CHAPTER 1

What Makes a Face Attractive and Why: The Role of Averageness in Defining Facial Beauty

Adam J. Rubenstein, Judith H. Langlois, and Lori A. Roggman

Scientists and philosophers have written and conjectured about beauty for centuries. Today one can find over 2,500 different research studies by using the search terms physical attractiveness (which includes the term beauty) and physical appearance in psychological journals alone. Yet, much of this research has been conducted without a conceptual definition of what attractiveness is or why judges seem to agree about it. This chapter will discuss facial attractiveness, our theory of what characterizes facial attractiveness, and why certain faces have more appeal than others, even to young infants. Findings from both adult and infant research suggest that there may be a universal standard of attractiveness, and we will discuss what aspect of faces seems to be universally perceived as attractive. We will also speculate about the mechanisms that could account for the development of attractiveness preferences and for the surprising fact that even young infants seem to recognize attractiveness in faces.

UNIVERSAL PREFERENCES FOR ATTRACTIVENESS?

Before we begin, however, it is necessary to dismiss a myth about facial attractiveness, “Beauty is in the eye of the beholder.” This adage dates back at least to the third century BC, “Beauty is not judged objectively, but according to the beholder’s estimation” (Theocritus, The Idyll). The adage is commonly heard, taught to children by almost all mothers and grandmothers, and is part
of received wisdom. Simply put, the adage states that “different people have different ideas about what is beautiful” (Spears, 1993, p. 45)—based on individual experiences, we each develop idiosyncratic opinions about attractiveness and about who is and isn’t attractive. Thus, using this adage to predict human behavior, even people raised in common environments should encounter unique faces that would affect their concept of attractiveness. Furthermore, if the adage is true, people from different backgrounds and experiences, such as those from different cultures, should develop very different definitions of attractiveness. And, if the adage is true, there is no point to this chapter or even to this volume because there is no common ground for discussion.

Empirical evidence, however, has shown the adage to be false. Using meta-analysis, we examined 130 samples of attractiveness ratings from 94 different studies in the face perception literature (Langlois et al., 2000). This meta-analysis quantitatively assessed the agreement (or disagreement) of thousands of people, both young and old, male and female. The primary studies examined in the meta-analysis largely used Likert-type ratings to assess both the agreement of adults and the agreement of children from within the same culture as well as agreement among adults from different ethnic and cultural groups. Contrary to the adage, the results indicated high agreement about attractiveness from even very different types of raters (see Table 1.1). Effective reliabilities, which assess the reliability of the mean of the judges’ ratings and, thus, can be generalized to random raters (Langlois et al., 2000; Rosenthal, 1991), were analyzed in four separate analyses.1 The average within-culture effective reliability was \( r = .90 \) for ratings of adult faces and \( r = .85 \) for ratings of child faces, both \( p < .05 \). The average cross-ethnic (ratings of different ethnic groups by raters living in the same culture) effective reliability was \( r = .88 \). Cross-cultural (ratings of different ethnic groups by raters from different cultures) effective reliability was even higher, \( r = .94 \). These reliabilities for cross-ethnic and cross-cultural ratings of attractiveness were statistically significant, meaningful, and show consistent agreement among raters regardless of particular experience with different types of faces.

Our meta-analysis also examined variables that might moderate agreement, such as year of publication, sample size, gender of the person being judged, and the situation in which the person was judged (e.g., rating photographs vs. in situ ratings). Somewhat to our surprise, not a single moderator variable had consistent or substantial effects on levels of agreement. While one might assume that methodological differences or group membership could influence attractiveness ratings, it instead appears that a fundamental characteristic of the human face is primarily responsible for agreement regarding facial attractiveness.

These results indicate that beauty is not simply in the eye of the beholder. Rather, raters agree about the attractiveness of both adults and children, even across cultures. Our findings are consistent with those of Feingold (1992), who did a meta-analysis of reliability coefficients from samples of U.S. and Ca-
<table>
<thead>
<tr>
<th>Population</th>
<th>Number of rated samples</th>
<th>Number of attractiveness raters</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-culture ratings of adults</td>
<td>88</td>
<td>1694</td>
<td>.90</td>
</tr>
<tr>
<td>Within-culture ratings of children</td>
<td>28</td>
<td>1182</td>
<td>.85</td>
</tr>
<tr>
<td>Cross-cultural ratings of Adults</td>
<td>17</td>
<td>12146</td>
<td>.94</td>
</tr>
<tr>
<td>Cross-Ethnic ratings of Adults</td>
<td>9</td>
<td>659</td>
<td>.88</td>
</tr>
</tbody>
</table>

Table 1.1 Effective reliability of attractiveness ratings (Langlois et al., 2000).

nadian adults and obtained an average effective reliability of $r = .83$. Together, the meta-analyses of Feingold and Langlois and colleagues (2000) suggest a possibly universal standard by which attractiveness is judged. They also seriously question the common assumption that attractiveness ratings are culturally unique and merely represent media-induced standards.

Agreement regarding attractiveness is not limited to ratings made by adults and school-age children. Research conducted within the past 15 years has demonstrated a previously inconceivable fact: Young infants are aware of attractiveness and exhibit preferences for attractive faces that mirror those of adults (Kramer, Zebrowitz, 1995; Langlois et al, 1987; Langlois, et al 1991; Rubenstein, Kalakanis, & Langlois, 1999; Samuels & Ewy, 1985; Samuels, Butterworth, Roberts, Graupner, & Hole, 1994; Slater, Von der Schulenburg, Brown, & Badenoch, 1998). These studies (which used children ranging from newborns up to 25 months of age) use a method of investigation called the visual preference design in which infants are simultaneously presented with two faces, one attractive and one unattractive. The time the infant spends looking at each face is measured. The assumption underlying the paradigm is that looking longer to one face over another indicates a preference for that face on the part of the infant. The results of these visual preference studies with infants are impressively consistent: Infants as young as a few days of age prefer to look longer at faces adults judge as attractive. These findings are also robust, having been replicated numerous times using different samples of faces (e.g., multiple
samples of female Caucasian faces) and faces from different groups (e.g., female and male Caucasian faces, female African American faces, infant faces).

Do the visual preferences of infants really represent a preference? This is an important question and should not be left to assumption. Therefore, Langlois, Roggman, and Rieser-Danner (1990) asked whether or not the visual preferences displayed by infants extended to other infant behaviors. In the first study, they asked a professional theatrical mask maker to design and construct attractive and unattractive masks for a woman who would later interact with infants as a "stranger." The masks were very lifelike and thin so that they moved with the stranger's face—she could smile, blink her eyes, and so on. The stranger played with 60 one-year-old infants using a strict, rehearsed script so that her behavior would be consistent for all infants. These interactions between the stranger and the babies were coded by observers who could not see the stranger's face and who, therefore, could not be biased by the attractiveness manipulation. In addition, the stranger was not aware of which mask she was wearing—all shiny surfaces were occluded and the masks were identical on the inside—so she could not behave differently based on how she believed she looked. Furthermore, coders evaluated the stranger's behavior in the two conditions and there were no significant differences in her behavior toward the infants. The results showed that infants' preferences for attractiveness extend beyond visual preferences. The infants more frequently avoided the stranger when she was unattractive than when she was attractive and they showed more negative emotion and distress in the unattractive than in the attractive condition. Furthermore, boys (but not girls) approached the female stranger more often in the attractive than in the unattractive condition, perhaps foreshadowing the types of interactions that will occur later when the boys are older!

