Journal of Experimental Psychology:
Human Perception and Performance
2007, Vol. 33, No. 6, 1420–1430

0096-1523/07/$12.00 DOI: 10.1037/0096-1523.33.6.1420

Dissociating Averageness and Attractiveness: Attractive Faces Are Not Always Average

Lisa M. DeBruine, Benedict C. Jones, and Layla Unger
University of Aberdeen

Anthony C. Little
University of Stirling

David R. Feinberg
Harvard University

Although the averageness hypothesis of facial attractiveness proposes that the attractiveness of faces is mostly a consequence of their averageness, 1 study has shown that caricaturing highly attractive faces makes them mathematically less average but more attractive. Here the authors systematically test the averageness hypothesis in 5 experiments using both rating and visual adaptation paradigms. Visual adaptation has previously been shown to increase both preferences for previously viewed face types (i.e., attractiveness) and their perceived normality (i.e., averageness). The authors used a visual adaptation procedure to test whether facial attractiveness is dependent upon faces' proximity to average (averageness hypothesis) or their location relative to average along an attractiveness dimension in face space (contrast hypothesis). While the typical pattern of change due to visual adaptation was found for judgments of normality, judgments of attractiveness resulted in a very different pattern. The results of these 5 experiments conclusively support the proposal that there are specific nonaverage characteristics that are particularly attractive. The authors discuss important implications for the interpretation of studies using a visual adaptation paradigm to investigate attractiveness.

Keywords: attractiveness, averageness, visual adaptation, faces

Supplemental materials: http//dx.doi.org/10.1037/0096-1523.33.6.1420.supp

When Sir Francis Galton (1878) first created composite images of faces by projecting face photographs of many different individuals onto a single piece of photographic film, he noticed that these composite faces tended to be more attractive than the individual face photographs from which they were manufactured. More recently, psychologists have used computer graphic methods to manufacture more realistic composite faces (sometimes referred to as *average faces* or *prototype faces*) that have the average characteristics of their constituent faces. Composite faces manufactured using these computer graphic methods also tend to be perceived as more attractive than the mean attractiveness rating of their constituent images (Langlois & Roggman, 1990; Rhodes, Sumich, & Byatt, 1999; Rhodes et al., 2001; Valentine, Darling, & Donnelly, 2004), leading some researchers to conclude that "attractive faces are only average" (Langlois & Roggman, 1990, p. 115). This is often referred to as the *averageness hypothesis* (Perrett, May, & Yoshikawa, 1994; Valentine et al., 2004).

As the number of images in a composite face increases, so do ratings of attractiveness and averageness (or, equivalently to averageness, ratings of normality or typicality). This is thought to reflect that the resulting faces are more prototypical when more faces are added (Little & Hancock, 2002; Rhodes, 2006). That is, images made up of larger numbers of faces better reflect the average of faces that have been seen. Comparison of new faces with this average may then lead to attraction to faces that are closer to this average (Langlois & Roggman, 1990). Although average faces are more symmetric and have smoother skin than individual faces, these attributes alone do not account for the attractiveness of average faces, as preferences for average shape remain when these factors are controlled (Little & Hancock, 2002; Rhodes et al., 1999; Valentine et al., 2004). The attractiveness of average faces may be a functionless byproduct of the visual system or it may be the result of adaptive preferences. Facial averageness may signal genetic heterozygosity, a predictor of a strong immune system, which from an evolutionary perspective will be attractive in potential mates (Rhodes, Harwood, Yoshikawa, Nishitani, & McLean, 2002; Thornhill & Gangestad, 1993).

As noted, one explanation put forward for the effect of averageness on facial attractiveness is that average faces more closely resemble mental representations of a typical face and therefore are processed more easily by the visual system. Visual adaptation

Lisa M. DeBruine, Benedict C. Jones, and Layla Unger, School of Psychology, University of Aberdeen, Aberdeen, Scotland, United Kingdom; Anthony C. Little, Department of Psychology, University of Stirling, Stirling, Scotland, United Kingdom; David R. Feinberg, Department of Psychology, Harvard University.

Lisa M. DeBruine was supported by National Science Foundation Grant 0401951, and Anthony C. Little was supported by a Royal Society University Research Fellowship. We thank Mike Burt, Lisa Welling, Claire Conway, and the students of the Aberdeen Face Research Lab for helpful comments.

Correspondence concerning this article should be addressed to Lisa M. DeBruine, School of Psychology, University of Aberdeen, Aberdeen, Scotland AB24 2UB, United Kingdom. E-mail: l.debruine@abdn.ac.uk

paradigms (Webster, Kaping, Mizokami, & Duhamel, 2004; Webster & MacLin, 1999) test the effect of visual exposure to faces with a common trait (e.g., increased eye spacing) on perceptions of novel faces and interpret changes in perception as reflecting changes in this mental average. Visual adaptation is thought to alter the mental average in the direction of the adapted faces such that novel, similar faces are perceived as more average after adaptation. Indeed, visual adaptation to faces has been shown to increase both the perceived normality and attractiveness of faces physically similar to those recently seen (Buckingham et al., 2006; Little, DeBruine, & Jones, 2005; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003), supporting the averageness hypothesis.

However, rather than attractiveness being solely a measure of closeness to an internal prototype, one component of attractiveness may be a measure of distance along certain dimensions in face space relative to an internal prototype (Perrett et al., 1994). As face shape deviates from average in one direction along this dimension, attractiveness will increase. However, deviations from average in the opposite direction on this "attractiveness dimension" would decrease attractiveness. Thus, attractiveness would depend on both distance and direction from average. Increasing distance from average in a positive direction along an attractiveness dimension will increase attractiveness, but increasing distance from average in the opposite direction will decrease attractiveness. As attractiveness would then depend on contrast from average such that exaggerated traits in one direction increase attractiveness and exaggerated traits in the opposite direction decrease attractiveness, we term this the *contrast hypothesis*.

Perrett et al. (1994) tested these competing hypotheses in an effort to establish whether averageness is the critical determinant of the attractiveness of faces. Perrett et al. (1994) found that composites of 60 faces were judged as less attractive than composites of the subset of the 15 most attractive faces from the original composite. More important, caricaturing (i.e., exaggerating) the shape differences between the average composites and the attractive composites in the direction of attractiveness produced faces that were even less average but even more attractive than either composite. This finding was interpreted as strong evidence that averageness is not necessarily the critical determinant of facial attractiveness and that highly attractive faces deviate systematically from average. In other words, varying face shape along an attractiveness dimension increases attractiveness even if it simultaneously decreases mathematical averageness.

This interpretation was criticized by Rubenstein, Langlois, and Roggman (2002), who suggested that if highly attractive faces are mathematically closer to average, then composites of 15 highly attractive faces should be mathematically more average than composites of 60 faces unselected for attractiveness. Rubenstein et al. also failed to replicate the results of Perrett et al. (1994) using a different sample of face images. Thus, evidence for and against the contrast hypothesis is equivocal.

