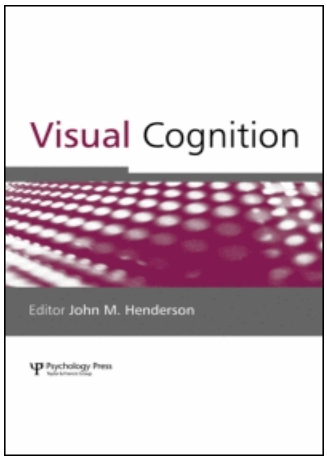

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

An analysis of body shape attractiveness based on image statistics: Evidence for a dissociation between expressions of preference and shape discrimination

First Published on: 09 March 2007
To cite this Article: Smith, Kathryn L., Tovée, Martin J., Hancock, Peter J. B., Bateson, Melissa, Cox, Mike A. A. and Cornelissen, Piers L. , 'An analysis of body shape attractiveness based on image statistics: Evidence for a dissociation between expressions of preference and shape discrimination', Visual Cognition, 1 - 27
To link to this article: DOI: 10.1080/13506280601029515
URL: http://dx.doi.org/10.1080/13506280601029515

VISUAL COGNITION, 0000, 00 (00), 000–000

Psychology Press
Taylor & Francis Group

An analysis of body shape attractiveness based on image statistics: Evidence for a dissociation between expressions of preference and shape discrimination

Kathryn L. Smith and Martin J. Tovée
University of Newcastle, Newcastle Upon Tyne, UK

Peter J. B. Hancock
University of Stirling, UK

Melissa Bateson, Mike A. A. Cox, and Piers L. Cornelissen
University of Newcastle, Newcastle Upon Tyne, UK

We develop an image-driven approach to the question of what makes the shape of a woman's body attractive. We constructed a set of 625 images of female bodies by factorially recombining four independent descriptors of shape derived from a principal components analysis of the variation in natural body shape, and had observers rate these images for attractiveness. We then modelled observers' attractiveness ratings with polynomial multiple regression, using the same shape descriptors as explanatory variables. The resulting model agrees well with existing models based on simple anthropometric indices of shape; however, some interesting new findings emerge. There was considerable variation in the shape of bodies that were judged to be equally attractive. Further experiments confirmed that observers could detect these subtle variations in shape suggesting a dissociation between attractiveness judgement and shape discrimination.

What makes a woman's body shape attractive? So far, the majority of attempts to answer this question have focused on exploring the impact of variation in a few simple anthropometric measures of body shape on

Please address all correspondence to Dr Martin Tovée, Psychology Department, Henry Wellcome Building, Framlington Place, Newcastle University, Newcastle Upon Tyne NE2 4HH, UK. E-mail: m.j.tovee@ncl.ac.uk

This research was funded by a grant from the Medical Research Council, a Wellcome Trust JIF award, and a Royal Society University Research Fellowship to MB.

http://www.psypress.com/viscog DOI: 10.1080/13506280601029515

judgements of attractiveness. Although a clear picture has emerged from these studies, we argue here that the approach has some important limitations, and that a parallel image-driven approach to attractiveness based on an analysis of naturally occurring variation in body shape could yield new insights into what constitutes an attractive body.

The use of simple anthropometric measures of body shape in studies of attractiveness is popular, because they provide a crude means of quantifying important variation in body shape, and are easy to obtain for large samples of women. Two of the most commonly used measures are body mass index (BMI), which is calculated as the ratio of a woman's weight (kg) to her height (m) squared, and is a measure of overall fatness, and waist–hip ratio (WHR), which is calculated as the ratio of the circumference of the waist to the circumference of the hips, and is a measure of fat distribution or curvaceousness (Weeden & Sabini, 2005). Many studies have confirmed that for optimum attractiveness a female body should be slim, but not skinny, with a BMI of approximately 19–20 (Tovée, Maisey, Emery, & Cornelissen, 1999), and have a small waist relative to the hips with a WHR of approximately 0.7 (Singh, 1993a; Streeter & McBurney, 2003; Tovée et al., 1999). Between them BMI and WHR explain the majority of the variance in attractiveness. For example, Tovée et al. (1999) produced a model based on these two measures that explained 76% of the total variance in attractiveness judgements.

Theoretical support for BMI and WHR as potential cues to attractiveness has come from the evolutionary psychology literature. Evolutionary theory predicts that males should have been selected to prefer as their mates females with the maximum reproductive potential, and that females should therefore have been selected to honestly signal their reproductive potential to males. Both BMI and WHR have been shown to correlate with independent aspects of female health and fertility (e.g., Wass, Waldenstrom, Rossner, & Hellberg, 1997; Zaadstra et al., 1993), supporting the theory that BMI and WHR could function as sexual signals providing honest information to prospective mates about different aspects of a woman's reproductive potential (Furnham, Petrides, & Constantinides, 2005; Henss, 1995; Singh, 1993a, 1993b, 1994a, 1994b, 1995).

However, despite the appeal of the above story, we believe that the approach of focusing on simple anthropometric measures of body shape has a number of potential limitations associated with it that need to be recognized and addressed if further advances are to be made in the science of body attractiveness.

PROBLEMS DUE TO CORRELATIONS BETWEEN EXPLANATORY VARIABLES

A major difficulty in determining the relative contributions of BMI and WHR to judgements of attractiveness lies in the fact that these two explanatory variables tend to be correlated in both natural and simulated stimulus sets. In the original studies of WHR using line drawn figures, WHR was varied by altering the torso width around the waist (e.g., Furnham, Tan, & McManus, 1997; Henss, 1995; Singh, 1993a, 1993b, 1994a, 1994b, 1995). However, as the value of the WHR rises, so also does that of the apparent BMI, making it impossible to say whether changes in attractiveness ratings are made on the basis of WHR, BMI or both (Tovée & Cornelissen, 1999; Tovée et al., 1999). The same problem is also found in studies using modified photographic images of women, where WHR is artificially altered by thickening or narrowing the waist or hips. Altering torso width also alters apparent body mass, so WHR and BMI are once again covaried (e.g., Henss, 2000; Rozmus-Wrzesinska & Pawlowski, 2005; Streeter & McBurney, 2003). To overcome the unintentional confounds introduced in studies using artificial stimuli, a number of studies have used samples of unaltered photographic images of real women, where BMI and WHR are both known. Multivariate analysis of attractiveness ratings of these image sets suggests that although both WHR and BMI are significant predictors of female attractiveness, BMI is a far more important factor (Tovée & Cornelissen, 2001; Tovée et al., 1999; Tovée, Reinhardt, Emery, & Cornelissen, 1998). However, because WHR and BMI are strongly correlated in natural image sets it is ultimately impossible to distinguish the actual cues to attractiveness used by raters.

PERCEPTION OF ATTRACTIVENESS MAY BE MORE COMPLEX THAN IS CURRENTLY IMPLIED

There are many different ways of quantifying a shape as complex as the human body, and just because BMI and WHR correlate with judgements of attractiveness, this does not mean that the perceptual system is using these as direct cues to attractiveness. Indeed it is possible that judgements of attractiveness are driven by a more holistic appraisal of body shape, or by cues that we have so far failed to identify. Clear evidence that there is more to an attractive body shape than just its BMI and WHR is provided by a multivariate analysis of attractiveness ratings performed on a set of unaltered photographic images of women (Tovée, Hancock, Mahmoodi, Singleton, & Cornelissen, 2002). In this analysis BMI emerged as the most important source of variation in attractiveness judgements. However, even

for images with very similar BMIs there was consistent variation in attractiveness judgements that was not explained by any of the anthropometric measures (including WHR) entered into the analysis. This result strongly suggests that observers' perceptual judgements are driven by a consistent set of image features, but that these features are not captured by the simple anthropometric indices of shape explored to date. Thus, knowing a woman's BMI and WHR is not sufficient information to accurately predict how attractive her body will be judged.

Reverse engineering

BMI and WHR alone do not uniquely specify the outline of a torso. It therefore follows that by using existing anthropometric models of attractiveness we are unable to reverse-engineer a body with a chosen attractiveness rating. The ability to visualize the shape changes responsible for changes in attractiveness would be one useful outcome of a model that contained a full specification of the shape of the torso. One of the aims of the current paper is therefore to determine the shape configurations that can produce graded changes in our observers' perception of attractiveness.

