Personality and Social Psychology Review
2003, Vol. 7, No. 3, 194–215



# Trait Impressions as Overgeneralized Responses to Adaptively Significant Facial Qualities: Evidence from Connectionist Modeling

**Leslie A. Zebrowitz**
*Department of Psychology*
*Brandeis University*

**Jean-Marc Fellous**
*Computational Neurobiology Laboratory*
*Salk Institute*

**Alain Mignault**
*Department of Psychology*
*McGill University*

**Carrie Andreoletti**
*Department of Psychology*
*Brandeis University*

*Connectionist modeling experiments tested anomalous-face and baby-face overgeneralization hypotheses proposed to explain consensual trait impressions of faces. Activation of a neural network unit trained to respond to anomalous faces predicted impressions of normal adult faces varying in attractiveness as well as several elderly stereotypes. Activation of a neural network unit trained to respond to babies' faces predicted impressions of adults varying in babyfaceness as well as 1 elderly stereotype. Thus, similarities of normal adult faces to anomalous faces or babies' faces contribute to impressions of them quite apart from knowledge of overlapping social stereotypes. The evolutionary importance of appropriate responses to unfit individuals or babies is presumed to produce a strong response preparedness that is overgeneralized to faces resembling the unfit or babies.*

We are enjoined not to judge a book by its cover, and we are cautioned that beauty is only skin deep. These warnings suggest that our natural proclivity is in fact to judge people by their appearance and to prefer those who are beautiful. Hundreds of research articles have demonstrated that this is so. There are strong consensual impressions of attractive individuals, who are perceived to possess the positive traits of social and intellectual competence, dominance, and health. Babyfaceness also produces strong and consensual impressions that are independent of attractiveness. We attribute childlike traits to baby-faced people of all ages, perceiving them as less dominant and less strong as well as warmer and more naive than their more mature-faced peers. Attractiveness and babyfaceness influence not only impressions of people but also their social outcomes. Moreover, the social consequences of facial appearance represent large effects, comparable in magnitude to the effects of an individual's gender or personality traits. It is clear from this research that people are singularly unsuccessful in adhering to conventional wisdom about facial appearance (for reviews see Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992; Langlois et al., 2000; Montepare & Zebrowitz, 1998; Zebrowitz, 1997).

The ubiquitous tendency to judge others by their appearance and the consensus shown in those judgments is remarkable. Why do people use facial appearance in first impressions, and how do they achieve a consensus? It was long assumed that positive impressions of attrac-

This research was supported by NSF Grant 97-08619 to the first and second authors. The first author thanks Eliot Smith for his excellent workshop on connectionist modeling at the 1996 meeting of the Society of Experimental Social Psychology, which inspired this research, as well as his comments on an earlier draft of this article. Appreciation is also extended to Rahul Pande for his extensive help with the connectionist modeling, to Mandy Smith for her help selecting faces and measuring facial distances, and to Jen Bunk, Katya Berezin, Coreen Farris, Deena Grant, Marina Kiskovich, So Young Lee, Nandita Murukutla, Valerie Pasternak, Jordan Robbins, and Vinavak Sharma, for their help with various aspects of the data collection and analysis. Finally, the authors thank the anonymous reviewers for their helpful comments on an earlier draft of this article.

Requests for reprints should be sent to Leslie A. Zebrowitz, Brandeis University, Department of Psychology, MS 062, Waltham, MA 02454–9110. E-mail: zebrowitz@brandeis.edu

tive individuals reflected cultural influences, but this explanation begs the question of the origin of those cultural values. Moreover, research suggests that a cultural explanation is insufficient to explain stereotypes of attractive or baby-faced people. Evidence that some universal process is involved is provided by the findings that even young infants prefer to look at faces of attractive or baby-faced adults (e.g., Kramer, Zebrowitz, San Giovanni, & Sherak, 1995; Langlois, Ritter, Roggman, & Vaughn, 1991). Stereotyped impressions of attractive and baby-faced individuals emerge as early as the preschool years (Dion, 1973; Keating & Bai, 1986; Montepare & Zebrowitz-McArthur, 1989), and there is considerable cross-cultural agreement in judgments of attractiveness and babyfaceness and associated traits (Dion, 2002; Zebrowitz, Montepare, & Lee, 1993). Three hypotheses have been proposed to account for the universality of face impressions.