Our Experiments

The averageness hypothesis would be supported if the magnitude, but not the direction, of change from average influences attractiveness such that increasing distance from the average face always decreases attractiveness. The contrast hypothesis would be supported if any dimension could be found for which both the direction and magnitude of change from average were important determinants of attractiveness, with increasing distance from the average face increasing attractiveness on one side of the mean and decreasing attractiveness on the other side.

In order to test the averageness hypothesis against the contrast hypothesis, we constructed a continuum of faces (see Figure 1) that vary along a single attractiveness dimension, defined by the linear shape differences between an average of 60 female faces and an average of 15 highly attractive female faces (see Figure 2). This method is the same used by Perrett et al. (1994) to manipulate attractiveness. Previous work on the specificity of adaptation effects has suggested that separable prototypes exist for upright and inverted faces (Rhodes et al., 2004), male and female faces (Little et al., 2005), and White and East Asian faces (Jaquet, Rhodes, & Hayward, 2006). Thus, it is likely that averageness is judged relative to the average of specific categories, so here we use faces of the same age, sex, and race (as do Langlois & Roggman, 1990; Perrett et al., 1994; Rubenstein et al., 2002; Rhodes et al., 2003).

Nonaverage facial characteristics, such as femininity (Perrett et al., 1998) and neoteny (D. Jones, 1995), have been shown to contribute to judgments of attractiveness for female faces. However, studies showing that nonaverage levels of characteristics such as femininity and neoteny are attractive have not compared judgments of attractiveness with judgments of normality (D. Jones, 1995; Perrett et al., 1998). Evidence for attentional biases toward attractive faces (Maner et al., 2003; Shimojo, Simion, Shimojo, & Scheier, 2003) raises the possibility that the mathematical average

Figure 1. A subset of faces used in Experiments 1–4, ranging from –600% to 600% attractiveness dimension manipulation in steps of 200%. The continuum of images used in Experiments 1–4 were 25 images manipulated in 50% steps. A color version of this figure is available on the Web at http://dx.doi.org/10.1037/0096-1523.33.6.1420.supp

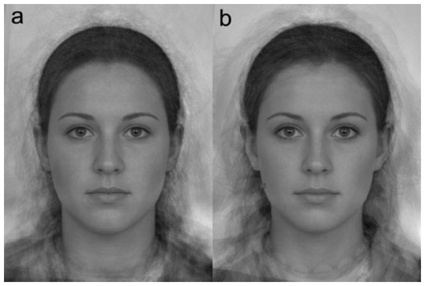

Figure 2. The composite of 60 white female faces (a) and the composite of the 15 faces from this set that were rated as most attractive (b). A color version of this figure is available on the Web at http://dx.doi.org/10.1037/0096-1523.33.6.1420.supp

of faces in a population is significantly different from the mathematical average of the faces seen by most people in that population. For example, Maner et al. (2003) have shown that both men and women selectively attend to attractive female faces, resulting in overestimations of the frequency of attractive faces seen. Thus, an attentional bias toward a certain type of face could result in a biased mental prototype, potentially reconciling preferences for faces that are systematically different from the population average (e.g., Perrett et al., 1994) with preferences for average faces. One prediction from this is that perceived averageness would be biased relative to mathematical average in the same way that attractiveness judgments are.

In Experiment 1, we analyzed normality ratings of these faces to confirm that our stimuli are perceived as less normal the more they deviate from mathematical average. We also compared these normality ratings with attractiveness ratings of the same stimuli in order to determine whether attractiveness and normality judgments show the same pattern (supporting the averageness hypothesis) or whether the image with maximum attractiveness has a more positive value on the continuum than the image with maximum normality (supporting the contrast hypothesis).

In Experiment 2, we analyzed forced-choice measures of preference for pairs of faces that differ equally from the mathematically average face in opposite directions. If the averageness hypothesis is correct, these preferences should be at chance. If the contrast hypothesis is correct, the images with the higher value on the continuum should be preferred.

Overexaggerating any dimension will eventually result in a face that is outside the range of normal human face shape. In Experiment 3, to establish the point at which caricaturing an attractive shape results in a face that is so far from average that concurrent preferences for averageness outweigh preferences for the attractive shape dimension, we analyzed forced-choice measures of preference for pairs of adjacent faces (faces that differ by one step on the attractiveness continuum). If the averageness hypothesis is correct, this tipping point will be at mathematical average. If the contrast hypothesis is correct, the tipping point will be at a positive value along the attractiveness continuum.

In Experiment 4, we analyzed forced-choice measures of preference for all possible pairs of faces on the continuum ranging from –200% to 200%. This was done to control for the possibility that the results of Experiment 2 were solely a result of a skewed (nonsymmetric) distribution of perceived normality. If the averageness hypothesis is correct, the more average face (i.e., the one with the smaller absolute value of manipulation) will always be preferred. If the contrast hypothesis is correct, then the face with the higher value on the continuum will be preferred most of the time.

In Experiment 5, we used a visual adaptation paradigm (e.g., Little et al., 2005; Rhodes et al., 2003; Webster & MacLin, 1999) to test the effects of exposure to attractive or unattractive individual faces on attractiveness and normality judgments of novel attractive and unattractive faces. If the averageness hypothesis is correct, viewing attractive individuals will increase both the attractiveness and normality of novel attractive faces and decrease both the attractiveness and normality of novel unattractive faces, while viewing unattractive individuals will do the opposite. If the contrast hypothesis is correct, normality judgments should follow the same pattern described above, but attractiveness judgments should follow a different pattern, with exposure to attractive individuals decreasing the attractiveness of both novel attractive and unattractive faces and exposure to unattractive individuals increasing their attractiveness.

Experiment 1: Attractiveness and Normality Ratings

Although one can use computer graphic methods to generate a mathematically average face from a population, the face that is perceived as most average may be systematically different from this. Exposure to highly attractive faces in the media or attentional biases to attractive faces (Maner et al., 2003; Shimojo et al., 2003) may result in biased experience such that the average of experienced faces is more attractive than the actual average of faces from that population. This could reconcile the averageness hypothesis with the findings of Perrett et al. (1994) if the face that is perceived as most average is closer to the attractive composite than to the average composite.

To confirm that stimuli on the attractiveness continuum are perceived as less normal the more they deviate from mathematical average and to further validate the use of "normality" judgments for assessing perceived averageness, we had participants rate all faces on the continuum for perceived normality. We also compared these normality ratings with attractiveness ratings of the same stimuli in order to determine whether attractiveness and normality judgments show the same pattern (supporting averageness hypothesis) or whether the image with maximum attractiveness has a more positive value on the continuum than the image with maximum normality (supporting the contrast hypothesis).