An image-driven model of attractiveness

In this paper we attempt to address the above deficiencies inherent in existing models of attractiveness by developing a novel image-driven approach to modelling the shape of the human body. We build a mathematical description of body shape that captures the more subtle variation in body shape that is missed by the crude anthropometric indices used to date. Instead of making a priori assumptions about the factors that are likely to contribute to attractiveness judgements, as the starting point for our model we use objective information about female body shape from unaltered photographs of real women reported by Tovée et al. (2002). Their analysis was based on 60 front-view digital photographs of real women sampled from the "normal" BMI range (specified by the sample mean ± 0.5 *SD*, i.e., 18.0–25.8) (see Methods section for biometric details). The image of each woman's torso was divided into 31 slices of equal thickness, and a waveform was generated by plotting the width of each slice against its position in the body. Principal Components Analysis was then used to derive independent descriptors of this waveform. This analysis suggested that female body shape can be adequately described by just four principal components. As can be seen in Figure 1, the first component (PC1) represents changes in overall body width and is highly correlated with BMI. The second component (PC2) captures a significant change in the

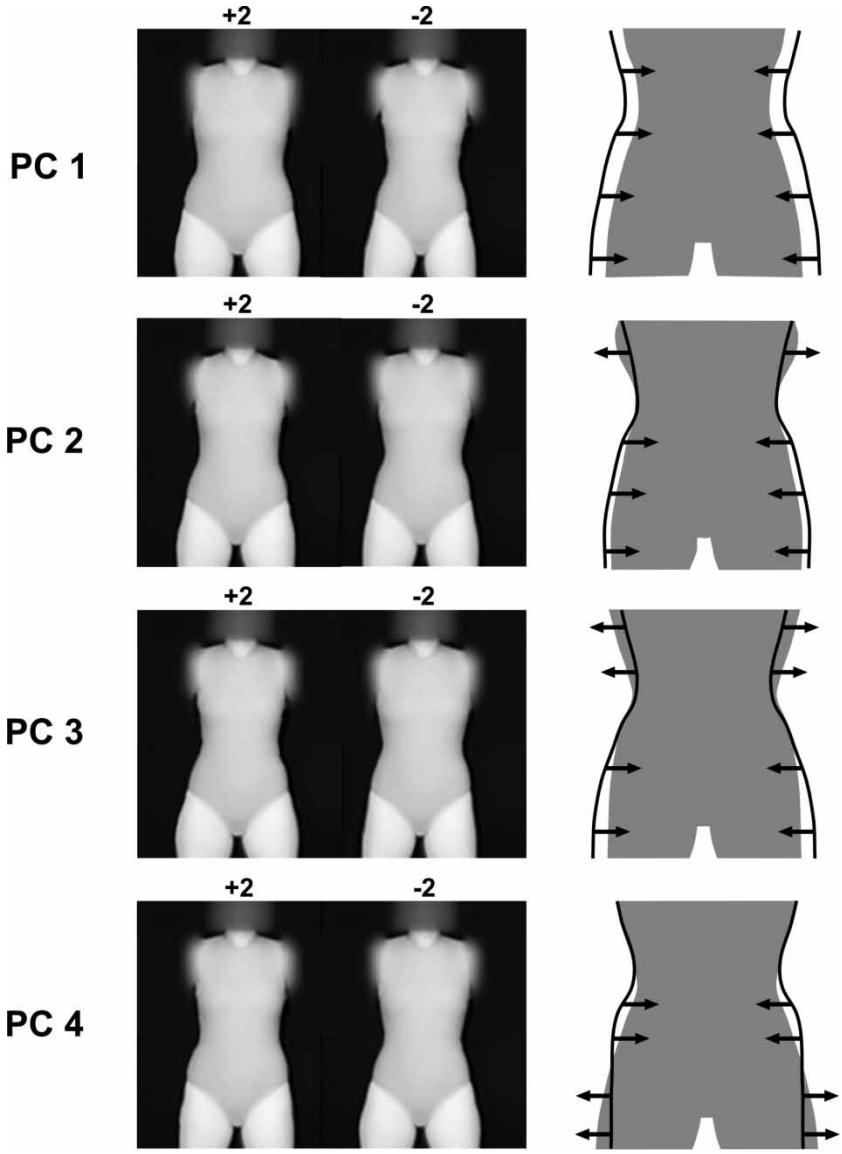

Figure 1. This figure illustrates the relationship between each principal component and torso shape. The left and middle column of images represent values of PC1–PC4 at +2 and −2 (i.e. the two ends of the PC dimensions). The right column shows a cartoon illustrating the change in shape across each PC dimension. The black outline of a torso shape shows the shape at +2, and the grey silhouette the body shape at −2. The black arrows show the direction of movement of body shape from one end of the PC dimension (the +2 shape) towards the other end (the −2 shape).

shape of the hip region, varying from thickset to slender. In addition, PC2 captures increasing chest diameter relative to the waist and hips. As a result, the cumulative effect of changing PC2 is to increase the bust size and slim the hips. The third component (PC3) keeps waist width constant, providing a fulcrum around which the chest and hip widths fluctuate: low values of PC3 are associated with wide hips and a narrow chest, while high values of PC3 capture the opposite effect. The fourth component (PC4) captures simultaneous fluctuations in hip and waist width. It has no effect on the chest. Increases in PC4 can be thought of as adding flesh to the iliac crest while simultaneously removing flesh inferiorly at the level of the greater trochanter. Therefore, positive values of PC4 give a "boxy" appearance to the hip region, and negative values are associated with a low slung, sloped appearance. In identifying four independent descriptors of body shape, Tovée et al.'s analysis supports our earlier assertion that it is likely to be overly simplistic to restrict the analysis of body attractiveness to just BMI and WHR; although WHR is clearly important in defining body shape, a particular value of WHR does not uniquely specify a body shape.

In this paper we use the four principal components derived by Tovée et al. (2002) in order to construct a new set of stimulus bodies. This is possible, because the principal components are statistically independent shape descriptors that can be linearly recombined to create new shapes, each of which is uniquely specified by its particular combination of principal components. In this way, we are able to create a carefully controlled set of body shapes that are based on the natural shape variation in the original sample, and which take into consideration the contours of the entire torso. However, unlike the original 60 images, which corresponded to a sparse sample of all possible body shapes, our new stimulus set of 625 images evenly samples the entire possible stimulus space. Our next step is to obtain attractiveness ratings for the entire new stimulus set, and then use the values of the four principal components as explanatory variables in a multiple regression analysis to model the observers' rating behaviour. A specific advantage of this approach is that since the principal components are by definition independent of one another there are no problems of interpretation arising from correlated independent variables. The final result is a perceptually based model of attractiveness that allows us to predict the attractiveness rating for a body with any given combination of principal component features. The model also gives us the ability to reverse-engineer a body with a given attractiveness rating. The rationale underlying this approach is summarized in Figure 2.