In a second study, Langlois et al. (1991a) created dolls that were identical except for their faces. One doll had an attractive face while the other doll had an unattractive face. Twelve-month-olds were given dolls to play with and the length of time they played with each doll was recorded. The results revealed that the infants played with the attractive doll for almost twice the amount of time as they played with the unattractive doll. These two studies combined strongly indicate that infants' preferences for attractiveness are not limited to looking behavior. The quantity of visual responses and attention to attractive and unattractive faces is consistent with the frequency and duration of other behaviors toward faces, indicating that longer looking is a valid measure of preference.

Prior to the discovery that even young infants prefer attractive faces, it was widely assumed that preferences for attractive faces were learned. Received wisdom declared that attractiveness preferences were gradually developed through exposure to the preferences of our parents and peers as well as to the concept of beauty idealized by the media. However, studies showing that preferences for attractive faces are evident early in life and studies showing near universal preferences for attractive faces in adults seemingly eliminate the grad-
ual socialization perspective. Both adults and infants appear to possess a common standard by which they judge attractiveness. Either infants are innately attracted to certain faces more than others or infants acquire their attractiveness preferences very early in life as part of basic processes of perceptual development.

The need for identifying how we develop attractiveness preferences becomes obvious when we look at the strong impact it has on our lives. Langlois and colleagues (2000) found in a meta-analysis of the attractiveness literature that attractive people have an advantage over unattractive people because they are viewed as having better social skills (e.g., Kuhlenschmidt & Conger, 1988), more competence at their jobs (e.g., Shapiro, Struening, Shapiro, & Barten, 1976), and as possessing leadership qualities that are lacking in others (e.g., Cherulnik, 1989). These differing opinions of attractive and unattractive people are manifested in actual treatment: attractive people receive higher salaries (e.g., Hamermesh & Biddle, 1994), elicit more cooperation (e.g., West & Brown, 1975), are rewarded more often (e.g., Raza & Carpenter, 1987), and in general have more positive interactions than unattractive people (e.g., Langlois, Ruggman, & Rieser-Danner, 1990). There even seem to be differences in actual behavior in that attractive individuals are better adjusted (e.g., Cash & Smith, 1982), possess higher self-esteem (e.g., O'Grady, 1989), have more dating experience (e.g., Curran & Lippold, 1975), and are more extroverted than unattractive individuals (e.g., Garcia, Stinson, Ickes, Bisonnette, & Briggs, 1991). How do these behavioral differences come about? Are they the results of attractiveness-based differential treatment? By understanding the pathway by which attractiveness preferences are developed, we may better understand how and when attractiveness is used as a cue in social interactions, and, subsequently, better understand how to ameliorate attractiveness biases and stereotypes.

The need to more fully understand the mechanism through which attractiveness preferences form is intensified by research investigating the early development of attractiveness stereotypes. Physical attractiveness may have an effect on the formation of stereotypes during infancy, even before socialization influences can become significant. Recent research has demonstrated that infants as young as 12 months of age are sensitive to the fact that positive valence is associated with attractive people and negative valence is associated with unattractive people (Rubenstein, 1999). Using an intermodal matching procedure, 12-month-olds were simultaneously shown video clips of attractive and unattractive faces while listening to voices speaking in either a pleasant or an unpleasant tone of voice. The infants looked longer at the attractive face when listening to the positive voice and looked longer at the unattractive face when listening to the negative voice. These results indicate that by the end of the first year of life infants associate pleasantness with attractive faces and unpleasantness with unattractive faces. Thus, by 12 months of age, the rudiments of the "beauty is good" stereotype that pervades our society seems to be in place.
WHAT MAKES A FACE ATTRACTIVE?

If standards and preferences for facial attractiveness are innate or acquired very early in life, and if there is considerable agreement about attractiveness both within and across cultures, then what exactly is facial attractiveness and why should it be so preferred by infants, children, and adults alike? The question of what makes a face attractive is an important one. One of the first attempts to define facial attractiveness came from the ancient Greeks who believed that aesthetic preferences could have their origins in mathematics. Specifically, the Greeks focused on the mathematical concept of the "Golden Proportion," a ratio of roughly 1:1.6 that is used to describe the relationship between different parts of physical structures (Huntley, 1970). Natural objects such as the structure of flowers or the spiral in a seashell exemplify the existence of the golden ratio in nature (see Figure 1.1). The ratio has been used by architects to create structures that are pleasing to the eye, such as the Parthenon of ancient Greece. Even Renaissance artists such as Leonardo Da Vinci attempted to put the Golden Proportion to work in painting the human face by using it as a guide for the placement of facial features.

Research conducted in the late 1800s provided a first hint of another way to conceptualize facial attractiveness. Francis Galton (1833, 1888) wondered if certain groups of people had certain facial characteristics. To find the answer, he created photographic composite images of the faces of vegetarians and criminals to see if there was a "typical" facial appearance of the vegetarian or the criminal (although we can speculate about Galton’s reasons for creating the “criminal” face, we must leave to the reader’s imagination his reasons for

Figure 1.1. Golden Proportion in nature.
creating the "vegetarian" face!). Galton overlaid multiple images of faces onto a single photographic plate so that each individual face contributed roughly equally to a final composite face. While the resulting "averaged" faces did little to allow the a priori identification of either criminals or vegetarians, Galton observed that the composite image was more attractive than the component faces. Similar observations were made in 1886 and 1887 by Stoddard, who created composite faces of members of the National Academy of Sciences and graduating seniors of Smith College.

These anecdotal observations of a nineteenth-century English anthropologist and an American psychologist proved insightful. Langlois and Roggman (1990) systematically examined whether mathematical averageness is linked with facial attractiveness. Langlois and Roggman randomly selected photographs of 96 female Caucasian faces and 96 male Caucasian faces that were then divided into three different groups (for generalization purposes) of 32 faces for each gender. Each of 32 photographs in a group were computer scanned and digitized. Although perceived as a perceptual whole, facial images actually are composed of multiple rows and columns of small dots (pixels) that vary in values of brightness and color. When mathematically averaging two images, pixel values are added and divided so that the resulting composite image represents the mathematical average of the individual faces (see Figure 1.2). In this manner, composite images of 2-, 4-, 8-, 16-, and 32 faces were created for each group (see Figure 1.3). These mathematically averaged faces and their component faces were rated for attractiveness by 300 judges on a 5–point

Figure 1.2 Pixel values of faces from Langlois et al. (1994). The box represents the numerical values of the individual pixels on the black-and-white image. The shading of each pixel has a corresponding numerical value, called a gray value. When mathematically averaging faces, overlapping pixels from the two images are averaged to produce a mathematically average image.
Figure 1.3. Progression of mathematically averaged faces from 2 faces to 32 faces.
Likert scale (1 = very unattractive, 5 = very attractive). The findings were statistically significant and robust. For each of the groups of male and female faces, the 16-and 32-face averaged faces were rated as higher in attractiveness than the mean attractiveness ratings of the component faces and as more attractive than the composites consisting of fewer faces. These results illustrate that a face representing the central tendency or average of a population is attractive.