Method

Stimulus manufacture. Following Perrett et al. (1994), we defined the attractive shape dimension as the linear two-dimensional shape difference between the average of 60 White female faces (see Figure 2a) and the average of the 15 faces from this set that were rated as most attractive (see Figure 2b). Twenty raters (10 women) judged the full set of 60 faces on a 7-point scale as having a mean attractiveness of 3.36 ($SD = 0.62$) and the most attractive 15 as having a mean attractiveness of 4.23 ($SD = 0.34$). Interrater reliability was high (Cronbach's $\alpha = .86$).

Technical details for the manufacture of composite faces can be found in Rowland and Perrett (1995). Briefly, 189 landmark points were delineated on each of the 60 original faces. The mean *x*- and *y*-coordinates were calculated for all 60 faces and separately for the 15 most attractive to produce average (see Figure 2a) and attractive (see Figure 2b) composites. These composite faces were averaged with their mirror-reversed versions to produce perfectly symmetrical composite faces. Next, the vector differences between each of these paired coordinates were calculated.

For Experiments 1–4, we generated 25 images by manipulating the average female composite along the attractiveness dimension from −600% to 600% in 50% steps (see Figure 1 for a subset of these images).

Participants and procedure. Participants ($n = 232$; 120 women, mean age = 27.1, $SD = 9.5$) were asked to rate all 25 images for normality. They were asked, "In this experiment you will be asked to determine how normal each face looks. Choose your answer by clicking on the labeled buttons above each face." Buttons were labeled *very abnormal* = 0, *somewhat abnormal* = 1, *somewhat normal* = 2, or *very normal* = 3.

A different set of participants ($n = 227$; 130 women, mean age = 25.1, $SD = 7.8$) were asked to rate these same images for attractiveness. They were asked, "In this experiment you will be asked to determine how attractive each face looks. Choose your answer by clicking on the labeled buttons above each face." Buttons were labeled *very unattractive* = 0, *somewhat unattractive* = 1, *somewhat attractive* = 2, or *very attractive* = 3.

Clicking on a phrase started the next trial, and participants could not go back to change their answers. The order of trials was fully randomized. Participants completed each trial at their own pace.

Participants for this and all subsequent experiments were recruited through the Internet. The use of the Internet to investigate face judgments is justified by studies demonstrating that face preference tests produce equivalent results when administered online and in the laboratory (Buchanan, 2000; Feinberg et al., 2005; B. C. Jones et al., 2005; Wilson & Daly, 2003).

Participants in these Web experiments found the Web site through links from other Web sites (no participants were specifically targeted using e-mail or message boards). No compensation was offered for participation. Although specific data on country of residence and ethnic origin were not collected for the participants in our experiments, in general participants at our Web site come from many countries (approximately 30% United States, 20% Great Britain, 12% Germany, 7% France, 6% Canada, 25% other countries) and are mostly of European ethnicity (approximately 78% European, 11% West Asian, 8% East Asian, 3% African). Our Web experiments work on all modern browsers, and registered participants at our Web site mostly use Microsoft Internet Explorer (MSIE) and FireFox browsers (approximately 60% MSIE 6 for Windows, 26% FireFox for Windows, 4% Safari for Macintosh, 2% MSIE 7 for Windows, 2% FireFox for Macintosh, 6% other browsers).

Results

In order to determine the centers of the distributions (i.e., the percentage of image manipulation having the greatest normality or attractiveness), we fit Gaussian distributions to the averages of the normality ratings and attractiveness ratings (see Figure 3). We used the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) to search for the minimum sum of squares (i.e., χ^2, a measure of fit). The distribution for normality judgments was centered on −6.5% ($M = -0.065 \pm 0.065$, $\chi^2 = 0.085$, Equation 1) and the distribution for attractiveness judgments was centered on 99.2% ($M = 0.992 \pm 0.036$, $\chi^2 = 0.314$, Equation 2).

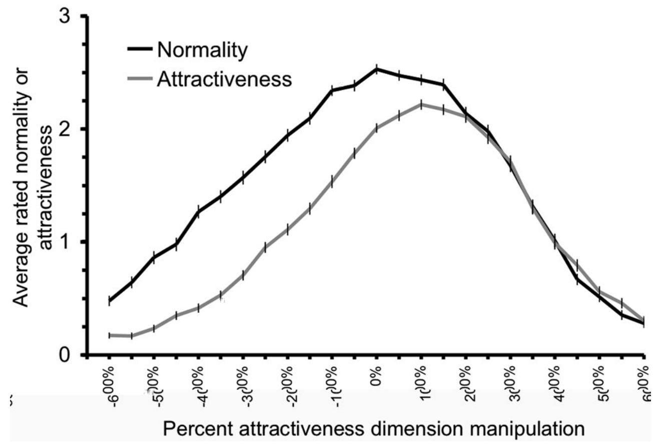

Figure 3. Normality and attractiveness judgments for faces with attractiveness dimension manipulations from −600% to 600%. Ratings were made on the following scale: 0 = *very abnormal* or *very unattractive*, 1 = *somewhat abnormal* or *somewhat unattractive*, 2 = *somewhat normal* or *somewhat attractive*, 3 = *very normal* or *very attractive*. Error bars show standard error of the mean.

Ratings of the images closest to the centers of the normality and attractiveness distributions were compared using a repeated-measures analysis of variance (ANOVA), with image manipulation as the repeated factor (0% or 100%) and rating type as the between-subjects factor (normality or attractiveness). A significant interaction between image manipulation and rating type, $F(1, 457) = 17.8$, $p < .001$, $\eta^2 = .037$, reflected that ratings of normality were higher for the 0% image manipulation than the 100% image manipulation, $t(231) = 1.76$, $p = .080$, $d = .109$, while ratings of attractiveness were lower for the 0% image manipulation than the 100% image manipulation, $t(226) = -4.41$, $p < .001$, $d = .297$.

Discussion

Although some have argued that the average of 15 highly attractive faces is actually a better representation of average than the average of 60 faces unselected for attractiveness (Rubenstein et al., 2002), the results of Experiment 1 confirm that the 0% image on the continuum is perceived to be most average, although not the most attractive. These results support the predictions from the contrast hypothesis rather than those from the averageness hypothesis.

Experiment 2: Attractiveness of Images Equidistant From Average

If the averageness hypothesis is correct, preferences when choosing between pairs of faces that differ equally from the mathematically average face in opposite directions should be at chance. If the contrast hypothesis is correct, the image with the higher value on the continuum should always be preferred.

Method

There were 491 participants (265 women, mean age = 25.8, $SD = 10.5$) who indicated their preference for 12 pairs of images that were equidistant from the mathematical average but opposite in terms of their location along the attractiveness dimension (i.e., 50% vs. –50% to 600% vs. −600%).

Participants were asked, "Please indicate which face you think is more attractive (and how much you prefer it) by clicking on one of the phrases above the face you prefer." They chose the more attractive of the two images and indicated whether this image was *much more attractive*, *more attractive*, *somewhat more attractive*, or *slightly more attractive*, resulting in scores ranging from 0 (the image with the negative value on the attractiveness dimension was judged as *much more attractive*) to 7 (the image with the positive value on the attractiveness dimension was judged as *much more attractive*).