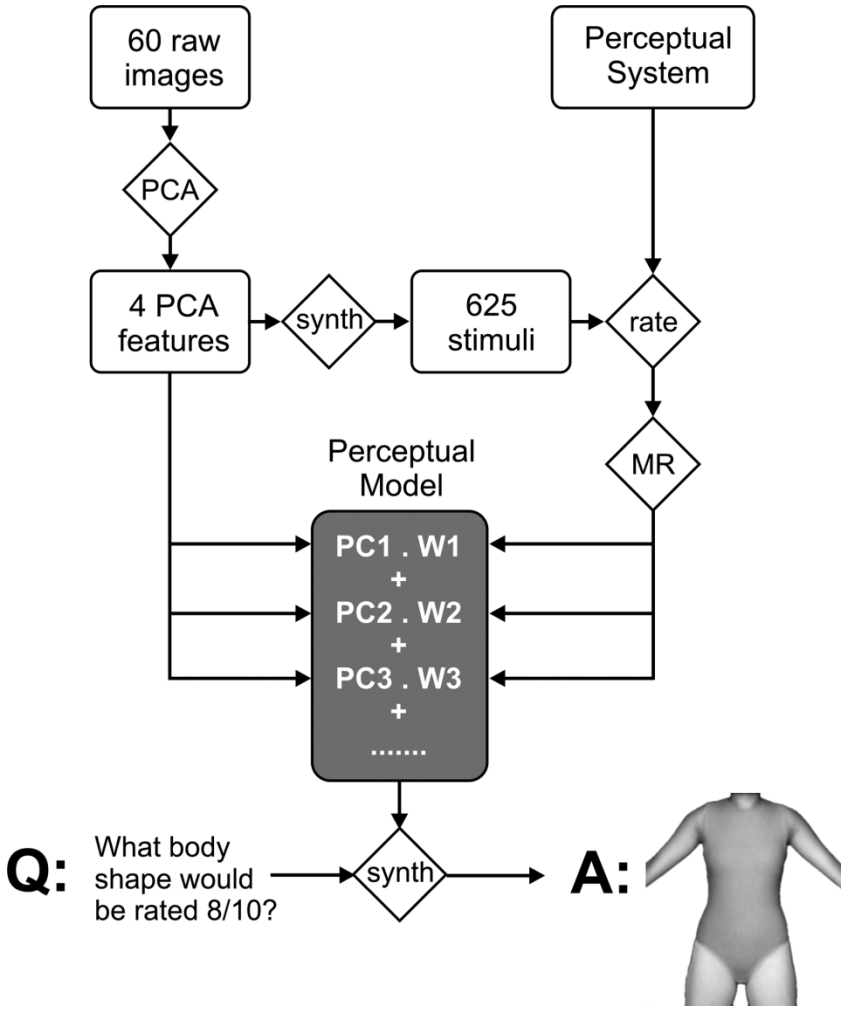

Figure 2. Flow diagram to illustrate the relationships between: the principal components analysis of the starting images; synthesis of 625 stimulus images and their rating for attractiveness; multiple regression modelling of attractiveness ratings to generate a perceptual model; reverse-engineering the shape of a body of a given attractiveness using the perceptual model. PCA = principal components analysis; synth = synthesis; rate = attractiveness rating experiment; MR = multiple regression. W1, W2, and W3 represent the weightings for each component in the perceptual model.

EXPERIMENT 1

Method

Stimulus creation. Our stimuli were based upon the PCA of body shape reported by Tovée et al. (2002), who measured the width of 31 horizontal slices across the torso from 60 images of women in front view. This image set comes from a larger sample of 457 images of women who were all recruited in Newcastle Upon Tyne, UK, and the surrounding area, mainly from the staff and students of the Newcastle Medical School (reported in Tovée et al., 1999). They ranged in age from 18 to 45 (mean age = 26.1 years, $SD = 6.7$). The BMI and WHR ranges from this sample are similar to values reported by other population studies (e.g., Marti et al., 1991). Our set of 60 images represented those women whose BMI fell 0.5 *SD* either side of the sample mean, corresponding to the range 18.0–25.8. The range of WHR for the image analysis subset was 0.66–0.84, representing 1.9 *SD* below the sample mean and 1.4 *SD* above the sample mean. Previous studies have shown that BMI is a strong predictor of attractiveness ratings (e.g., Tovée et al., 1998, 1999), and by holding this variable within a relatively narrow range, the image set will maximize the chances of other shape features influencing attractiveness judgements.

All the new stimulus images were created by manipulating a standard, or reference body. This was created by morphing together the 60 images from the shape analysis reported in Tovée et al. (2002). To do this, a number of points were automatically selected on each body based on the slice points. A consistent set of triangles for these landmarks was calculated, marking out the entire body. The morphing procedure (using Matlab with a customized morphing code) aligned all the landmark points for these bodies to create an average position for them all. The bodies were then morphed to this position and the pixel values averaged across them all to determine the values for the standard body. Further details of the morphing technique can be found in Hancock, Bruce, and Burton (1998).

To create new stimuli from the standard image, its basic shape was again identified through the body slicing method. The four PCs then specified how each point along the standard's body edge should be moved. The points at the edges of each slice were moved as specified by each principal component so as to produce images one and two standard deviations either side of the mean. So, for example, PC1 affects the overall width of the body and the resultant images cover the range of normal variation in body width. The standard body was then morphed to these new positions to create the new stimulus. We used all possible combinations of the four PCs (at -2, -1, 0, $+1$, $+2$ *SD*) to create a total of 625 images in the new image set.

The morphing technique was not applied to the arm region of the standard image. Therefore, although the rest of the torso changed across stimuli, the arms did not. Pilot experiments revealed that the mismatch between arm and torso widths for large values of PC1 particularly enhanced the detection of changes in this cue. To avoid this potential bias in the data, we therefore applied a standard mask to obscure the arms of all 625 stimuli (see Figure 1).

Stimulus anthropometric properties. Although the procedure we adopted should only permit patterns of shape change defined by the variability in the natural images, we wanted to ensure that the synthesized images were still biologically plausible. We therefore extracted standard anthropometric measures from each image, and checked that the ranges of these measures overlapped with those from previous studies using images of real women.

Using Optimas Image Analysis Software version 6.1 (Media Cybernetics, Inc., Maryland, USA), we estimated the BMI of all 625 images in the stimulus set by measuring BMI_{PAR}, i.e., the area of the body divided by the perimeter length around the edge of the body. BMI_{PAR} has been shown to be an accurate proxy for BMI (Tovée et al., 1999). The hip, waist, and chest widths of the images were also measured in pixels, and used to estimate the WHR and WCR as seen from the front view of the body. The horizontal locations for these slices were at: 19%, 44%, and 62% of the total image height respectively. WHR and WCR values derived from measurements taken across the front of bodies in a photograph differ from those derived from circumferential measurements taken from real bodies (see Tovée & Cornelissen, 2001). Therefore, to compare the biometric measures taken from our new stimulus set with the anthropometric data from other image sets, it was necessary to apply the following conversion factors: 1.0552 and 0.8748 to WHR and WCR, respectively. We found that the range of anthropometric measures estimated from the current stimulus set in this way was representative of those used in previous studies (e.g., Fan, Liu, Wu, & Dai, 2004; Thornhill & Grammer, 1999; Tovée et al., 1999). Specifically: $BMI_{PAR} = 20.85–23.55$; $WHR = 0.676–0.827$; $WCR = 0.722–0.905$.

The Pearson correlations between BMI_{PAR} and WHR, BMI_{PAR} and WCR, and WHR and WCR were: $r = .37$, $p < .0001$; $r = .27$, $p < .0001$; $r = -.03$, $p > .1$ respectively. Table 1 shows the correlations between each of the four PCs and the proxy anthropometric measurements.

Observers and procedure. The images were rated individually for attractiveness on a scale of 0–9 by 40 observers: 20 males and 20 females (mean age = 24.5 years, $SD = 4.9$). The images were presented in eight-bit greyscale on a 17-inch LCD monitor (1280×1024 native pixel resolution)

TABLE 1
The correlations between each of the four PC factor scores and the proxy anthropometric measures

	BMI_{PAR}	WHR	WCR
PC1	−0.990**	0.094*	0.353**
PC2	0.031	−0.063	−0.510**
PC3	−0.072	−0.793**	0.235**
PC4	0.040	0.553**	0.694**

*$p < .05$, **$p < .005$, ***$p < .0005$.

positioned approximately 50 cm from the observer, each image measuring 6 cm × 5.5 cm on the viewing screen. The images were presented in five randomized blocks of 125 images each using Superlab Pro version 2 for Windows (Cedrus Corporation, 1999). Observers' responses were recorded by keypress and logged automatically by the stimulus computer. Prior to the test series, the observers were shown a set of 81 representative images to familiarize them with the range of shape variation.

Results

Reliability and gender differences. We found good interrater agreement between observers. Winer intraclass correlations for the mean of k scores was .91 for all observers ($> .79$ for males and females separately). Cronbach's alpha for all observers was .94 ($> .85$ for males and females separately). We also found a high correlation between the male and female attractiveness ratings ($rs = .89$ $p < .001$). Therefore, we pooled data across all observers for further analysis.

Attractiveness modelled with PC scores. Figure 3 shows plots of attractiveness as a function of PC score. Each data point represents the average attractiveness rating across observers and images for each value of each PC. Attractiveness was most strongly influenced by PC1 and least by PC4. In addition, Figure 3 also clearly demonstrates that these relationships are nonlinear, thereby justifying the inclusion of quadratic terms in the multiple regression analysis.