## Good Genes Hypothesis

The good genes hypothesis proposed by evolutionary theorists provides an accuracy explanation for an influence of attractiveness on impressions. This hypothesis holds that attractive faces signal mate quality, and preferences for attractive individuals evolved because they enhance reproductive success (see Berry, 2000, for a review of pertinent theories). On this account, impressions of attractive people as healthier, more intelligent, and more socially skilled than their less attractive peers are accurate. However, existing research has provided little support for the thesis that attractiveness signals health (Kalick, Zebrowitz, Langlois, & Johnson, 1998; Shackelford & Larsen, 1999), although it is possible that the relation is attenuated in modern societies (Gangestad & Buss, 1993). There is some evidence for accuracy in impressions of attractive individuals as more popular, intelligent, and dominant (Langlois et al., 2000; Zebrowitz, Hall, Murphy, & Rhodes, 2002; but see Feingold, 1992, for lack of evidence). However, even the small degree of accuracy that has been documented does not necessarily reflect the coevolution of facial attractiveness and adaptive traits. Rather, such relations can also be explained by a variety of nonevolutionary mechanisms. Most notably, there may be self-fulfilling prophecy effects whereby stereotypes of attractive people influence their social environments, which in turn may influence their resultant traits (Snyder, Tanke, & Berscheid 1977; Zebrowitz, Collins, & Dutta, 1998; Zebrowitz et al., 2002). If a self-fulfilling prophecy accounts for accuracy in judging people who vary in attractiveness, then the question remains as to where the prophecy comes from. The question of where the expectations come from also applies to a self-fulfilling prophecy account for accuracy in judging people who vary in babyfaceness. Moreover, research investigating the accuracy of such judgments has revealed that more often than not, baby-faced adults have psychological qualities that are opposite to the stereotype. This finding has been attributed to a self-defeating prophecy effect, whereby baby-faced adults overcompensate for undesirable expectations (Zebrowitz, Andreoletti, Collins, Lee, & Blumenthal, 1998; Zebrowitz, Collins, et al., 1998).

Although there is little evidence to support the good genes hypothesis within the range of faces that have been studied, the fact is that extremely unattractive faces often mark people with congenital or genetic anomalies that severely impair their fitness. Individuals with anomalies, such as Down syndrome, fetal alcohol syndrome, and schizophrenia, not only suffer from poor health or social and intellectual incompetence, but also they are marked by faces that are atypical or asymmetrical—two hallmarks of unattractiveness (e.g., Campbell, Geller, Small, Petti, & Ferris, 1978; Cummings, Flynn, & Preus, 1982; Guy, Majorski, Wallace, & Guy, 1983; Krouse & Kauffman, 1982; Paulhus & Martin, 1986; Streissguth, Herman, & Smith, 1978; Thornhill & Møller, 1997). In view of this, a "bad genes" account of facial preferences may be more useful than the good genes account. Those who avoided mates with extremely unattractive faces would have increased their reproductive success as well as the survival of their offspring. Such a mechanism is consistent with the conclusion of Grammer, Fink, Juette, Ronzal, and Thornhill (2002) that the cognitive decision making in attractiveness ratings may be best simulated by a strategy of "simply avoid the worst," as well as with evidence that negative stimuli have a more powerful influence than positive stimuli on human judgments and emotions across a wide range of contexts (cf. Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001). It is also consistent with the argument that there was little need to select the most attractive mates to ensure viable progeny in our evolutionary past because the vast majority of sexually mature people were (and still are) capable of producing healthy offspring (Hazan & Diamond, 2000). Although a bad genes account provides a plausible explanation for accurate negative-trait impressions of faces at the far negative end of the attractiveness spectrum, it provides no obvious explanation for variations in impressions within the normal range of attractiveness where faces are not differentiated by any genetic anomaly.