Clicking on a phrase started the next trial, and participants could not go back to change their answers. The order of trials and side of presentation of images in each trial were fully randomized. Participants completed each trial at their own pace. Data were analyzed using one-sample t tests comparing preferences to chance (3.5).

Results

The image with the larger value on the attractiveness dimension was preferred in all cases (all ts > 11.2, $p < .001$, $d = 0.51$–1.28). Note that all of these effects remained significant following correction for multiple comparisons (Holm, 1979). From Figure 4, one can also see that the strength of preference increased as the size of the manipulation increased from 50% to 200% and decreased for manipulations of over 200%, although there was always a significant preference for the face that was higher on the continuum.

Discussion

The results of Experiment 2 show that, for a broad range of values along the attractiveness continuum, the face with the positive value is preferred to the face with an equal absolute difference from mathematical average in the opposite direction (e.g., 200% vs. –200%). Thus, our findings from Experiment 2 again support the predictions from the contrast hypothesis rather than those from the averageness hypothesis.

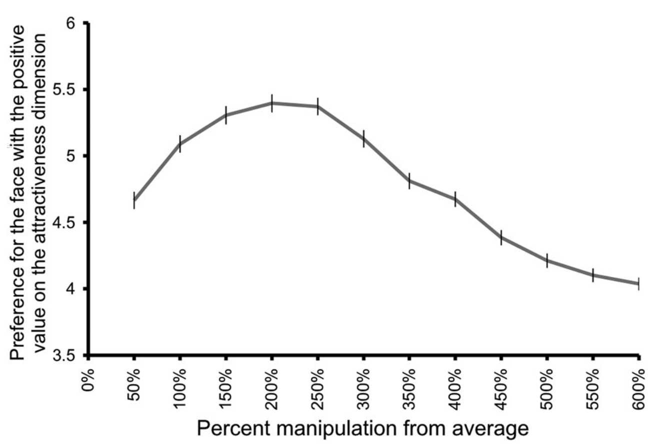

Figure 4. The strength of preferences (on a 0–7 scale) for the image with the larger value on the attractiveness continuum for pairings of equal distance from the average face. The *x*-axis indicates the percentage of manipulation for both images, one positive and one negative (e.g., 200% shows preference for 200% over –200%). Error bars show standard error of the mean.

Experiment 3: Attractiveness of Adjacent Images

Since exaggerating any dimension will eventually result in a face that is outside the range of variability for normal humans, we also analyzed forced-choice measures of preference for pairs of adjacent faces (faces that differ by one step on the attractiveness continuum). At some point, caricaturing an attractive shape will result in a face that is so abnormal that concurrent preferences for averageness will outweigh preferences for the attractive shape dimension. This experiment will determine whether this point is at mathematical average (supporting the averageness hypothesis) or at a positive value along the attractiveness continuum (supporting the contrast hypothesis).

Method

There were 418 participants (241 women, mean age = 27.0, $SD = 11.3$) who chose the more attractive face from all 24 pairs of adjacent images. Each pair of images differed by 50% of the vector differences in shape between the average face and the attractive composite. Otherwise, the experimental procedure was identical to that of Experiment 2. The dependent variable was also the same as for Experiment 2: strength of preference for the image with the larger value on the continuum on a 0–7 scale. Data were analyzed using one-sample t tests comparing preferences to chance (3.5).

Results

Participants preferred the image with the larger value on the attractiveness dimension for all image pairs where the preferred image was manipulated in attractiveness by 150% or less (150% vs. 100%: $t = 1.98$, $p = .048$, $d = .010$; all other ts > 4.95, $p < .001$, $d = .245$–$.528$), even when the preferred image was less average (see Figure 5). Participants preferred the more average image with the smaller value on the attractiveness dimension when the image with the larger value on the continuum was manipulated by 200% or more (all ts < -3.07, $p < .002$, $d = .149$–$.588$). Note that all of these effects remain significant following correction for multiple comparisons (Holm, 1979).

Discussion

The results of Experiment 3 show that indefinitely increasing a face's value on the attractiveness dimension will not indefinitely increase its attractiveness. While this is incompatible with a strict interpretation of the averageness hypothesis (e.g., "attractive faces are only average"; Langlois & Roggman, 1990), these results support the idea that averageness and the attractiveness dimension make independent contributions to attractiveness.

Experiment 4: Attractiveness of All Possible Pairs

The slightly right-skewed shape of the distribution of normality ratings from Experiment 1 could have affected our results for Experiment 2, making comparisons of faces equally different in

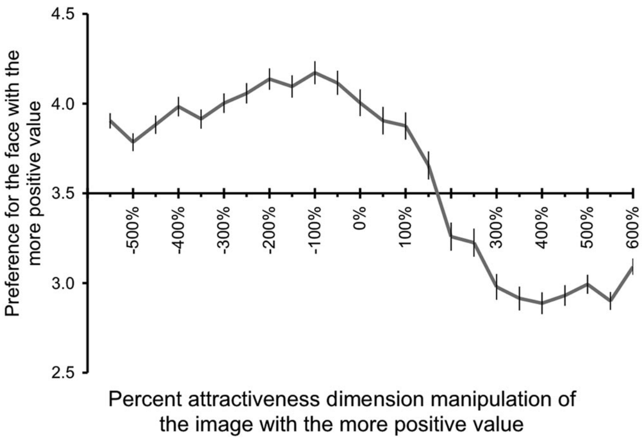

Figure 5. The strength of preferences (on a 0–7 scale) for the image with the larger value on the attractiveness continuum for adjacent pairings. The *x*-axis indicates the level of manipulation for the image with the larger value on the attractiveness continuum, which was compared to the image with a 50% lower level of manipulation (e.g., 200% shows preference for 200% over 150%). Error bars show standard error of the mean.

shape from mathematical average unequally different in perceived distance from an average mental prototype. Therefore, we tested preferences for all unique pairings of images to obtain a detailed perspective on how averageness and value on the attractiveness continuum interact. We limited this experiment to images with manipulations between –200% and 200% in order to limit our conclusions to faces within a plausible range of human variation (following findings from Experiment 1). If the averageness hypothesis is correct, the more average face (i.e., the one with the smaller absolute value of manipulation) will always be preferred. If the contrast hypothesis is correct, then the face with the higher value on the continuum will be preferred most of the time.

Method

There were 502 participants (240 women, mean age = 26.2, $SD = 8.7$) who indicated the more attractive from all possible pairs of the nine images on the continuum with manipulations ranging from –200% to 200%. This resulted in 36 unique pairings. Otherwise, the experimental procedure was identical to that of Experiments 2 and 3. The dependent variable was also the same as for Experiments 2 and 3: strength of preference for the image with the larger value on the continuum on a 0–7 scale. Data were analyzed using one-sample t tests comparing preferences to chance (3.5).