Figure 4 shows a matrix of 3 × 3 contour plots. Each individual plot shows the relationship between PC1 (x-axis), PC2 (y-axis), and attractiveness (z-axis: dark grey to white = low to high attractiveness). In each case the raw data have been smoothed by adjacent averaging. Values of PC3 change across the rows of the matrix, and values of PC4 change across the

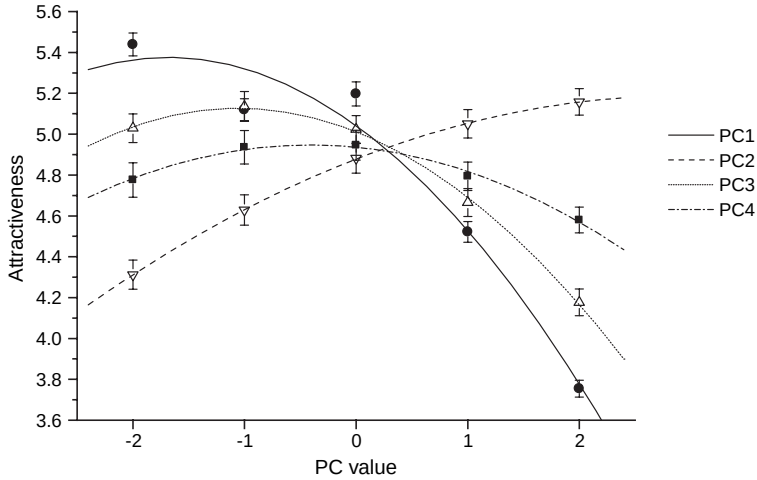

Figure 3. Scatterplots of the relationships between PC1 (filled circles), PC2 (open, inverted triangles), PC3 (open triangles), PC4 (filled squares), and mean attractiveness. Error bars represent one standard error of the mean.

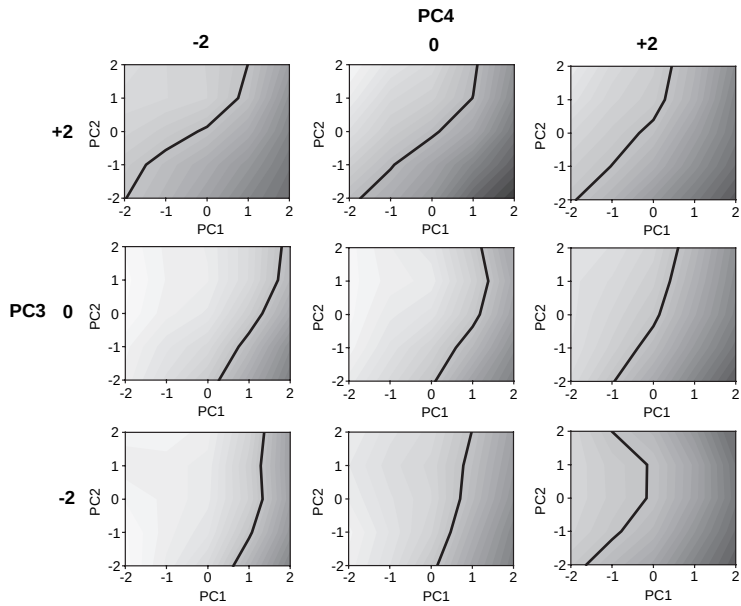

Figure 4. This figure shows a 3×3 matrix of contour plots. Each individual plot shows the relationship between PC1 (x-axis), PC2 (y-axis), and attractiveness (z-axis), where low to high attractiveness is indicated by the greyscale change from dark grey to white. The columns of the matrix correspond to PC4 and the rows to PC3. In each plot the black contour represents an attractiveness of 4.

columns of the matrix. Every plot also contains a high-lighted contour for attractiveness at level 4 (please note, this is an arbitrary choice for purposes of illustration). It is clear that the shape of the highlighted contour changes across the first row of the matrix. Critically, the nature of this shape change is different across successive rows down the matrix. In other words, the effects of PC4 (i.e., the column) on attractiveness is dependent on the values of PC3 (i.e., the row). This pattern, which is evident in the raw data, strongly supports the need to include interaction terms between the four PCs in the multiple regression analysis.

To build the model, we ran the multiple regression analysis using PROC REG in SAS (SAS Institute Inc., North Carolina, USA). All explanatory variables were centred (Altman, 1991). We first estimated the variance in attractiveness ratings explained by a model which contained linear terms only: PC1 ... PC4, for which total $R^2 = 71\%$. Next we added the quadratic terms PC12 ... PC42, for which total $R^2 = 82.7\%$. Then, we added in all possible one-way interactions between PC1 and PC4 (i.e., PC12 = PC1 × PC2, PC13 = PC1 × PC3 ... PC34 = PC3 × PC4), for which total $R^2 = 90.3\%$. Finally, we used PROC REG to optimize the model by simultaneously minimizing Mallow's Cp statistic, maximizing total R^2, and only permitting explanatory variables significant at $p < .05$. According to these criteria, the best-fit model still explained 90.3% of the total variance in attractiveness ratings and is:

$$\begin{aligned}\text{Attractiveness} = {} & 4.81 + 0.091 \times \text{PC1} + 0.15 \times \text{PC2} - 0.062 \times \text{PC3} + 0.047 \\ & \times \text{PC4} - 0.029 \times \text{PC12} - 0.009 \times \text{PC22} - 0.026 \times \text{PC32} \\ & - 0.016 \times \text{PC42} + 0.0037 \times \text{PC13} - 0.0027 \times \text{PC14} \\ & + 0.013 \times \text{PC23} - 0.0048 \times \text{PC24} + 0.025 \times \text{PC34}\end{aligned}$$

As a last step we ran a commonality analysis to estimate the variance attributable to effects associated with PC1, PC2, PC3 and PC4. So, for example, to estimate any effects generally attributable to PC1, we compared the full model to a model with PC1, PC12, PC12, PC13, and PC14 removed. Accordingly, the percentages of variance associated with PC1 ... PC4 were: 50.1%, 14.8%, 24.9%, and 8.4%, respectively.

Attractiveness and anthropometric measures. We wanted to compare our image-driven model of attractiveness with a model based on anthropometric measures. First, to generate scatterplots comparable to those in Figure 3, we split the ranges of BMI_{PAR}, WHR, and WCR into five equal intervals, and used these to calculate the mean attractiveness for all five levels of each variable.

Figure 5 shows plots of the relationships between BMI_{PAR}, WHR, WCR, and attractiveness. Then, using the same model-building procedures

as above, we estimated the variance in attractiveness ratings explained by a model which contained linear terms only: BMI_{PAR}, WHR, and WCR, for which total $R^2 = 66.3\%$. Next we added in quadratic terms and optimized the model by simultaneously minimizing Mallow's Cp statistic, maximizing R^2, and only permitting explanatory variables significant at $p < .05$. It was inappropriate to include interaction terms in the model, particularly between BMI_{PAR} and WHR, because they are well correlated in this image set. According to these criteria, the best-fit anthropometric model explained 78.3% of the total variance in attractiveness ratings and is:

$$\text{Attractiveness} = 4.8 + 295.4 \times \text{WHR} - 201.6 \times \text{WHR}^2 + 12.21 \times \text{BMI}_{\text{PAR}} - 0.29 \times \text{BMI}_{\text{PAR}}^2 - 9.1 \times \text{WCR}$$

Finally, a commonality analysis showed the percentages of variance in attractiveness accounted for by BMI_{PAR}, WHR, and WCR were: 35%, 12%, and 17%, respectively.

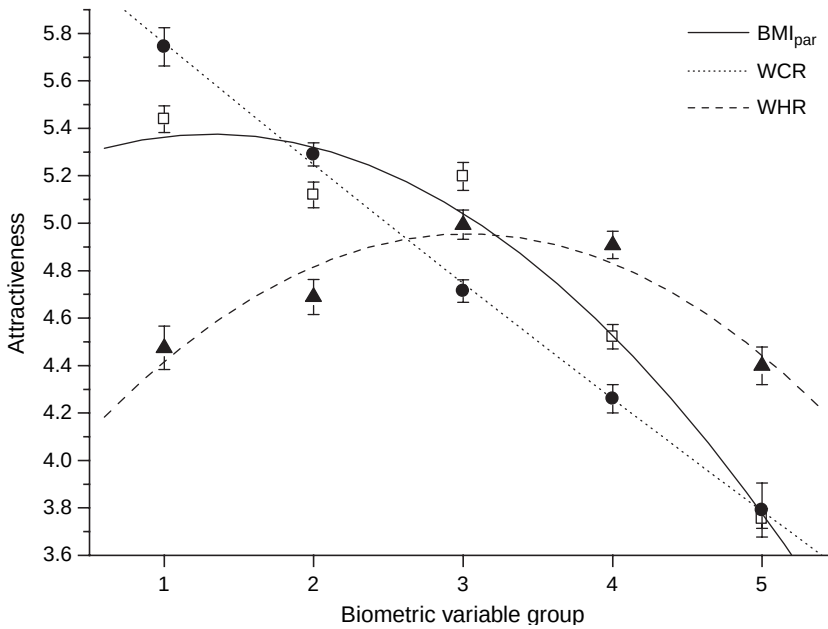

Figure 5. This figure shows scatterplots of the relationships between BMIPAR (open squares), WHR_{front} (filled triangles), WCR_{front} (filled circles), and mean attractiveness. Error bars represent one standard error of the mean.

