtained a statistically significant result, the modest relationship between attractiveness ratings and CRF means that attractiveness will be very insensitive to big changes in CRF status. For example, as we have stated in Methods, for American college students, the norm for women ranged from 28 ml/min/kg at low fitness to 49 ml/min/kg for high fitness (DeVires & Housh, 1994). At a 0.19 correlation with attractiveness, a simple regression shows that changing one’s fitness from low to high changes one’s attractiveness by only one point (from 4.5 to 5.5). By comparison, changing the percentage of body fat from just underweight to just overweight alters the attractiveness rating by 5 points.
(from 2.5 to 7.5). This fivefold difference in effect size suggests that cues to cardiovascular health are not very salient when the body is inactive, as in our video clips.

Since CRF is a good predictor of long-term health and longevity (e.g., Batty & Thune, 2000; Blair et al., 1996; Farrell et al., 2002; Mora et al., 2003), it is plausible that it should contribute to judgements of attractiveness. We suggest that cues to CRF may only become salient when someone is undertaking an activity that places cardiovascular strain upon them. For example, Brown et al. (2005) have shown that biological motion cues (in this case from dancing) can increase the saliency of cues relevant to sexual selection. Therefore, future experiments using video clips of women moving (e.g., dancing, running, or walking) may show much stronger correlations between attractiveness and cardiovascular health. Alternatively, cardiovascular health and body fat composition may have been closely linked in our ancestral population. One could determine the former by reference to the latter, which may be more perceptually salient. However, as current findings show, in our modern industrialized society, body fat and cardiovascular fitness can be decoupled, potentially making fitness judgements less straightforward and more inaccurate.

Acknowledgments

The authors would like to thank the Northumbria University Sports Science Department for the use of their facilities and equipment. We would also like to thank Melissa Bateson for her comments on the manuscript. This research was supported by an MRC studentship to K.L.S. and Wellcome Trust funding through the JIF initiative to P.L.C.

References