Current research into the attractiveness of averaged faces has moved from a computer overlaying procedure in which pixel values are averaged to a computer morphing procedure. Keypoint markers are placed on an image of the face to designate the location, size, and configuration of the facial features (e.g., Benson & Perrett, 1993; see Figure 1.4). Facial images are then combined based upon the average of the keypoint locations to create averaged faces. The keypoint morphing technique produces averaged faces that lack the "ghosting" due to blur or soft focus often associated with the original averaging studies of Langlois and Roggman (1990). While the morphing procedure is still a form of mathematical averaging, it has the benefit over the pixel averaging procedure that it allows for more accurate alignment of facial features in order to create higher quality images. Using this methodology, Rhodes and Tremewan (1996) digitized photographs of faces and extracted line drawings from the images. The line drawings were mathematically averaged together. The resulting averaged line drawings were rated as higher in attractiveness than were the individual nonaveraged faces. Rubenstein, Langlois, and Kalakanis (1999) went beyond line drawings of faces and applied the morphing technique to real faces. Like the Langlois and Roggman study, a

Figure 1.4. Example of keypoint placement on the human face.
sample of 32 female Caucasian faces was randomly selected from a large database of faces. These faces were then averaged using the keypoint technique to create 2-, 4-, 8-, 16-, and 32-face averages that were rated for attractiveness along with their component faces. The results indicated a strong pattern of preference for the averaged faces that mirrored the findings of the earlier Langlois and Roggman study.

Up to this point, the research described has focused on the attractiveness of Caucasian faces. It is clear, however, that preferences for mathematically averaged faces go beyond Caucasian faces to faces from other ethnic groups. Pollard (1995) generalized the finding that averaged faces are attractive from primarily Caucasian to a cross-cultural sample of judges. Pollard created mathematically averaged faces of male Caucasian and female Caucasian faces by creating photographic overlays using timed exposures. These faces were then rated for attractiveness by judges from four different countries: New Zealand, India, China, and Nigeria. Despite the differences in cultural backgrounds, high cross-cultural agreement was found for the attractiveness ratings of the faces. However, the use of photographic overlays is not the best methodology to create mathematically averaged faces because of the inherent imprecision involved in the timing of photographic exposure. In contrast, Rhodes, Harwood, Yoshikawa, and Nishitani (Chapter 2, this volume) used the morphing procedure described above and also found cross-cultural preferences for mathematically averaged faces. Rubenstein, Langlois, Kalakanis, Larson, and Hallam (1997) extended the finding that averaged faces are attractive faces to a sample of female Asian faces. Using a morphing procedure, samples of 32 female Caucasian faces and 32 female Asian faces were separately combined to form a mathematically averaged face for each ethnic group. The averaged faces and their component faces were evaluated by Caucasian and Asian judges. High cross-cultural agreement was found in the attractiveness ratings from both ethnic groups, indicating that the Caucasian and Asian judges considered the same faces to be attractive, regardless of the ethnicity of the faces being judged. Both the averaged female Caucasian faces and averaged female Asian faces were rated as attractive, demonstrating that the average of a population is attractive regardless of the ethnicity of the faces from which they are created and regardless of the ethnicity of the raters. Thus, adults from different cultural backgrounds consider averaged faces to be attractive.

The finding that averaged faces are attractive and preferred is not limited to just preferences of adults. Even 6-month-old infants consider averaged faces to be attractive (Rubenstein et al., 1999). In a study parallel to Langlois and colleagues (1987, 1990), infants were shown pairs of a female Caucasian 32–face averaged composite face and a female Caucasian unattractive individual face. The attractiveness of the faces was determined by adult judges. The infants’ looking time to each face was measured and the results showed that infants look significantly longer at averaged faces compared to unattractive faces.
Thus young infants demonstrate a preference for averaged faces just as adults do.

It is important to be clear about what we mean by an “average” face. A mathematically averaged face is not an average or common face. It is not average in facial attractiveness. The term “average” (more accurately referred to as “averaged” face) refers only to the physical configurations of faces created by averaging multiple individual faces together mathematically. These mathematically averaged faces have features and a configuration representative of the population average. A mathematically averaged face is not average in perceived attractiveness or other psychological dimensions associated with faces. Rather, these averaged faces and any individual faces that are close in configuration to the mathematical average are extreme, not average, in attractiveness.

CRITICISMS OF THE AVERAGE HYPOTHESIS

The explanation of facial attractiveness as the average of the population is not without criticism. Pittenger (1991) argued that averaging pixel values of a matrix of the whole face as we did rather than averaging spatial locations of anatomically defined features does not preserve the shapes and locations of anatomical features. Langlois, Roggman, Musselman, and Acton (1991b) showed that averaging matrices of gray values in effect creates a frequency map that is similar to a contour map in which contrast information is preserved. Such a map clearly does preserve the shape and location of anatomical features, as much work has previously shown (Ballard & Brown, 1982; Marr, 1977; Ramachandran, 1988). Furthermore, the keypoint average method, discussed above, does average anatomical features, and produces the exact same results as the pixel matrix averaging procedure with respect to judgments of attractiveness.

Pittenger’s (1991) second complaint was that, by averaging two-dimensional representations of faces normally perceived in three dimensions, we distorted distances along lines not perpendicular to the viewing direction. In a mathematical proof, Langlois and colleagues (1991b) showed that this concern was trivial and inconsequential, as has also been proven in other research (Acton & Bovick, 1990; Barrow & Tennebaum, 1981; Marr, 1982).

Pittenger’s (1991) final criticism was that he believed that a functional “optimal value” was preferable to using a measure of central tendency to understand what explains facial attractiveness. We replied then (Langlois et al., 1991b) and still believe today that such an argument about “optimal” values is circular and lacks both parsimony and scientific rigor: How is “optimal” to be defined and operationalized? Optimal with respect to what? Very widely spaced eyes might provide superior binocular vision and protruding ears might provide superior sound localization but neither are considered to be very attractive on a face. Likewise, by what set of weights are we to judge whether
"optimal" eye width is more or less important than "optimal" ear protrusion? Until the proponents of an "optimal" definition of attractiveness specify their construct, it cannot be falsified and we, therefore, do not see its utility.

Artifacts of Averaging

Other early criticisms of our theory that "averageness" produces attractive faces claimed that artifacts of the averaging process were better explanations for the attractiveness of mathematically averaged faces than was "averageness." For example, Alley and Cunningham (1991) suggested several artifacts of the averaging process that could result in increased attractiveness of averaged faces: blurring and smoothing that would remove blemishes, softening of focus that would increase the appearance of youthfulness, and an increase in symmetry that would remove unattractive asymmetries. Both logic and empirical testing show that these artifacts do not explain the attractiveness of averaged faces.