Visualizing the image-driven model. The best-fit image-driven regression model clearly captures observers' rating behaviour very well. However, we wanted to visualize what this model suggests a more attractive body shape looks like as compared to a less attractive one. In order to do this, we used the regression model to calculate the fitted attractiveness rating for each image, and then ranked these scores. The top panel of Figure 6 shows plots of PC1 . . . PC4 as a function of rank fitted attractiveness, from low to high along the x-axis. The thin lines illustrate how PC values fluctuate dramatically from one image to the next. In comparison, the thicker lines in each plot show the same relationship, but this time after smoothing (by adjacent averaging with a 40 point window). The smoothed plots show the general trends suggested by the multiple regression analysis (such as the inverse relationship between PC1 and attractiveness). However, the raw, unsmoothed plots show that there can be considerable variation in the PC values between one ranked image and the next. To illustrate the consequences of these small-scale fluctuations on the torso shape in the images, we sampled the three regions from the plots in Figure 6, corresponding to the superimposed grey bands A, B, and C. Each band contains six images, and represents a set of images which fell within a narrow range of attractiveness ratings. Band A represents fitted attractiveness values: 2.70–2.87; Band B = 4.60–4.62; Band C = 6.30–6.65. So within each band, the images have roughly the same attractiveness level. The lower panel in Figure 6 shows each of the six stimulus images (and their associated PC values) that belong to these "iso-attractiveness" bands. It is clear from Figure 6 that within each iso-attractiveness band there is considerable, albeit subtle, shape change from one image to the next. However, the fact that the images within a band were rated as equally attractive raises the question of whether observers were insensitive to these shape differences. Alternatively, observers may have been able to detect the differences, but still rated the images at the same attractiveness level (implying that different shape configurations can have the same attractiveness rating). To decide between these alternatives, we ran a second experiment in which we directly tested

Figure 6 (*see over*). The top panel shows plots of PC1 to PC4 as a function of rank fitted attractiveness, from low to high along the x-axis. The thin lines illustrate how PC values vary from one image to the next, and the thicker lines show the same relationship, but this time after smoothing. To illustrate the small-scale fluctuations, we sampled three regions from the plots, corresponding to the superimposed grey bands A, B, and C. Band A represents fitted attractiveness values: 2.70–2.87; Band B = 4.60–4.62; Band C = 6.30–6.65 (please note the width of the grey bands is not to scale). The lower panel shows each of the six stimulus images in each of these three iso-attractiveness bands A, B, and C. Directly below the image of each body is a set of four numbers corresponding to the values of PC1, PC2, PC3, and PC4, respectively, that were used in the synthesis of that image.

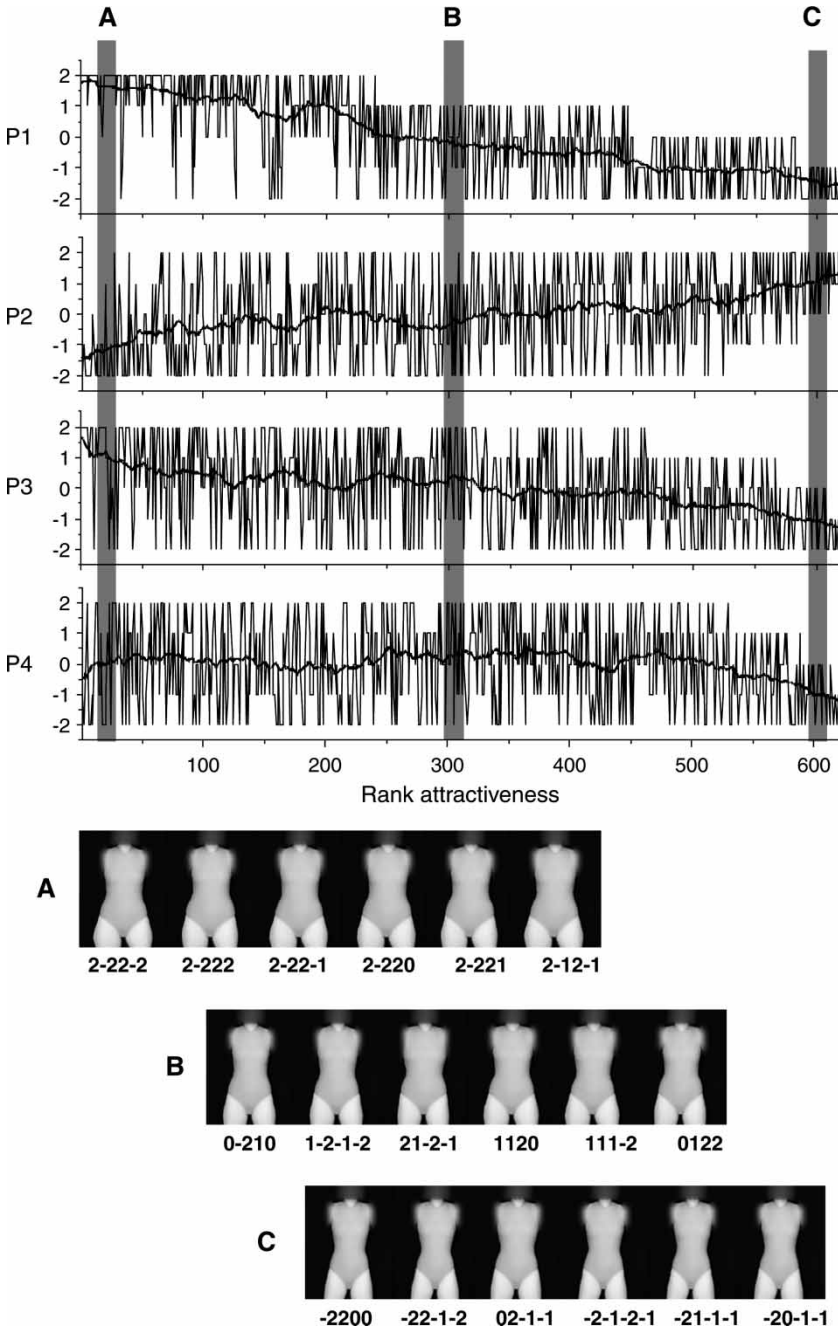

Figure 6. (caption on previous page)

our observers' ability to discriminate between the images in each iso-attractiveness band.

EXPERIMENT 2

Method

Observers and procedure. The 18 images (i.e., 6 from each iso-attractiveness band) were used in two 2-alternative forced-choice (2AFC) tasks; a same–different task and a match-to-sample task. The order in which observers carried out the tasks was counterbalanced. As before, the images were presented in eight-bit greyscale on a 17-inch LCD monitor (1280 × 1024 native pixel resolution) with a neutral grey background. The screen was positioned approximately 50 cm from the observer. Each image measured 6 cm × 5.5 cm on the viewing screen.

For the same–different task, on every trial a pair of images was presented in the middle of the screen for 1500 ms. The images were separated from each other by 3 cm. The pair of images then disappeared, leaving only the neutral grey background. Observers were asked to respond by key press, and without time pressure, whether they thought the two images were the same or not. The task comprised 90 "different" trials (i.e., 30 for each of the three iso-attractiveness bands) and 90 "same" trials, presented at random. For the "different" trials, each of the six images within an iso-attractiveness band was paired with every other image, twice (e.g., Image 1 on the left, Image 3 on the right on one trial, followed by Image 3 on the left and Image 1 on the right on another trial). The 90 "same" trials were constructed by pairing each image within an iso-attractiveness band with itself, five times.