Results

Consistent with Experiment 3, the only pairing for which the face with the smaller value on the continuum was preferred was 150% versus 200%, although this preference was not significantly different from chance, $t(501) = -0.84$, $p = .40$, $d = .035$. For all other pairings, the face with the larger value on the continuum was preferred. This was not significant for 200% versus 100%, $t(501) = 0.12$, $p = .90$, $d = .006$, but was significant for other values (all ts > 2.9, $p < .004$, $d = .183$–2.00; see Figure 6). Note that all significant effects remain significant following correction for multiple comparisons (Holm, 1979).

Discussion

The results of the detailed comparison in Experiment 4 determined that the image with the larger value on the attractiveness continuum was almost always preferred within a plausible range for human face shapes, supporting the predictions from the contrast hypothesis rather than those from the averageness hypothesis. Thus, the face with the larger value on the attractiveness continuum was preferred even when the face with the smaller value was unquestionably perceived as more average.

Experiment 5: Effects of Visual Adaptation on Attractiveness and Normality

Visual adaptation paradigms have been extensively used to investigate how faces are encoded (Jiang, Blanz, & O'Toole, 2006; Leopold, Rhodes, Muller, & Jeffery, 2005; Little et al., 2005; Rhodes et al., 2004; Webster et al., 2004; Webster & MacLin, 1999). Faces are thought to be represented in the brain by neurons that code their values on relevant dimensions in relation to a prototype or average face (Giese & Leopold, 2005; Leopold, O'Toole, Vetter, & Blanz, 2001; Rhodes & Jeffery, 2006; Tsao & Freiwald, 2006). For instance, if face roundness were such a dimension, faces of average roundness would be encoded as having a value on this dimension near zero, while rounder faces would have a positive value and narrower faces would have a negative value. Visual adaptation to faces is thought to alter the prototype, changing the perception of subsequent faces. Thus, after exposure

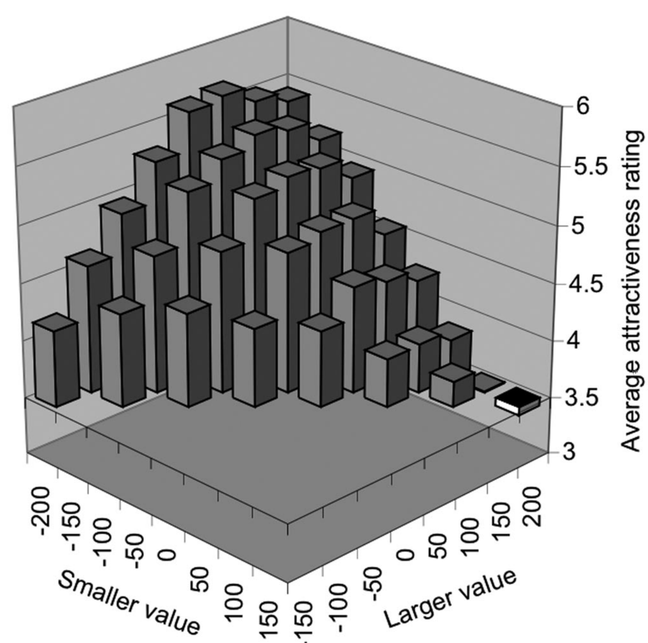

Figure 6. The strength of preferences for the image with the larger value on the attractiveness continuum for all unique pairings between –200% and 200% on the continuum. Values greater than 3.5 indicate that the face with the larger value on the continuum was preferred.

to many round faces, one's mental prototype would have a rounder face and faces that would have been previously encoded with a value on this dimension near zero will now be seen as having a negative value (i.e., being more narrow than average).

Here we used a visual adaptation paradigm (e.g., Little et al., 2005; Rhodes et al., 2003; Webster et al., 2004; Webster & MacLin, 1999) to differentiate between the averageness hypothesis and the contrast hypothesis by testing the effect of viewing particularly attractive or unattractive faces on attractiveness and normality judgments of highly attractive or unattractive composite faces. Exposure to attractive faces will increase the attractiveness of the mental prototype, while exposure to unattractive faces will decrease its attractiveness. Thus, after exposure to attractive faces, the attractive composite will be perceived as closer to average and the unattractive composite will be perceived as farther from average. Conversely, after exposure to unattractive faces, the attractive composite will be perceived as farther from average and the unattractive composite will be perceived as closer to average.

If averageness is the sole determinant of attractiveness, viewing attractive individuals will increase both the attractiveness and normality of novel attractive faces and decrease both the attractiveness and normality of novel unattractive faces (supporting the averageness hypothesis). However, if attractiveness increases as the magnitude of difference from average is increased along an attractiveness dimension within multidimensional face space (Valentine, 1991), normality judgments should follow the same pattern described above, but attractiveness judgments should follow a different pattern, with exposure to attractive individuals decreasing the attractiveness of both novel attractive and unattractive faces and exposure to unattractive individuals increasing their attractiveness (supporting the contrast hypothesis).

To elaborate, the contrast hypothesis predicts that viewing individuals with positive values on the attractiveness continuum will alter the internal prototype so that the contrast between this new prototype and faces with positive values is less than before, and the contrast between this new prototype and faces with negative values is greater than before. This would decrease the perceived attractiveness of faces with both positive and negative values on the attractiveness continuum. Viewing faces with negative values on the attractiveness continuum will increase the contrast between the altered prototype and faces with positive values and decrease the contrast between the altered prototype and faces with negative values, increasing the perceived attractiveness of faces with both positive and negative values on the attractiveness continuum.

Figure 7 illustrates hypothetical effects of adaptation on attributes that are either distributed such that the image with the highest value is near zero and values decrease as images become less average (zero-centered inverted U) or distributed such that the value of that attribute increases as the value on the attractiveness continuum increases (linearly increasing). After adaptation to faces with positive values on the attractiveness continuum, perceptions of any attribute that is distributed in a zero-centered inverted U should shift such that faces with positive values on the attractiveness continuum increase their value on this attribute, while faces with negative values on the attractiveness dimension decrease their value on this attribute. However, adaptation to faces with positive values on the attractiveness continuum should shift perceptions of any attribute that is linearly increasing such that all faces decrease their value on this attribute. Adaptation to faces with negative values on the attractiveness dimension should have opposite effects. The results of Experiment 1 suggest that, at least within a realistic range of faces, normality judgments are distributed in a zero-centered inverted U, while attractiveness judgments are more linearly increasing.

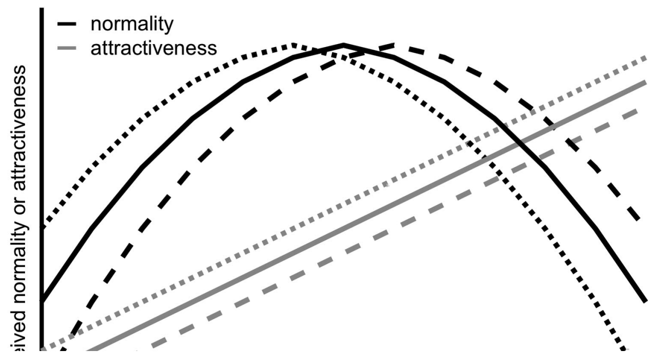

Figure 7. Hypothetical effects of adaptation on perceptions of normality and attractiveness. Solid lines represent perceptions before adaptation, dotted lines represent perceptions after adaptation to faces with negative values on the attractiveness continuum, and dashed lines represent perceptions after adaptation to faces with positive values on the attractiveness continuum.