Blurring and Smoothing

Blurring and smoothing of faces result when faces are averaged together and, therefore, composite faces can have a "soft focus," as well as smooth, uniform skin tones, free of blemishes. In our facial averaging studies, however, individual faces are always "smoothed and blurred" to match any smoothing and blurring that occurs in averaged faces as a result of the computer manipulation (Langlois & Roggman, 1990; Langlois et al., 1994; Rubenstein et al., 1999). Therefore, any differences documented between averaged and individual faces cannot be due to smoothing or blurring because the faces are equivalent on these dimensions. Second, we (Langlois et al., 1994) further examined artifacts of the averaging process by photographing the same individual multiple times and creating an averaged face from these different images of the same person. Thus, individual faces were image-processed in exactly the same manner as a mathematically averaged face consisting of multiple different faces. If "smoothing and blurring" and other artifacts of computer averaging procedures are responsible for attractiveness, rather than averaging together different faces, the face resulting from averaging together multiple images of the same person should also be attractive. However, this is not what we found. Averaging together multiple images of the same person did not produce an attractive face, whereas averaging together multiple faces does (Langlois et al., 1994). Thus, the smoothing and blurring aspects of the averaging process cannot explain the attractiveness of the face. An averaged facial configuration is necessary and sufficient to create an attractive face.
Youthfulness

Youthfulness, although certainly related to perceived attractiveness, also is not adequate to explain the findings of Langlois and Roggman (1990). It is clear that, in general, young people are rated as more attractive than old people (e.g., Alley, 1988; Henss, 1991; Mathes, Brennan, Haugen, & Rice, 1985; Zebrowitz, Olson, & Hoffman, 1993). However, all the faces in the Langlois and Roggman sample were young college students. Thus, for youthfulness to account for the results of averaging, youthfulness would have to be correlated with attractiveness within this young sample. Langlois and colleagues (1994), therefore, had judges rate the perceived age of averaged faces as well as the age of the individual faces that comprised the averaged faces. There was no significant relation between youthfulness and attractiveness for either male or female faces ($r = .04$, n.s.; $r = .13$, n.s., respectively), demonstrating that the perceived youthfulness of the faces was not a criterion in determining the attractiveness of this particular sample of faces. Thus, although we certainly agree that young faces are generally rated as more attractive than old faces (e.g., Zebrowitz et al., 1993), youthfulness per se does not account for the attractiveness of averaged faces when they are derived from young faces. To reiterate, although we agree that youthfulness is a component of attractiveness when different age cohorts of faces are compared with each other, we disagree that youthfulness accounts for the attractiveness of averaged faces within age cohorts.

Symmetry

A number of critics have argued that symmetry, not averageness, accounts for the attractiveness of averaged faces. Alley and Cunningham (1991) claimed (but provided no empirical data in support of their claim) that the symmetry created by averaging faces together is what accounts for their attractiveness. Of course, symmetry in a face is important—no reasonable person would deny that severely distorted, asymmetrical faces are not attractive. Many craniofacial deformities involve severe distortions of symmetry in faces. It is also true that mathematically averaged faces are high in symmetry, indicating a relationship of some sort between averageness and symmetry. However, the real issue is as follows: Can symmetry explain the attractiveness of mathematically averaged faces? And, is symmetry sufficient to produce attractiveness in faces? Like youthfulness, to conclude that symmetry is the defining factor of attractiveness in normal faces, there must be a strong relationship between symmetry and attractiveness in normal faces. Furthermore, if symmetry is defining, perfectly symmetrical faces should be more attractive than faces that are not perfectly symmetrical.

To test whether or not symmetry defines attractiveness, Langlois and colleagues (1994) evaluated the attractiveness of symmetrical faces in two different ways. First, they evaluated the attractiveness of perfectly symmetrical, mirror-image faces (chimeras). Chimeras were created on a computer by dividing a face down a central, vertical line and replacing one side with a mirror im-
age of the other (i.e., replacing the left side of the face with a mirror image of the right side). Both “left” and “right” chimerical faces were created. The chimerical faces and the original faces from which they were created were rated for physical attractiveness on a 5-point Likert scale. The results showed that the original (asymmetrical) faces were rated as significantly higher in attractiveness than the perfectly symmetrical chimerical faces. Inspection of the faces confirmed the validity of the statistics: Some perfectly symmetrical faces can be quite unattractive and some asymmetrical (within the normal range) can be quite attractive. (See Figure 1.5.)

Second, because the chimeric method of creating symmetrical faces can introduce structural abnormalities that create faces abnormal in appearance (Rhodes, Roberts, & Simmons, 1999), Langlois and colleagues (1994) created symmetric faces by averaging together a normal face and its mirror image. If the greater attractiveness of averaged faces is due simply to the symmetry produced by the averaging procedure, the mirror-image averaging technique of producing symmetrical faces should also produce highly attractive faces. However, this not what the results showed. Rather, composite faces made by averaging together many faces were judged to be significantly more attractive than composite symmetrical faces made by averaging together the same individual and its mirror image. This evidence refutes the claim that averaged faces are attractive because they are symmetrical rather than because they represent the central tendency of the population.

Grammer and Thornhill (1994) investigated both symmetry and averageness as potential explanations of attractive faces, and, like Alley and Cunningham (1991), claimed that attractiveness is defined by symmetry, not averageness. They created four composite faces made from four faces each, two composite faces made from eight faces each, and one composite face made

Figure 1.5. Example of individual face (left) and the corresponding symmetrical chimera (right). Used with permission.
from 16 faces for both male and female faces. After correlating level of averageness (4-, 8-, or 16-faces) with attractiveness ratings, Grammer and Thornhill concluded that averageness was not associated with high attractiveness. In contrast, after removing the impact of averageness on the face, facial symmetry was related to perceived facial attractiveness. Thus, Grammer and Thornhill concluded that symmetry was more important than averageness in facial attractiveness. However, by creating primarily low-level averaged faces (four 4-face composites, two 8-face composites, and only one 16-face composite), Grammer and Thornhill analyzed the effect of averageness in a sample of averaged faces that did not approach the population average. No statistician would agree that sampling four or eight items from a large population is sufficient to approach the population mean and any impact of an averageness effect from the 16-face composite (the only face that would begin to approach a representation of the population average) would have been negated by analyzing the averaged faces as a group. Langlois and Roggman (1990) and Langlois and colleagues (1994) claim only that faces close in configuration to the mathematical average of a population are attractive, not that averaging together any small sample of faces will produce either attractiveness or averageness. Indeed, Langlois and Roggman found that attractiveness was highest for 32-face composites, whereas the lower-level composites were not significantly different in attractiveness from individual faces. By using faces that were primarily low-level composites (4- and 8-faces), Grammer and Thornhill failed to test the averageness hypothesis as proposed by Langlois and Roggman.

Rhodes, Sumich, and Byatt (1999) directly manipulated both the symmetry and averageness of faces, creating images of high and low mathematical averageness by warping individual faces so that their configurations were made closer to that of a mathematically average face or further from that of a mathematically average face. They also created perfectly symmetrical versions of the images by morphing each face with its mirror image. Their findings indicated that symmetry impacted attractiveness independently of averageness. However, the impact of averageness on facial attractiveness appeared to be more robust than that of symmetry. Furthermore, recent research (Rhodes et al., 1999) has demonstrated that two-face morphs rated high in averageness were rated as being more attractive than symmetrical images of faces, lending strength to the claim that averageness has a robust impact on the perception of facial attractiveness.

Finally, the importance of symmetry is also called into question by infant research. Using a visual preference paradigm, Samuels and colleagues (1994) presented infants with normal faces and with symmetrical chimerical faces. Infants ranging in age from 4 to 25 months of age preferred adult-rated attractive faces over merely symmetrical faces. If symmetry defines attractiveness, the infants should have preferred the symmetrical faces. However, given the questionable findings using chimerical faces in adult studies, research utilizing
mirror-image averaging will be necessary to confirm that infants prefer attractive faces over symmetrical faces.

Thus, although it appears that symmetry is related to the attractiveness of a human face, it is not essential. Although a mathematically averaged face will be symmetrical, a symmetrical face is not necessarily highly attractive or close to the mathematical average of a population of faces. Furthermore, a highly attractive face is not necessarily highly symmetrical. This conclusion is consistent with the oft-heard comment of movie stars that insist on being photographed only from their “good side.”