For the match-to-sample task, on each trial a single "target" image was presented in the middle of the screen for 1500 ms. As soon as the target disappeared, two images were presented in the middle of the screen, separated from each other by 3 cm. One was the "target" and the other was a "foil" taken from the same iso-attractiveness band. The correct choice appeared at random in either the left or the right position, with equal probability. Again, observers were asked to respond by key-press, and without time pressure, whether the target was to the left or to the right, at which point the images disappeared. The task comprised 180 trials in total, corresponding to the same image pairings as for the 90 "different" trials in the same–different task (detailed above), but where each image of each pair was presented as the target on separate occasions. These two tasks were carried out by 10 observers (five male and five female), mean age of 23.4 years, $SD = 6.62$.

Results

The probabilities of observers responding correctly in the same–different and match-to-sample tasks were calculated and converted into *d*-prime scores (Gescheider, 1997), as summarized in Table 2.

In this case, *d*-prime provides a bias free estimate of sensitivity to shape difference for each task. Univariate tests of location also showed that observers responded significantly better than chance in both tasks (i.e., $t = 14.1$, $p < .0001$ and $t = 10.1$, $p < .0001$ for same–different and match-to-sample, respectively). These findings show that our observers were perfectly capable of discriminating the subtle shape differences between images within each of the iso-attractiveness bands. Therefore, this strongly suggests that when observers were rating the same images for attractiveness, information about these shape differences either was not actually used in the judgement process, even though it was available, or that the information was used, but the conclusion arrived at was the same.

To further explore this difference between preference and discrimination performance we asked what the chances would be of an observer correctly assigning any one of the six images to its corresponding distribution of attractiveness, i.e., picking the correct one of the six normal distribution curves within a band shown in Figure 7. To run this virtual 6-AFC experiment, we extracted the distributions of raw attractiveness ratings from Experiment 1 for each iso-attractiveness band (i.e., 40 ratings for each of the 18 images). We assumed that the prior chances of any one of the six images being drawn from its iso-attractive set were equal (i.e., 1/6—as with no other information all images are equally likely), and we applied Bayes' theorem to compute the conditional probabilities for each image being correctly assigned to its distribution (see Appendix). If our argument is correct—that the finest scale shape changes seem to be ignored when attractiveness judgements are made—this calculation should show performance no better than guessing.

TABLE 2
The probabilities of observers responding correctly in the same–different and match-to-sample tasks in Experiment 2

		p		*d-prime*	
Task	*Iso-attractiveness band*	*Mean*	*SD*	*Mean*	*SD*
Same–different	A	.67	0.07	1.94	0.23
	B	.72	0.11	2.14	0.39
	C	.73	0.05	2.14	0.19
Match-to-sample	A	.69	0.08	2.03	0.24
	B	.68	0.11	2.01	0.40
	C	.66	0.11	1.92	0.36

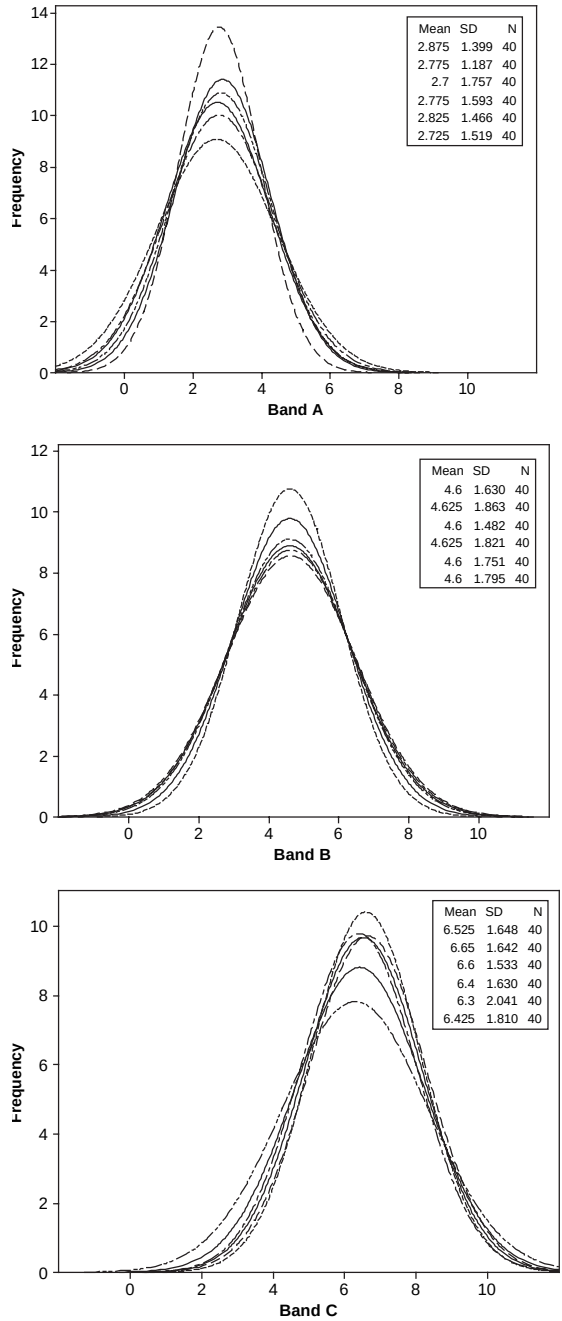

Figure 7. The normal distribution curves of attractiveness for each of the six images in each iso-attractiveness band.

TABLE 3
The conditional probabilities and *d*-prime scores associated with the distributions of attractiveness in Experiment 1 calculated separately for the three iso-attractiveness bands

	Conditional probability		*d-prime*	
Iso-attractiveness band	Mean	SD	Mean	SD
A	.17	0.02	0.30	0.11
B	.17	0.02	0.29	0.07
C	.18	0.05	0.35	0.13

Figure 7 illustrates the normal distribution curves of attractiveness for each of the six images in each iso-attractiveness band. Table 3 shows the mean of the conditional probabilities that were calculated from the mean of each distribution in Figure 7, separately for the three iso-attractiveness bands. In order to compare these results directly to the 2-AFC same–different and match-to-sample experiments, we also calculated the equivalent *d*-prime scores for the conditional probabilities, as if they had been derived from a 6-AFC experiment. These results confirm that an observer would do no better than chance at identifying which distribution a particular image came from, and suggest that a genuine discrepancy exists between preference and discrimination in these data.

EXPERIMENT 3

The results of Experiment 2 suggest a dissociation between attractiveness preference and shape discrimination. However, the sizes of the stimulus sets in the experiments from which this conclusion is drawn are very different: 625 for the attractiveness ratings versus 18 for the discriminations. Therefore, an alternative reason why observers appeared indifferent to subtle shape changes when rating images for attractiveness may simply be related to the differences in the scale of the tasks they undertook. For example, they may only have attended to the most salient differences between images when rating the larger stimulus set. Additionally, if attractiveness judgements of the smaller image set are made in a 2-AFC paradigm, rather than with each image being rated in isolation (as in experiment 1), the observers might be able to reliably discriminate between bodies on the basis of attractiveness. Therefore, in Experiment 3, we tested whether the same apparent indifference to subtle shape change persisted when observers judged the attractiveness of the 18 images from the three iso-attractiveness bands in a 2-AFC task.

Method

Ten observers (five male, five female), mean age of 26.7 years, $SD = 8.5$, carried out a 2-AFC task, in which they had to decide whether the image on the left or the right of a pair of images was most attractive. The set up was the same as for Experiment 2, so that the pairs of images were presented in the middle of the LCD screen for 1500 ms, and they were separated from each other by 3 cm. When the images disappeared, leaving only the neutral grey background, observers were asked to express their preference (left or right image) by keypress. We used the same 18 images (i.e., 6 from each of the three iso-attractiveness bands) as in Experiment 2. Each image was paired with every other image and itself, twice—once with image "X" on the left and "Y" on the right, and a second time with "Y" on the left and "X" on the right (a total of 342 trials per observer).

Results

We analysed three kinds of comparison: within iso-attractiveness band comparisons between nonidentical images, between band comparisons, and same-image comparisons. For the within-band trials, we calculated the probability of each of the six images winning a trial. For between-band trials, we calculated the chances that Band C images beat Band B images; Band C images beat Band A images, and Band B images beat Band A. For same-image trials we calculated the probabilities of the images on the left being treated as more attractive than the images on the right.