Method

Stimulus manufacture. Plus and minus 150% of the vector differences between the average and attractive composites (see Figure 2) were applied to 20 individual female faces to produce attractive (see Figure 8, top row) and unattractive (see Figure 8, bottom row) versions. The shape of each face was altered, but the color of the original images was not altered. The 20 attractive versions were then averaged to produce an attractive composite face (see Figure 8, top right) and an unattractive composite face (see Figure 8, bottom right).

This method is computationally identical to that used by Perrett et al. (1994) to produce high 50% composite faces. In Perrett et al. (1994), 50% of the vector differences between the average and attractive composites were applied to the attractive composite, forming the high 50% faces. Here, 150% of the vector differences between the average and attractive composites were applied to 20 individual faces that were then averaged to form the attractive composite. Because Perrett et al. (1994) labeled their faces relative to the attractive composite and we labeled our faces relative to the average composite, our 150% manipulation is equivalent to their 50% manipulation.

Participants. Participants in the two adaptation experiments were 318 people (141 women, mean age = 28.9, SD = 12.4). There were 148 participants who judged the attractiveness of faces (75 in the attractive exposure condition and 73 in the unattractive exposure condition), and 170 who judged the nor-

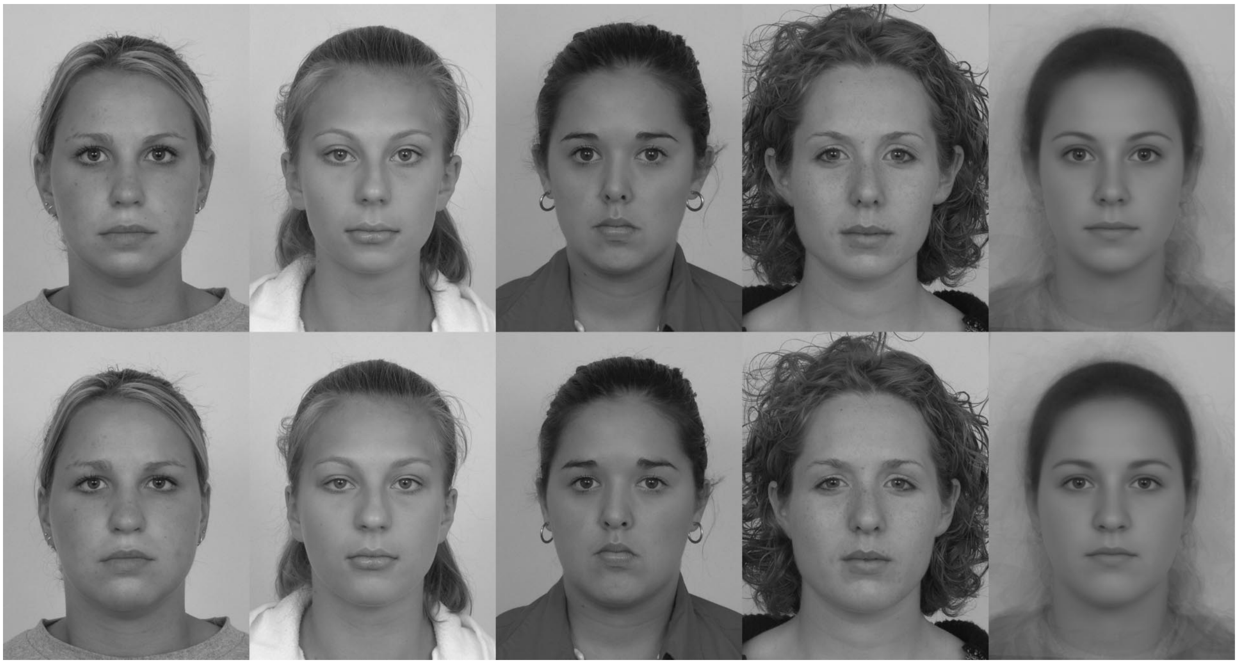

Figure 8. Examples of individual faces increased in attractiveness (top row) and decreased in attractiveness (bottom row), and the composite of 20 individuals increased in attractiveness (top right) and decreased in attractiveness (bottom right). A color version of this figure is available on the Web at http://dx.doi.org/10.1037/0096-1523.33.6.1420.supp

mality of faces (85 each in the attractive and unattractive exposure conditions).

Procedure. In the preadaptation phase, participants were asked to rate the attractiveness or normality of the attractive and unattractive composite faces (see Figure 8, far right images) on a scale from 1 to 7. Specifically, participants in the attractiveness rating task were asked, "Please rate how attractive you think this face is by typing a number 1 (*very unattractive*) to 7 (*very attractive*)," and participants in the normality rating task were asked, "Please rate how normal you think this face is by typing a number 1 (*not normal*) to 7 (*very normal*)." Pressing the enter key after typing in a rating started the next trial and participants could not go back to change their answers. The order of trials was fully randomized. Participants completed each trial at their own pace.

Next, participants completed an adaptation phase, which consisted of passive viewing of 20 unfamiliar female faces that had been manipulated by either 150% (attractive) or –150% (unattractive) as described above (see Figure 8). Each face was presented for 3 s (a total of 1 min).

In the postadaptation phase, participants repeated the first rating task.

Results

Initial analysis. All data were analyzed using a mixed-design ANOVA, with phase (preadaptation, postadaptation) and composite attractiveness (attractive, unattractive) as repeated factors and exposure type (attractive, unattractive) and judgment (attractiveness, normality) as between-subjects factors. A four-way interaction among phase, composite attractiveness, exposure type, and judgment, $F(1, 314) = 3.97$, $p = .047$, $\eta^2 = .012$, confirmed that exposure to attractive and unattractive individual faces did not affect attractiveness judgments and normality judgments of attractive and unattractive faces in the same way. All other significant effects were qualified by this four-way interaction.

In other words, the pattern of results for attractiveness judgments differed from that for normality judgments, falsifying the averageness hypothesis. Next, we carried out separate analyses for attractiveness and normality judgments to determine the nature of these different patterns.

Attractiveness judgments. Attractiveness judgments were analyzed using a mixed-design ANOVA, with phase (preadaptation, postadaptation) and composite attractiveness (attractive, unattractive) as repeated factors and exposure type (attractive, unattractive) as the between-subjects factor. A main effect of phase, $F(1, 146) = 8.01$, $p = .005$, $\eta^2 = .052$, was qualified by an interaction between phase and exposure type, $F\ (1, 146) = 19.5$, $p < .001$, $\eta^2 = .118$ (see Figure 9). A main effect of face attractiveness, $F(1, 146) = 327.7$, $p < .001$, $\eta^2 = .692$, confirmed that the attractive composite was judged as more attractive than the unattractive composite. No other effects were significant (all Fs < 3.1, $p > .08$, $\eta^2 < .021$).