Facial Extremes

Some critics of the averaging hypothesis claim that extreme facial features rather than averageness account for attractiveness. The exaggerated feature argument rests in evolutionary theory and observation in which certain characteristics of males, such as large antlers or bright colors, are “attractive” to females of the species (almost always, nonhuman species). Presumably, as a result of female preferences, features like large antlers are subject to positive selection pressures. Interestingly, however, there are many differences and contradictions between the human and nonhuman literatures. In nonHumans, it is the male rather than the female who possesses these “extreme” features. In humans, in contrast, evolutionary theory posits that attractiveness and “extremes” are more important for females than males (Buss & Barnes, 1986; Buss & Schmitt, 1993; Jackson, 1992).

Furthermore, the logic of the exaggerated feature position is suspect because the “extreme/exaggerated” argument suffers from the same problem of circularity as the “optimal” argument discussed earlier. The exaggerated features theory does not specify how large the features must be or when a feature changes from being “exaggerated” to “abnormal” in human faces. Furthermore, even within the normal range of feature sizes, eyes that may be large for one face may appear as “small” for another face (see Figure 1.6). What is crucial is not the size of any individual feature but rather the entire configuration of the face and all its constituent parts. For human faces, exaggerated or large features in and of themselves cannot be a parsimonious explanation of attractiveness because one can easily imagine a face with huge eyes or a huge nose or a huge mouth that would not be attractive. Thus, the generalization from nonhumans to humans of the exaggerated feature hypothesis is dubious, at least with respect to the face. Furthermore, as we will see in the next section, the averageness hypothesis is not inconsistent with evolutionary theory.

Nonetheless, using evolutionary theory as the underlying theoretical rationale, Perrett, May, and Yoshikawa (1994) created a mathematically averaged face from a group of 16 highly attractive faces taken from a population of 60 faces. Perrett and his colleagues then exaggerated the facial features of the 16-face composite by calculating differences between it and a mathematically
Figure 1.6. Do extreme features produce attractive faces? Faces A and B have the exact same internal features but placed within a different head size. Most raters judge Face A as having eyes too large (bulging), a nose too large, and too large a mouth. They judge the entire face as unattractive, demonstrating that extreme features such as large eyes are not necessarily attractive. Most raters judge Face B, with identical features as Face A, to have eyes that are about average in size, and both a small nose and mouth. They judge the configuration of Face B to be more attractive (although not highly attractive) than Face A despite having identically sized internal features. Thus, it is the configuration of the entire face not the size of individual features that determines facial attractiveness.

An averaged face created from all 60 faces. Keypoints on the “high attractive” averaged face were selected and compared to the corresponding keypoints on a mathematically averaged face representative of the population. From this comparison, the differences between the two faces were then exaggerated by 50%, creating a “caricature” face. Thus, if the eyes of the “high attractive” composite face were normally larger than the population average, the size of the eyes would be further exaggerated through the caricaturing process. Perrett and colleagues reported that Caucasian and Asian female faces with extreme facial features were more attractive than composite faces created by averaging together only very attractive female faces or composite faces comprised of female faces of all attractiveness levels (this last type being analogous to the Rubenstein et al. [1999], methodology). Their conclusions, however, were more likely due to artifacts of their methodology than evidence against the averaging hypothesis.

First, note that averaged faces were exaggerated to create the caricature faces, making averageness the foundation of the caricature faces. Thus, logically, creating faces that are deviations from this base or creating averaged faces from only very attractive faces cannot disconfirm our theory. Because more attractive faces are closer to the mathematical average of the population, creating averaged faces
from only attractive faces merely yields a face even closer to the average because exclusion of unattractive faces reduces variation in the population. Second, note that Perrett’s claims were overgeneralized: their results for female faces were not replicated by their own results for male faces. Finally, Perrett and colleagues (1994) used a forced-choice task that emphasized very small differences in facial appearance because judges were not allowed to rate faces as equally attractive. Research in the judgment and decision-making literature shows that forced-choice measures (such as that used by Perrett) and Likert-scale judgment measures (such as that used by Langlois) can, and quite frequently do, yield different results dependent on the judgment situation (see Payne, Bettmann, & Johnson, 1993, for a review). The debate concerning the more appropriate measure of true preferences is long-running (Guilford, 1954; Torgerson, 1958). Although far from resolved, some evidence indicates that forced-choice measures are more appropriate for judgments of highly specific characteristics based on highly specific dimensions such as which pair of eyes are further apart. In contrast, Likert-scale judgments are more appropriate when holistic judgments, such as attractiveness, are made (e.g., Acredolo, O’Connor, Banks, & Horobin, 1989). Because most face perception research indicates that we process faces as wholes instead of processing individual features (e.g., Ellis, Burton, Young, & Flude, 1997; Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998), a Likert scale may be more appropriate for judging attractiveness.

Nevertheless, to evaluate empirically Perrett and colleagues’ (1994) claims, we created computer-generated faces in an identical fashion. Thus, we created the following types of Caucasian (male and female) and Asian (female) faces: Individual nonaveraged faces, population averaged (made from randomly selected faces of all levels of attractiveness, as per Rubenstein et al. [1999]), high averaged (made from only very attractive faces, as per Perrett et al.), and caricature (made by exaggerating facial features, as per Perrett et al.)

Unlike Perrett, we used several different methods of evaluating attractiveness. First, a group of judges (N = 135) rated the faces for attractiveness using a nonforced-choice task in which one choice allowed judges to indicate that the two faces were equally attractive. Another group of judges (N = 135) rated the faces on a Likert scale (1 = very unattractive, 5 = very attractive), the standard method of assessing attractiveness. A third group of judges (N = 300) indicated whether or not they believed the images depicted two different people or different photographs of the same person (see Figure 1.7).

Analyses of the Likert ratings showed that there were no significant differences in attractiveness among the caricatures, high averaged faces, and population averaged faces (F(2, 264) = 2.72, n.s.). Furthermore, a priori power analyses ensured that these null results were not due to low statistical power (recall also that a large number of raters participated so that even small differences could have proved significant). Planned comparisons of the caricature faces with the high averaged and population averaged faces revealed no significant differences in the perceived attractiveness of the faces. When judges were
Figure 1.7. Comparison of mathematically average face (left) and caricature face.

asked to evaluate the faces with the extreme features (caricature faces) versus the averaged faces made from only very attractive faces (high averaged), they did not prefer either to the other. Instead they selected the “equally attractive” option more often than chance ($X^2(1) = 90.32, p < .001$). Indeed, judges believed these faces were merely different pictures of the same individual ($X^2(1) = 180.62, p < .001$). When judges were asked to evaluate the faces made from only very attractive faces (high averaged) versus faces made from randomly selected faces of all levels of attractiveness (population averaged), they did not prefer either to the other. Instead, they again selected the “equally attractive” option more often than chance ($X^2(1) = 14.03, p < .001$), and were more likely to believe that these faces were pictures of the same individual, although this pattern was not significant.

These results do not support Perrett and colleagues’ (1994) findings and instead argue against the “extreme” view of attractiveness. Furthermore, visual inspection of the faces in Figure 1.6 shows that the differences between our averaged faces and Perrett and colleagues’ caricature/extreme faces are trivial: The faces look much more alike than different. This should not be surprising given that the Perrett faces with “exaggerated” features used averaged faces as the foundation for creating exaggerated faces. Thus, although it may be possible to modestly enhance the attractiveness of a mathematically average face, Perrett’s results support rather than refute our theory. This same logic extends to other work in which averaged faces are “feminized,” such as Perrett and colleagues (1998) and Rhodes, Hickford, and Jeffery (2000).