Table 4 shows the mean probabilities (and their standard deviations) for the analyses outlined above. Table 4 also shows the results of univariate tests

TABLE 4
A summary of three sets of calculations for the within- and between-band comparisons made from Experiment 2

	Mean probability of an image winning a trial	*SD*	*T-test of $\mu = 0$*
Same-image trials	.50	0.20	$p = 0.8$
Within-band trials			
A	.50	0.13	$p = 1.0$
B	.50	0.20	$p = 1.0$
C	.50	0.13	$p = 1.0$
Between-band trials			
C > A	.92	0.06	$p < 0.0001$
C > B	.83	0.10	$p < 0.0001$
B > A	.82	0.08	$p < 0.0001$

of location. For the same-image trials, these tests demonstrate that observers were as likely to respond to the image on the left as the image on the right. This confirms that observers were not biased in responding to left or right. For the within-band analyses, it is clear that, on average, any particular image had a 50:50 chance of "winning" a trial. This result suggests that images within an iso-attractiveness band are being treated as equivalently attractive. This is consistent with both the original ratings data (Experiment 1) and the Bayesian analysis of these data in Experiment 2. Finally, the between-band analysis confirms that, within the context of a 2-AFC task, the same patterns of preference emerge as when observers are asked to rate images for attractiveness on a Likert scale. To summarize, the results from Experiment 3 suggest that the dissociation between preference and discrimination that we found in Experiment 2 is genuine, and cannot be attributed to differences in task structure and the size of the stimulus set.

Discussion

In this study, we develop a novel, image-driven approach to the question of what makes the shape of a woman's body attractive. To address this question, we constructed a set of female images by factorially recombining four feature dimensions, and had observers rate these images for attractiveness. The four dimensions we used were derived from a PCA of photographs of women who were standing in a fixed pose in front view. We then modelled observers' attractiveness ratings with a polynomial multiple regression, using the same PC dimensions as explanatory variables. The final product is a perceptual model of attractiveness based on the natural pattern of variation in the shape of female bodies.

Our approach was motivated by three limitations inherent in existing anthropometric models of attractiveness. First, our image-driven model solves the problem of correlated explanatory variables by using statistical descriptors of shape that are by definition independent of one another. Second, in modelling the shape of the entire torso our approach has highlighted subtle variations in body shape missed by the simpler anthropometric models. Finally, our model allows for the first time the possibility of reverse-engineering a body with a given attractiveness rating. One consequence of this is that we can actually visualize the shape changes responsible for graded changes in our observers' perception of attractiveness. To give a flavour of the myriad of possibilities opened up by our new approach, Figure 8 shows five bodies drawn from our shape-space of 625 synthesized images. As we point out in the introduction, it is has been difficult to determine the relative contributions of BMI and WHR to attractiveness judgements because these two variables tend to be correlated

Figure 8. A sequence of five of the synthesized images representing minimal change in BMI_{PAR} (i.e., 4% of the variation in the complete image set) and maximal change in WHR (i.e., 69% of the variation in the complete image set). See text for further details.

in both natural and simulated stimulus sets. In this set of five bodies, the BMI is held constant, while the WHR of the images increases from a minimum of 0.65 on the left to a maximum of 0.78 on the right. This figure illustrates the complex shape changes that are necessary to effect a dissociation between BMI and WHR, and that this dissociation is possible using our PCA approach.

At first appearance our approach might seem trivial, because the information that we put into the stimuli was the same as that which we used to model observers' responses. However, this potential criticism misses the point. By creating a stimulus set that comprises all possible combinations of the four PCs, we have extensively sampled the biologically plausible "body-shape space". The body shapes that we created are based on the physical characteristics of the original image set, and these configurations are not determined by any attractiveness judgements. So we are sampling shape space, and not perceptual space. By asking our observers to rate these bodies for attractiveness, we are transposing from shape space to perceptual space. The role of the statistical modelling is to estimate the weights (i.e., regression coefficients) that allow us to make this transposition. To the extent that the resultant model is stable, and accounts for some 90% of the variance in attractiveness ratings, this endeavour succeeded—at least within the limits of our stimulus set. This result shows proof of concept, and suggests that a broadly similar technique could be used to model the relationship between expressions of preference and complex shape variation for other kinds of object. Though beyond the scope of this paper, it would also be valuable to confirm whether we would arrive at a similar solution were we to base the original image analysis on an alternative basis function (e.g., independent components analysis).

The PC model shows very clearly that the change in body shape that drives the perception of attractiveness most strongly is the width of the body (see Figure 6). This result is in good agreement with the anthropometric model, which showed that the strongest predictor of attractiveness was

BMI_{PAR}, which accounted for $\sim 50\%$ of the total variance explained. In comparison, WHR and WCR accounted for smaller, though significant proportions of the variance. These results are consistent with previous studies that have used unaltered photographic images of women. In these studies, BMI is the primary predictor of attractiveness, with shape cues such as WHR and WCR having weaker, but significant effects (Tovée & Cornelissen, 2001; Tovée et al., 1998, 1999, 2002). It is important to note that the results from the current analysis, which are based on an arbitrary statistical image decomposition, should converge on a similar set of physical characteristics as suggested by evolutionary psychology. This suggests that the anthropometric features evolutionary psychologists have picked to explain human attractiveness behaviour, although not necessarily optimal or "correct", are highly correlated with the features actually used. However, there are a number of caveats to this conclusion. For example, the images in this study are impoverished, so the number of physical features available to an observer is limited. Therefore, by using more realistic, 3-D colour representations, other features might come into play. In addition, the current study does not tell us directly what (if any) subset of information in the stimulus images was actually used by observers; whether they attended to the entire image or merely a restricted region or set of regions. Other techniques, like selective masking and eye-movement recording, would be required to address these supplementary questions.

This study has explored variation in torso shape and not variability in leg shape. This is because previous studies have not found the legs to be a significant predictor of attractiveness. For example, Tovée et al. (2002) found that changes in leg width and shape were unrelated to attractiveness. Additionally, several studies have failed to find significant relationships between attractiveness and the ratio between torso length and leg length (TLR) (e.g., Smith et al., 2007; Tovée et al., 1998, 1999). As leg width, shape, and length do not seem to be significant predictors of attractiveness judgements, we concentrated on torso variability in this study.

Perhaps the most interesting aspect of this study is the fact that there is significant variation in the body shapes that are rated as equally attractive. So although Figure 2, which summarized our intentions and expectations going into the study, shows a single outcome when generating a body shape for a particular attractiveness level, our results suggest that in reality multiple configurations can be generated that will be rated as equally attractive. To our knowledge, this phenomenon has not been discussed in the literature before. An obvious explanation is that the shape differences are too subtle to be detected. However, the results of Experiments 2 and 3 suggest this is not the case. Instead, the results suggest that there is a dissociation between preference for body shape (i.e., what observers rated as attractive) and the ability to discriminate between subtle differences in shape. Even

within a subset of bodies, which had all been rated as being of the same level of attractiveness, observers were able to reliably discriminate between them on the basis of their shape. This demonstrates unequivocally that the fact that different PC combinations were treated as equally attractive was not due to insensitivity to the subtle changes in body shape. Instead, it seems that different configurations can actually have the same attractiveness rating.

What might this dissociation between attractiveness preference and discrimination mean? When observers are in attractiveness preference mode, one possibility is simply that "attractiveness space" has multiple maxima, i.e., different configurations of physical features can produce the same level of attractiveness. This suggests that one or more tradeoffs might exist for different body size and shape cues. For example, a body with a higher BMI may be compensated for by a more curvaceous WHR and WCR, and this configuration might be rated as attractive as a body with a more attractive BMI and less curvaceous WHR and WCR. Thus, there may be many routes to being assigned a particular attractiveness level, a theoretical point that is well accepted in the mate choice literature (Wittenberger, 1983). This possibility fits with other findings in human mate choice suggesting that we trade off different attributes. For example, women have been shown to trade off creative intelligence against wealth when choosing potential partners (Haselton & Miller, 2006). Alternatively, observers may be using a very limited subset of body shape cues when they judge attractiveness. Therefore, when in preference mode, as opposed to discrimination mode, they may actually ignore or filter out subtle shape changes that are not directly relevant to the judgement they are making. This interpretation is consistent with the power of BMI, a crude shape discriminator on the basis of body fat distribution, to explain attractiveness ratings. However, it is not sufficient, because BMI and the like do not work well when their range is restricted—so both possibilities may still be true. It may be that when "major" cues are available, such as when there is a wide range of BMI values, then these cues may be used almost exclusively and more subtle cues may be disregarded. But when the BMI range is restricted, the information that the attractiveness judgements are made on may default to a range of more subtle "minor" shape cues.