To interpret the interaction between phase and exposure type, we carried out independent-samples t tests on the change in attractiveness judgments from pre- to postadaptation tests (i.e., post-minus preadaptation) comparing participants who were exposed to attractive individuals to those who were exposed to unattractive individuals. This was done separately for judgments of the attractive and unattractive composites. Exposure to unattractive faces

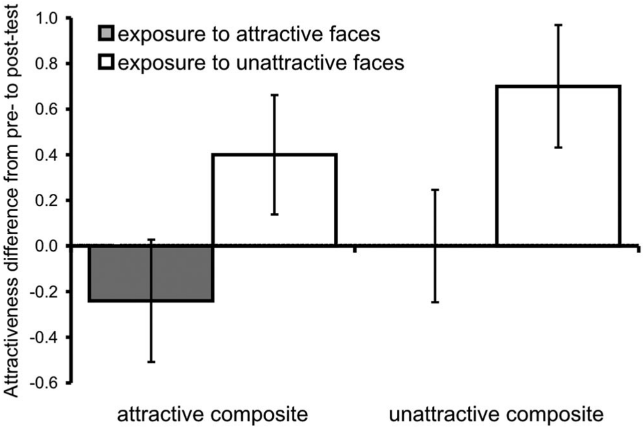

Figure 9. The effects of exposure to attractive and unattractive individuals on attractiveness ratings of attractive and unattractive composite faces. Error bars show standard error of the mean.

increased attractiveness more than exposure to attractive faces for both the attractive composite, $t(146) = 2.79$, $p = .006$, $d = .323$, and the unattractive composite, $t(146) = 3.42$, $p = .001$, $d = .395$.

Normality judgments. Normality judgments were analyzed using the same mixed-design ANOVA, with phase (preadaptation, postadaptation) and composite attractiveness (attractive, unattractive) as repeated factors and exposure type (attractive, unattractive) as the between-subjects factor. The only significant effect was an interaction among phase, composite attractiveness, and exposure type, $F(1, 168) = 6.55$, $p = .011$, $\eta^2 = .038$ (see Figure 10). No other effects were significant (all $F\text{s} < 2.17$, $p > .14$, $\eta^2 < .014$).

To interpret the interaction among phase, composite attractiveness, and exposure type, we carried out independent-samples *t* tests on the change in normality judgments from pre- to postadaptation tests (post- minus preadaptation) comparing participants who were exposed to attractive individuals to those who were exposed to unattractive individuals. This was done separately for judgments of the attractive and unattractive composites. For the attractive composite, the change in normality judgments increased from pre- to postadaptation tests more after exposure to attractive faces than after exposure to unattractive faces, $t(168) = 2.29$, $p = .023$, $d = .119$, while this pattern was in the opposite direction for the unattractive composite, $t(168) = -1.10$, $p = .273$, $d = .249$.

Discussion

While the averageness hypothesis predicted that the same pattern of results would be found for judgments of attractiveness and normality following visual adaptation (Rhodes et al., 2003), here we found different patterns. For judgments of attractive faces, adaptation to attractive or unattractive individual faces had opposite effects on judgments of attractiveness and normality. Although it was hypothesized that exposure to attractive faces would decrease the attractiveness of both attractive and unattractive composites, there was no change in the attractiveness of the unattractive composite, possibly because of a floor effect. Exposure to unattractive faces had a smaller effect on the normality of composites than did exposure to attractive faces. This may be due to attentional biases to attractive faces (Maner et al., 2003; Shimojo et al., 2003).

This pattern conformed to the predictions from the contrast hypothesis. Although Rhodes et al. (2003) have shown that exposure to distorted faces affected both perceived normality and attractiveness in similar ways and Little et al. (2005) and Buckingham et al. (2006) have shown that exposure to systematically manipulated faces increases preferences for novel, similar faces, here we find that exposure to attractive individuals is associated with a decrease in the attractiveness of novel, similar faces. While this finding is inconsistent with the averageness hypothesis, it is a specific prediction of the contrast hypothesis. Increasing perceived averageness increases attractiveness for some dimensions (such as the dimensions altered in previous research) but decreasing averageness by exaggerating certain characteristics in a specific direction increases attractiveness for other dimensions.

General Discussion

The results of the four face rating experiments (Experiments 1–4) and the visual adaptation experiment (Experiment 5) strongly support the contrast hypothesis, demonstrating that there is a dimension in face space that differentiates faces based on attractiveness, independently of the effects of averageness. That is, moving away from average in one direction along this dimension will increase attractiveness, while moving away from average in the opposite direction will decrease attractiveness.

Some have argued that the average of 15 highly attractive faces is actually a mathematically better representation of average than the average of 60 faces unselected for attractiveness (Rubenstein et al., 2002). This implies that the tipping point at which adjacent faces along the continuum are equally attractive is the location of perceived average. The results of Experiment 1 confirm that the 0% image on the continuum is perceived to be most average, although not the most attractive. The results of Experiment 2 show that, for a broad range of values along the attractiveness continuum, the face with the positive value is preferred to the face with an equal absolute difference from mathematical average in the opposite direction.

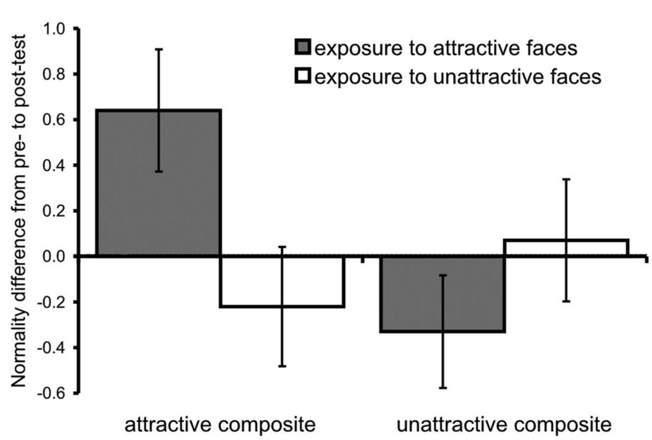

Figure 10. The effects of exposure to attractive and unattractive individuals on normality ratings of attractive and unattractive composite faces. Error bars show standard error of the mean.

Of note, increasing the value on the attractiveness dimension of a face indefinitely will not increase its attractiveness indefinitely. Indeed, the results of Experiment 3 support the idea that averageness and the attractiveness dimension make independent contributions to attractiveness. For manipulations between 0% and 150%, the face with the greater attractiveness dimension manipulation was preferred to the more average face, although for relatively unnatural manipulations (≥200%) the more average face was preferred.