Differences between Physical and Psychological Representations

In the most recent published criticism of averaging, Busey (1998) reported on differences between actual and predicted locations of morphed (averaged) faces in psychological face space. He reasoned that, because the morph of the
images of two parent faces is represented by the average physical locations of the two parents (in terms of pixel locations), the morph should also be represented by the centroid of the psychological dimensions (e.g., perceived youthfulness, pudginess, etc.) associated with the parent faces as extracted from similarity ratings. Thus, a morph should represent 50% of each parent’s physical attributes and judges should rate morphs as equally similar to each parent. Instead, Busey found deviations between the actual and “predicted” locations. That is, raters did not judge the morphs to be exactly in between the parent faces on the psychological dimensions. His report concluded that these deviations posed significant problems for research using image-processing techniques to derive a face that represents the centroid of a population.

Busey (1998) had raters judge the similarity between pairs of faces, including some morphed faces, and used multidimensional scaling (MDS) procedures to determine psychological dimensions of the morphs, their parent faces, and random pairs of faces. He converted these similarity ratings into a Euclidean space, where the distance between points represented the degree of psychological disparity between faces. He then interpreted the resulting clusters of points by inferring six “psychological” dimensions, such as age and adiposity, on which raters seemed to have differentiated the faces. This multidimensional space, based on similarity ratings, converted to Euclidian space, is what Busey termed “psychological face space.”

We note several problems with these procedures and conclusions. First, the major thesis that physical and psychological space should be equivalent is inconsistent with our theory. Following such an argument, for example, would lead to the prediction that morphs should be judged to be in between their two parents in attractiveness. However, it is clear that when many faces are averaged or morphed together, the composite is more attractive than the individual “parents.” This “deviance” or discrepancy between mathematical averaging and psychological dimensions is in fact the very basis for our hypothesis that Euclidean averageness is a major component in psychological attractiveness. Thus, Busey’s (1998) findings confirm, not undermine, what we reported in 1990 and 1994.

Second, by making claims such as “age appears to be the primary dimension along which people organize faces,” Busey (1998) implies that his face space can be generalized to all faces and all types of judgments about faces. The article argues that this particular psychological face space can be used to evaluate “experiments in which the physical prototype created by combining faces is assumed to correspond to a psychological prototype created by unknown psychological combination mechanisms” (p. 476). Yet, Busey used only bald men as stimuli. What if women’s faces had been included? Wouldn’t gender emerge as a primary dimension? Indeed, in other MDS studies of psychological space, different sets of dimensions emerge showing that MDS solutions are highly dependent on input (Hirschberg, Jones, & Haggerty, 1978; Pedelty, Levine, & Shevell, 1985). Other studies have shown that faces are represented differ-
ently depending upon the psychological task at hand (Gerrig, Maloney, & Tversky, 1991; Markman, 1999). It is very unlikely that facial representation is a stored, fixed "thing" that is timeless and divorced from context and task. Any single MDS solution is only one of many possible representations of how the human mind represents faces.

Summary

We end this section by reiterating what we claim and what we do not claim about the importance of averageness in human faces. Our claim is that averageness is essential to facial attractiveness. Without averageness, a face will not be judged as attractive. However, we do not claim that averageness is exclusive. Other factors influence the perceived attractiveness of the human face. Youthfulness and symmetry, among others, have all been shown to enhance the attractiveness of a face in certain situations. However, although these factors can enhance attractiveness, none are sufficient causes of attractiveness. For example, a face can be both youthful and unattractive. It is only in combination with averageness that youthfulness creates a highly attractive face. Averageness is the only characteristic discovered to date that is both necessary and sufficient to ensure facial attractiveness—without a facial configuration close to the average of the population, a face will not be attractive no matter how smooth, youthful, or symmetrical. Averageness is fundamental.

MECHANISMS

We turn now to a discussion of the mechanisms underlying preferences for attractive faces. Why do people prefer attractive faces and how might various explanatory mechanisms relate to averageness? Given that even young infants prefer attractive faces, it is very unlikely preferences are socialized, gradually learned, and due merely to media exposure. Instead, two other theoretical accounts suggest potential answers.

Evolution and Innate Preferences for Attractive Faces

Very young infants exhibit preferences for attractive faces. Are such preferences innate or acquired? Evolutionary theory provides an explanation for how preferences for attractive faces might be built into the young organism from birth. Indeed, the averageness hypothesis does not oppose an evolutionary point of view and is in fact quite compatible with it. Stabilizing or normalizing selection is more ubiquitous than directional selection in evolution. Normalizing selection selects for average-size physical characteristics in the species-typical range and selects against extremes (e.g., Barash, 1982). Average physical characteristics, according to this view, are indicative of health and genetic fitness (Thornhill, 1998). Mating with healthy attractive partners
would increase the probability of successful reproduction and genetic trans-
mission: Potential offspring from such a union would have a higher chance of
survival. Thus, an evolved tendency to prefer average physical characteristics
may underlie preferences for attractive “averaged” faces. If such preferences
have evolved, perhaps they are also innate and present at birth.

The evidence on whether or not contemporary attractive people are health-
ier than unattractive people is mixed. When strangers are asked to make attrib-
utions of health based solely on a photograph of the face, attractive people are
consistently assumed to be healthier than unattractive people (Buss & Barnes,
1986). However, studies using only impressions of strangers based on photo-
graphs are not adequate to test the health hypothesis. Because of the
well-known “beauty is good” stereotype, judges have a positive bias toward at-
tractive individuals. Therefore, associations between perceptions of facial at-
tractiveness and perceptions of health in these rating studies probably reflect a
positive halo effect rather than a true relation between attractiveness and real
health.

To investigate whether or not health and facial attractiveness are related,
studies must examine objective measures of health. Only a few have done so,
probably because of the difficulty of obtaining samples for which actual (as op-
posed to perceived) measures of health are available. We were able to locate
five studies in which some kind of nonattributional measure of health as a func-
tion of facial attractiveness was available. Reis, Wheeler, Nezlek, Kernis, and
Spiegel (1985) collected a number of health indices but found few significant
associations between attractiveness and health in college students. A single
moderate relation was found in which unattractive students visited the student
health center more frequently than did attractive students. Hansell, Sparacino,
and Ronchi (1982) found a relation between attractiveness and health as as-
sessed by blood pressure. Unattractive people had higher blood pressure than
attractive people. However, the interpretation of both visits to a student health
center and blood pressure are problematic. Are less attractive people funda-
mentally less healthy than attractive people? Or do unattractive people visit the
health center and have higher blood pressure because life is more stressful for
them because they are judged and treated less favorably than attractive people
are (Langlois et al., 2000)?

Shackelford and Larsen (1999) obtained self-reports of health and objec-
tively assessed heart rate recovery in a study of 66 undergraduate women and
34 undergraduate men. Self-reported symptoms of the minor complaints as-
sessed in this study (headache, stuffy nose, muscle soreness) are neither objec-
tive nor important measures of real health and should not be used to evaluate
“health.” For heart rate recovery, a single significant correlation emerged for
the small sample of males but not for the larger sample of females. Thus, the re-
sults of this study provide weak and equivocal support for the proposed link
between health and attractiveness.
Using the archival records of a longitudinal sample ranging in age from 11 to 66, Kalick, Zebrowitz, Langlois, and Johnson (1998) found that, if anything, adolescent attractiveness falsely advertised health: Although raters who made health attributions judged attractive individuals to be more healthy than unattractive individuals, the correlation between facial attractiveness and a variety of objective composite health scores (derived from physician and hospital records) was actually slightly negative. The study, which benefited from a large sample size, also found no significant relation between attractiveness and the number of offspring. Thus, the results of the Kalick and colleagues study question the evolutionary assumption that attractiveness is related to actual health, at least in individuals born between 1920 and 1929.