Of course these images represent only a subset of the cues that are available from a real body. They are greyscale images and so lack colour, a potentially powerful cue to health and fitness. Additionally, it can be argued that 2-D images do not fully capture all the cues to depth and shape available from a 3-D body, and so underplay their importance in determining attractiveness. In real life, when all these cues to colour and shape are available, the role of cue trading may be more significant, as there are more cues to trade.

REFERENCES

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chance. *Philosophical Transactions of the Royal Society*, *53*, 370–418.

Fan, J. T., Liu, F., Wu, J., & Dai, W. (2004). Visual perception of female physical attractiveness. *Proceedings of the Royal Society, London: Series B*, *271*, 347–352.

Furnham, A., Petrides, K. V., & Constantinides, A. (2005). The effects of body mass index and waist-to-hip ratio on ratings of female attractiveness, fecundity and health. *Personality and Individual Differences*, *38*, 1823–1834.

Furnham, A., Tan, T., & McManus, C. (1997). Waist-to-hip ratio and preferences for body shape: A replication and extension. *Personality and Individual Differences*, *22*, 539–549.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals*. London: Lawrence Erlbaum Associates Ltd.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (1998). A comparison of two computer-based face recognition systems with human perceptions of faces. *Vision Research*, *38*, 2277–2288.

Haselton, M. G., & Miller, G. F. (2006). Women's fertility across the cycle increases the short-term attractiveness of creative intelligence compared to wealth. *Human Nature*, *17*, 50–73.

Henss, R. (1995). Waist-to-hip ratio and attractiveness: A replication and extension. *Personality and Individual Differences*, *19*, 479–488.

Henss, R. (2000). Waist-to-hip ratio and attractiveness of the female figure: Evidence from photographic stimuli and methodological considerations. *Personality and Individual Differences*, *28*, 501–513.

Marti, B., Tuomilehto, J., Salomaa, V., Kartovaara, L., Korhonen, H. J., & Pietinen, P. (1991). Body fat distribution in the Finnish population: Environmental determinants and predictive power for cardiovascular risk factor levels. *Journal of Epidemiology and Community Health*, *45*, 131–137.

Rozmus-Wrzesinska, M., & Pawlowski, B. (2005). Men's ratings of female attractiveness are influenced more by changes in female waist size compared with changes in hip size. *Biological Psychology*, *68*, 299–308.

Singh, D. (1993a). Adaptive significance of female physical attractiveness: Role of waist-to-hip ratio. *Journal of Personality Social Psychology*, *65*, 293–307.

Singh, D. (1993b). Body shape and women's attractiveness: The critical role of waist-to-hip ratio. *Human Nature*, *4*, 297–321.

Singh, D. (1994a). Ideal female body shape: The role of body weight and waist-to-hip ratio. *International of Journal of Eating Disorders*, *16*, 283–288.

Singh, D (1994b). Body fat distribution and perception of desirable female body shape by young black men and women. *International of Journal of Eating Disorders*, *16*, 289–294.

Singh, D. (1995). Female health, attractiveness and desirability for relationships: Role of breast asymmetry and waist-to-hip ratio. *Ethology and Sociobiology*, *16*, 465–481.

Smith, K. L., Cornelissen, P. L., & Tovée, M. J. (2007). Colour 3D bodies and judgements of human female attractiveness. *Evolution and Human Behavior*, *28*, 48–54.

Streeter, S. A., & McBurney, D. H. (2003). Waist–hip ratio and attractiveness: New evidence and a critique of "a critical test". *Evolution and Human Behavior*, *24*, 88–98.

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

Tovée, M. J., & Cornelissen, P. L. (1999). The mystery of female beauty. *Nature*, *399*, 215–216.

Tovée, M. J., & Cornelissen, P. L. (2001). Female and male perceptions of female physical attractiveness in front-view and profile. *British Journal of Psychology*, *92*, 391–402.

Tovée, M. J., Hancock, P. J. B., Mahmoodi, S., Singleton, B. R. R., & Cornelissen, P. L. (2002). Visual cues to female attractiveness: Waveform analysis of body shape. *Proceedings of the Royal Society, London: Series B*, *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, London: Series B*, *266*, 2111–2118.

Tovée, M. J., Reinhardt, S., Emery, J. L., & Cornelissen, P. L. (1998). Optimal BMI and maximum sexual attractiveness. *The Lancet*, *352*, 548.

Wass, P., Waldenstrom, U., Rossner, S., & Hellberg, D. (1997). An android body fat distribution in females impairs the pregnancy rate of in-vitro fertilization-embryo transfer. *Human Reproduction*, *12*, 2057–2060.

Weeden, J., & Sabini, J. (2005). Physical attractiveness and health in Western societies: A review. *Psychological Bulletin*, *131*, 635–653.

Wittenberger, J. F. (1983). Tactics of mate choice. In P. Bateson (Ed.), *Mate choice* (pp. 435–447). Cambridge, UK: Cambridge University Press.

Zaadstra, B. M., Seidell, J. C., van Noord, P. A. H., Velde, E. R., Habbema, J. D. F., Vrieswijk, B., et al. (1995). Fat and fecundity: Prospective study of effect of body fat distribution on conception rates. *British Medical Journal*, *306*, 484–487.

Manuscript received April 2006
Manuscript revised September 2006

APPENDIX

The terms E_i are referred to as priors, which are assigned before the experiment, associated with event A, is conducted. These independent events cover the sample space. A Venn diagram, designed to aid in interpreting the theorem is presented in Figure 9.

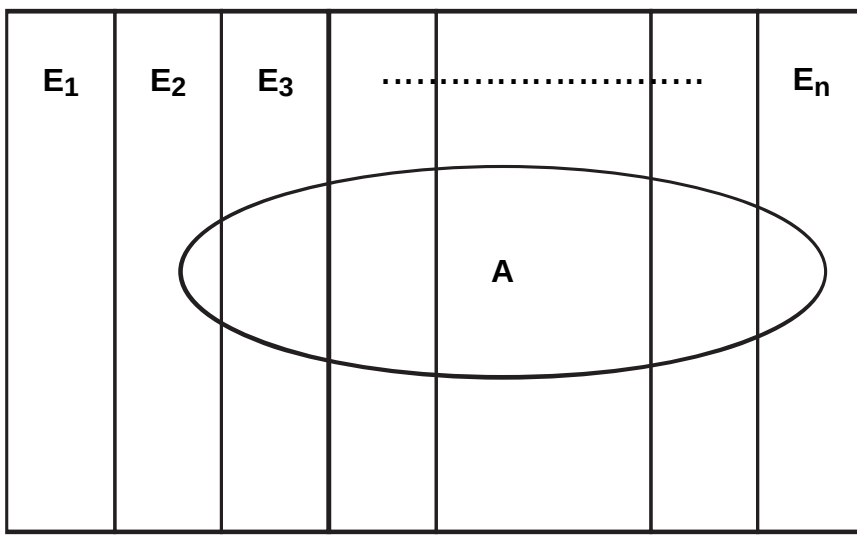

Figure 9. A Venn diagram to illustrate the application of Bayes' theorem.

TABLE 5
Mean and standard deviation for the six priors

i	1	2	3	4	5	6
μ_i	2.8750	2.7750	2.7000	2.7750	2.8250	2.7250
σ_i	1.3990	1.1870	1.7570	1.5930	1.4660	1.5190

Bayes' theorem (Bayes, 1763) says that if there are n disjoint events that cover the sample space, then

$$\text{Prob}(E_i|A) = \frac{\text{Prob}(E_i)\ \text{Prob}(A|E_i)}{\sum_{j=1}^{n} \text{Prob}(E_j)\ \text{Prob}(A|E_j)}$$

where the term in the denominator represents the action of taking the sum of all of the various options. This effectively gives the probability associated with the ellipse, while the numerator gives the probability of a specific prior as it intersects the ellipse.

In the situation considered here the priors are equally likely,

$$\text{Prob}(E_i) = \frac{1}{n}$$

while the individual conditional probabilities are approximated by a normal curve.

$$\text{Prob}(A|E_i) = \phi(A{:}\mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{A-\mu_i}{\sigma_i}\right)^2}.$$

For example, for $n = 6$, given the means and standard deviation shown in Table 5, then for $A = 1$, the probability of the priors are given in Table 6.

TABLE 6
Probability of the priors

i	3	6	4	5	1	2
	0.1856	0.1799	0.1757	0.1637	0.1516	0.1434

Where the outcomes have been arranged in order of likelihood, so Event 3 is the most likely.