The results of a detailed comparison in Experiment 4 determined that the image with the larger value on the attractiveness continuum was almost always preferred within a plausible range for human face shapes. This confirms the results of Experiments 1–3 and rules out the possibility that the pattern of results in Experiment 2 was simply an artifact of the slightly right-skewed shape of the distribution of normality judgments in Experiment 1.

In Experiment 5, while exposure to faces with a positive value on the attractiveness shape dimension increased the perceived averageness of the highly attractive composite of these faces, this exposure simultaneously decreased the attractiveness of the attractive composite. Thus, the effect of changing face shape along a dimension defined by the shape differences between an average face and an average of highly attractive faces is independent of the effect of similarity to the average of seen faces. This is because faces with negative values on this dimension will become more attractive as they move closer to the population average, and faces with positive values on this dimension will become more attractive as they move away from the population average, conclusively supporting the contrast hypothesis.

Our results complement the findings of Perrett et al. (1994) and are incompatible with the averageness hypothesis (Langlois & Roggman, 1990), which predicts that increasing the averageness of a face will always increase its attractiveness because it proposes that averageness is the critical determinant of attractiveness. Throughout this article, we have been referring to the averageness hypothesis in the most restrictive, original sense of "attractive faces are only average" (Langlois & Roggman, 1990). Our results suggest that while averageness is one component of attractiveness, there is at least one other component of attractiveness that is not explainable by averageness.

In our experiments, we used normality ratings to assess perceived averageness, following work on the effects of visual experience on face perception (Jeffery, Rhodes, & Busey, 2006; Little et al., 2005; Rhodes & Jeffery, 2006; Rhodes et al., 2003, 2004). Other researchers have used various judgments to assess perceived averageness, including ratings on scales anchored by the labels *typical–distinctive*, *typical–atypical*, *usual–unusual*, and *low-high unique* (Wickham, Morris, & Fritz, 2000). However, work on the role of averageness in attractiveness judgments has generally not investigated judgments of perceived averageness, simply assuming that perceived averageness will be equivalent to mathematical average.

Evidence for perceptual biases toward attractive faces (Maner et al., 2003; Shimojo et al., 2003) raises the possibility that perceived averageness is biased relative to mathematical average in the same way that attractiveness judgments are. In Experiment 1, the mathematically average face was perceived as the most normal face, while a different face was perceived as the most attractive. By explicitly testing for this possibility, our findings go beyond those of Perrett et al. (1998) and D. Jones (1995), who showed preferences for nonaverage levels of femininity and neoteny by demonstrating that preferences for a nonaverage characteristic do not reflect preferences for facial cues that are perceived as average.

While our interpretation of the results of Experiment 5 is based in literature on visual adaptation (Little et al., 2005; Rhodes et al., 2003), it is possible to interpret these findings in relation to cognitive comparison or contrast mechanisms (e.g., Cash, Cash, & Butters, 1983). Regardless of the specific mechanisms used, our data have the same important implications for studies of the effects of visual adaptation on normality and attractiveness of faces (e.g., Little et al., 2005; Rhodes et al., 2003). Rhodes et al. (2003) found similar effects of visual adaptation on judgments of normality and attractiveness of compressed and expanded facial features. They interpreted these findings as evidence that judgments of attractiveness are arbitrarily defined by the average of visual experience.

Such shifts would also show that average faces are attractive because

Jeffery, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. *Psychological Science, 178,* 501–505.

Jiang, F., Blanz, V., & O'Toole, A. J. (2006). Probing the visual representation of faces with adaptation: A view from the other side of the mean. *Psychological Science, 178,* 493–500.

Jones, B. C., Perrett, D. I., Little, A. C., Boothroyd, L., Cornwell, R. E., Feinberg, D. R., et al. (2005). Menstrual cycle, pregnancy and oral contraceptive use alter attraction to apparent health in faces. *Proceedings of the Royal Society of London B: Biological Sciences, 272,* 347–354.

Jones, D. (1995). Sexual selection, physical attractiveness, and facial neoteny. *Current Anthropology, 36,* 723–748.

Langlois, J., & Roggman, L. (1990). Attractive faces are only average. *Psychological Science, 1,* 115–121.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4,* 89–94.

Leopold, D. A., Rhodes, G., Muller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society of London B: Biological Sciences, 272,* 897–904.

Levenberg, K. (1944). A method for the solution of certain problems in least square. *Quarterly Journal of Applied Mathematics, 2,* 164–168.

Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face aftereffects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society of London B: Biological Sciences, 272,* 2283–2287.

Little, A. C., & Hancock, P. J. B. (2002). The role of masculinity and distinctiveness on the perception of attractiveness in human male faces. *British Journal of Psychology, 93,* 451–464.

Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., et al. (2003). Sexually selective cognition: Beauty captures the mind of the beholder. *Journal of Personality and Social Psychology, 85,* 1107–1120.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics, 11,* 431–441.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998, August 27). Effects of sexual dimorphism on facial attractiveness. *Nature, 394,* 884–887.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994, March 17). Facial shape and judgments of female attractiveness. *Nature, 368,* 239–242.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57,* 199–226.

Rhodes, G., Harwood, K., Yoshikawa, S., Nishitani, M., & McLean, I. (2002). The attractiveness of average faces: Cross-cultural evidence and possible biological basis. In G. Rhodes & L. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 35–58). Westport, CT: Ablex.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research, 46,* 2977–2987.

Rhodes, G., Jeffery, L., Watson, T., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science, 14,* 558–566.

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology, 14,* 2119–2123.

Rhodes, G., Sumich, A., & Byatt, G. (1999). Are average facial configurations attractive only because of their symmetry? *Psychological Science, 10,* 52–58.

Rhodes, G., Yoshhikawa, S., Clark, A., Lee, K., McKay, R., & Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: In search of biologically based standards of beauty. *Perception, 30,* 611–625.

Rowland, D., & Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications, 15,* 70–76.

Rubenstein, A. J., Langlois, J. H., & Roggman, L. A. (2002). What makes a face attractive and why: The role of averageness in defining facial beauty. In G. Rhodes & L. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (p. 1–34). Westport, CT: Ablex.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6,* 1317–1322.

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature, 4,* 237–269.

Tsao, D. Y., & Freiwald, W. A. (2006). What's so special about the average face? *Trends in Cognitive Sciences, 10,* 391–393.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* 43(A), 161–204.

Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin and Review, 11,* 482–487.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004, April 1). Adaptation to natural facial categories. *Nature, 428,* 557–561.

Webster, M. A., & MacLin, O. H. (1999). Figural after-effects in the perception of faces. *Psychonomic Bulletin and Review, 6,* 647–653.

Wickham, L. H. V., Morris, P. E., & Fritz, C. O. (2000). Facial distinctiveness: Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology, 91,* 99–123.

Wilson, M., & Daly, M. (2003). Do pretty women inspire men to discount the future? *Proceedings of the Royal Society of London B: Biological Sciences, 271,* S177–S179.

Received May 18, 2006
Revision received January 12, 2007
Accepted January 18, 2007 ■