Finally, Cronin, Spirduso, Langlois, and Freedman (1999) separated adult health and physical fitness into different components and obtained actual performance measures for each. Like the results of Kalick and colleagues (1998), no significant relation between attractiveness and health emerged. Interestingly, however, there was a statistically significant relation between facial attractiveness and physical fitness. Fitness (endurance, strength, balance, and flexibility) was defined as “the ability to carry out daily tasks with vigor and alertness, without undue fatigue, and with ample energy to enjoy leisure-time pursuits and to meet unforeseen emergencies” (U.S. Department of Health and Human Services, 1996). Health (resting heart rate, blood pressure rankings, and body fat percentages) was defined as a positive state of physical, mental, and social well-being or the absence of disease and illness (U.S. Department of Health and Human Services, 1996). Health status and fitness may differ in their relation to attractiveness ratings because they represent two different dimensions of performance. Health status, for example, is not the highest range of physical potential for a person; the average person may have adequate health, but the fit individual surpasses “good” or average health. Thus, a level of fitness beyond adequate health may be required to observe a relationship between physical function and attractiveness, at least in our modern, primarily healthy society. Obviously, very poor health would contribute to a person’s appearance, but, within the normal range of older adults, attractiveness predicts fitness better than resting heart rate, blood pressure rankings, and body fat percentages.

Another approach to testing mechanisms underlying the preference for attractiveness is to determine whether or not such preferences are innate. The term innate is used in different ways by different people. Here, we use the term as many developmental psychologists do and consistent with the preferred dictionary definition: present at birth, and not requiring any experience to be manifested. Logically, if preferences for attractive faces are not present at birth, evolutionary mechanisms cannot be ruled out because many innate characteristics are not expressed until later in development (i.e., puberty). However, if preferences for attractive faces are present at birth, many experiential accounts of how preferences for attractiveness develop can be ruled out. Thus, studying
newborn infants can yield insight into the mechanism responsible for the ubiquitous preferences for attractive faces.

Some research suggests that preferences for faces over nonfacial patterns are evident in newborns prior to any direct visual experience with faces (Goren, Sarty, & Wu, 1975; Morton & Johnston, 1991). In these newborn studies, infants are presented with geometric shapes configured in the shape of a human face or in a scrambled configuration. Infants pay significantly more attention to stimuli representing a facial configuration, indicating an innate predisposition to pay attention to the human face. Indeed, there may be a selective advantage for infants to prefer faces. Attending to facial stimuli is important for social interaction, so infants may have a built-in preference for faces (e.g., Linn, Reznick, Kagan, & Hans, 1982). Perhaps if infants are born with preferences for faces over nonfaces, they also are born with preferences for attractive over unattractive faces, given that attractive faces are better “examples” of faces than are unattractive faces.

To determine whether preferences for attractiveness are innate, it is necessary to conduct research with newborns that have no experience with faces. Needless to say, such research is very difficult to conduct. Some research has been conducted with newborns, but the evidence from these studies for an innate attractiveness preference is mixed. Slater and colleagues (1998) used a visual preference paradigm and measured looking times toward pairs of attractive and unattractive female faces in very young infants (average age = 72 hours). The infants looked longer at the attractive faces. However, two limitations of the study make innateness a questionable interpretation of the data. The first is that the infants in the study ranged in age from 14 hours to 151 hours and it is therefore likely that they had substantial experience with faces prior to participating in the study. Early and rapid learning about faces based on experience has previously been demonstrated by Walton and Bower (1994). Walton and Bower showed newborn infants individual faces and then presented them with a face that was the mathematical average of the individual faces. The newborns reacted to the averaged face as a familiar face, indicating that the infants were encoding facial information from very brief exposures. Thus, innate preferences for attractive faces can be claimed only if such rapid learning is controlled or eliminated.

The second limitation of the Slater and colleagues (1998) study is that the infants were held by an experimenter who supported the infant’s head and body and was not blindfolded, or in some other way prevented from seeing the stimuli, as is typical in most studies involving infant looking times (e.g., Langlois et al., 1987; Quinn, 1987; Younger, 1985). It also is possible, therefore, that the infants’ looking behavior was subtly affected by preferences of the experimenter who was holding the infant at the time of the study.

In her dissertation work at the University of Texas, Lisa Kalakanis (1997) tested two different samples of infants who were only about 15 minutes old. Kalakanis was present in the delivery room and recorded the infants’ experi-
ence with faces. Unlike Slater and colleagues, she used a visual tracking procedure like that of Morton and Johnson (1991). In this procedure, the infant is lying on a crib and follows a moving photograph of either an attractive or unattractive face. The experimenter has no opportunity to influence infant looking behavior. Furthermore, all tracking was videotaped and coded by observers who could not see which face the infant was following, a procedure that thereby eliminated the possibility of experimenter/observer bias. The results of the study were mixed: Although the infants did seem to track attractive faces further than unattractive faces in some pairs, most of the time they did not. Given the methodological problems of the Slater and colleagues study and the mixed results of the Kalakanis study, it is premature to conclude that preferences for attractive faces are innate and do not require exposure or experience. Additional studies with newborns are critical to resolve the issue.

Cognitive Theory and Experience-Based Preferences for Attractive Faces

The literature in cognitive psychology suggests another mechanism underlying the preferences of both infants and adults for attractive faces. Cognitive representations, called prototypes, may be defined as a central exemplar or “average” of a category. Preferences for prototypes have been shown in several studies using a variety of categories of stimuli: color categories (Martindale & Moore, 1988), object categories (Whitfield & Slatter, 1979), and musical categories (Smith & Melara, 1990). Prototypes are perceived as “typical” and are “good” examples of a category of stimuli (Evans, 1967). Furthermore, faces rated as more attractive are also rated as more typical than less attractive faces (Light, Holland, & Kayra-Stuart, 1981). Averaged faces, therefore, which are attractive, might be prototypic of the human face.

Prototypes are recognized as familiar even when never seen before—perhaps this is why both adults and infants prefer prototypical faces. Posner and Keele (1968) generated families of dot patterns, in which each dot varied in distance from a prototype (averaged from the examples) pattern of dots. Adults identified the prototype as familiar even when they saw it for the first time, indicating that they had cognitively “averaged” across the individual exemplars to form a prototype. Likewise, Bomba and Siqueland (1983) showed dot patterns to infants and found that they also responded to an averaged prototype as if familiar despite having never seen it before. Similar results have been found in infants viewing schematic drawings of faces. When shown a novel face with features that were averaged feature values from faces previously shown, infants looked longer at the face. Thus, the new face was treated as a familiar member of the category, or a prototype, even though they had not previously seen that particular configuration of facial features (Strauss, 1979).

Can infants extract a prototype face from complex images of real faces? Can infants cognitively average across faces they experience to form a prototype? If
they cannot, an information processing or cognitive averaging account is an unlikely explanation for their preferences for attractive faces.

We empirically evaluated the viability of a cognitive averaging explanation for attractiveness preferences (Rubenstein et al., 1999). Six-month-olds were tested in a familiarization procedure, similar to that used by Bomba and Siqueland (1983). In this paradigm, infants are familiarized to a group of category exemplars, in our case, normal faces. After they are familiarized, they are shown multiple test comparisons in which looking time is measured. The general idea is that the infants' preference for novel stimuli, after a familiarization procedure, will reveal which test stimuli seem familiar to them and which do not.

In our research, the infants were familiarized to eight unfamiliar, attractive female faces, presented sequentially and repeatedly. Once the infants were familiarized with the faces, they were shown test comparisons in which different types of individual faces were paired with the prototype face created by averaging together the familiarization faces. The test faces were all equal in attractiveness to ensure that, in this case, infants were responding to novelty or familiarity, not attractiveness. Both familiar and novel faces were paired with each other and with the prototype/averaged face in the test trials. If infants can and do abstract the averaged representation from multiple exemplars of faces, the average of those faces should seem familiar to them, even though they have never seen it before, because it would match their internal representation formed during familiarization. Thus, according to the assumptions of this procedure, if infants are capable of averaging across faces to form prototypes, they will demonstrate a preference for novel faces over the averaged face because the never-before-seen averaged face would be experienced as familiar.

The infants in this study demonstrated a pattern of looking showing that they perceived the averaged face as familiar, even though they had never previously seen this face. Thus, cognitive averaging can explain both preferences for attractive faces (prototype faces) and why averagedness is attractive. Through exposure to facial exemplars, even young infants can form a prototypical representation of the face.

The finding that infants look longer at novel, individual faces than at averaged faces is not contradictory with previous research showing that infants prefer both averaged and attractive faces (e.g., Langlois et al., 1991a; Rubenstein et al., 1999). First, recall that in the familiarization study, all the faces were equal in attractiveness, as judged by adults, so that the infants were selecting their preferred face based on it's familiarity, not on it's attractiveness. Second, the logic of the familiarization procedure is very different from that of the visual preference procedure. In the visual preference procedure, infants see two novel faces simultaneously. Babies, therefore, can choose to look at the novel face they prefer. In the familiarization procedure, the experimenter deliberately exposes infants to faces until the babies become bored with them—the faces become excessively familiar to the infants. Once bored with a
particular stimulus, infants will look at something novel, assuming they remember the previous, overly familiar stimuli. If babies are shown the face of their own mothers over and over again, they will become bored and look longer at a subsequently shown novel face. This does not mean they do not prefer their mother to the novel face. It merely means that they are able to recognize and remember something very familiar in this experimental context. Thus, the different predictions about infant looking behavior in these two different experimental paradigms are logically consistent and not contradictory (see Hunter & Ames, 1988, for a review of infant preferences for familiarity and novelty).

The proximal cognitive averaging account can nicely account for race and ethnic differences in preferences for attractiveness. Although raters from many different cultures and ethnic groups significantly agree about who is and isn’t attractive, and their rank orders of individual faces are highly similar, there are still group differences evident in the literature. Many studies demonstrate an in-group bias in which the in-group evaluates their own group as more attractive than an out-group (e.g. Moss, Miller, & Page, 1975). In-group bias would follow from different experience with different kinds of faces—the faces that are averaged together are different from group to group. Thus, although averaged faces (of 32 or more faces) from different ethnic groups are judged as attractive, ethnic differences in the facial features of the averaged face may still be preserved and preferred by the in-group. The reason that there is so much cross-group similarity in ranking the attractiveness of different types of faces is because faces from different ethnic groups are much more similar than they are different. And, any face, regardless of unique ethnic features, will be judged as attractive if it is close to the average of the central tendency of the population, according to our point of view.

The work showing that averaged faces are attractive, together with the work showing that infants form prototypes (averages) of faces, suggest that cognitive averaging may be the proximal ontogenetic mechanism underlying preferences for facial attractiveness. The work also suggests that preferences for attractive faces are not innate but rather that experience with faces is necessary for these preferences to develop. The amount of experience required, however, is not extensive because the brain of the human infant has probably evolved to process faces efficiently and to average across many different types of experienced stimuli. Thus, the work on cognitive averaging does not suggest that evolution is irrelevant to or in conflict with the proximal cognitive account. Cognitive averaging is an efficient information-processing strategy with likely adaptive consequences.

CONCLUSION

At the beginning of this chapter, we introduced a popular maxim concerning physical attractiveness: “Beauty is in the eye of the beholder.” We now
know through empirical research that this maxim is false. Furthermore, the theoretical perspective described in this chapter can parsimoniously explain why beauty is not in the eye of the beholder. Mathematical averageness is a necessary and fundamental characteristic of perceived attractiveness in the human face and the concept of averageness has theoretical deep roots in both evolutionary and cognitive psychology. The explanatory power of averageness theory is robust. Although other characteristics of the human face may be in some way associated with attractiveness (e.g., symmetry, youthfulness), none of these competing explanations can fully explain attractiveness preferences when averageness is removed from the equation nor can they explain the attractiveness of averaged faces.

Think of averageness like a cake—there are basic ingredients that are required before a cake can be a cake. Flour, sugar, eggs, and butter are fundamental—they are both necessary and sufficient to make a cake. There are other ingredients, however, that create cakes with individual identities: adding carrots produces carrot cake, adding chocolate produces chocolate cake. And finally, adding icing to the cake will make a sweeter cake. But without the flour, sugar, eggs, and butter, all you have is icing; without the flour, sugar, eggs, and butter, all you have is a carrot. For faces, characteristics like youthfulness and symmetry are analogous to icing and carrots—they are nice but not necessary. And, although we agree that adding youthfulness to averageness is surely better than adding wrinkles, without averageness, you are left with something that is nice but neither necessary nor sufficient.

What is the benefit of identifying the fundamental characteristic of a human face that makes it attractive and what is the benefit to understanding the manner in which attractiveness preferences are formed? Although averageness can explain how and why we prefer attractive faces, we must also focus on why it is important to understand what makes a face attractive. “Beauty is in the eye of the beholder” is not the only commonly heard piece of conventional wisdom that concerns attractiveness. For instance, “beauty is only skin deep” indicates that an individual’s appearance is not a sincere reflection of their internal personality traits. “Never judge a book by its cover” dictates that we should not base our treatment of others upon their outward physical appearance. As the first maxim was proven false, these two additional sayings are also not supported empirically—they imply truisms that are not true. People are judged by their covers as if their beauty reflected inner characteristics deeper than their skin. Physical attractiveness is a salient piece of information that helps define our social interactions.

Through investigating the characteristics that cause a face to be perceived as attractive, we may gain greater insight as to how the physical attractiveness stereotypes that so greatly affect our social interactions are formed and how they may be attenuated. We believe that our conception of attractiveness in terms of mathematical averageness and our explanation of the evolutionary and cognitive mechanisms behind the development of attractiveness preferences can
provide a substantial and necessary stepping stone to further address the impact of attractiveness on our society.

NOTE

1. Rosenthal (1991) indicates that two types of reliability statistics can be analyzed using meta-analysis: mean interrater reliabilities (used to estimate agreement between specific pairs of judges) and effective reliabilities (used to estimate the reliability of the mean of the judges’ ratings). In our analysis of effective reliabilities, mean interrater reliabilities were converted to effective reliabilities in order to investigate general agreement about facial attractiveness.

REFERENCES