## Perceptual By-Products Hypotheses

Several theorists have proposed that universally positive reactions to attractive faces may have evolved as the by-products of more general perceptual mechanisms that facilitate the recognition of individuals or objects from different viewpoints, or the abstraction of prototypes from structurally distinct classes, so as to generalize across visually similar exemplars (Endler & Basolo, 1998; Enquist & Arak, 1994; Enquist & Johnstone, 1997). Such mechanisms could contribute to the documented attractiveness of symmetrical faces and average faces (Halberstadt & Rhodes, 2000; Johnstone, 1994; Langlois & Roggman, 1990; Rubenstein, Langlois, & Roggman, 2002). It also has been argued that the attractiveness of other facial qualities may be a by-product of perceptual mechanisms that enable people to recognize sex, age, mature status, or positive emotion (Cunningham, Barbee, & Philhower, 2002; Enquist, Ghirlanda, Lundqvist, & Wachtmeister, 2002; Keating, 2002). Whereas these perceptual mechanisms have proven useful in explaining why certain structural features would make a face attractive, they do not in and of themselves explain why these features are associated with positive traits, such as social skills, intelligence, and health.

## Overgeneralization Hypotheses

Overgeneralization hypotheses, derived from the ecological theory of social perception, merge the wisdom of the good genes and the perceptual by-products hypotheses to provide an explanation for the traits attributed to faces that vary in attractiveness as well as other qualities (Gibson, 1979; McArthur & Baron, 1983; Zebrowitz, 1990, 1997; Zebrowitz & Collins, 1997). Like the good genes hypothesis, ecological theory holds that trait impressions of faces reflect evolutionarily adaptive preferences that are often accurate. But, like the perceptual by-products hypotheses, ecological theory also holds that our reactions to faces sometimes may be the erroneous by-product of perceptual biases that serve a general adaptive function. Putting the two tenets together, ecological theory predicts that adaptive, accurate trait impressions of certain faces are overgeneralized to other faces that are physically similar to them.

## Anomalous-Face Overgeneralization Hypothesis

The evolutionary importance of recognizing individuals with bad genes may have produced such a strong tendency to respond to their anomalous facial qualities that responses are overgeneralized to normal adults whose faces merely resemble those who are unfit. For example, the perception of normal individuals with somewhat asymmetrical or nonaverage faces as unattractive, unhealthy, and unintelligent (Grammer & Thornhill, 1994; Langlois & Roggman, 1990; O'Toole, Price, Vetter, Bartlett, & Blanz, 1999; Perrett et al., 1999; Rhodes, Proffitt, Grady, & Sumich, 1998, Rhodes & Tremewan, 1996; Rhodes et al., 2001; Zebrowitz et al., 2002; Zebrowitz, Voinescu, & Collins, 1996) may be an overgeneralized response that derives from the adaptive rejection of abnormal individuals who show marked asymmetry and nonaverageness and who do, in fact, lack health, intelligence, or other evolutionarily adaptive qualities. According to ecological theory, the errors shown in such overgeneralization effects occur because they are less maladaptive than those that might result from failures to respond to fitness information. However maladaptive it may be to reject healthy and fertile unattractive individuals as mates, it would be even more maladaptive to select those with craniofacial anomalies who do, in fact, lack the requisite health and fertility. This account differs from the good genes hypothesis not only by focusing on the evolutionary significance of attending to bad genes but also by making no claim for the accuracy of reactions to faces. It differs from the perceptual by-products hypotheses by explaining why particular faces are perceived to have particular traits. In sum, the anomalous-face overgeneralization hypothesis argues that the attractiveness "halo effect" is a perceptual by-product of negative reactions to individuals with bad genes. (For related discussions, see Kurzban & Leary, 2001; Neuberg, Smith, & Asher, 2000.)

## Baby-Face Overgeneralization Hypothesis

The overgeneralization of accurate and highly adaptive social perceptions may also explain impressions of people who vary in babyfaceness. The evolutionary importance of identifying babies may have produced such a strong tendency to respond to their facial qualities (Todd, Mark, Shaw, & Pittenger, 1980) that responses are overgeneralized to those whose faces merely resemble those of babies. For example, the perception of adults with round faces or big eyes as socially, intellectually, and physically weak (Montepare & Zebrowitz, 1998; Zebrowitz, 1997) may be an overgeneralized response that derives from the accurate and adaptive perception of babies, who do indeed lack the social, intellectual, and physical resources to care for themselves. Again, the errors shown in such overgeneralization effects occur because they are less maladaptive than those that might result from a failure to respond appropriately to babies. However maladaptive it may be to deny baby-faced individuals dominant leadership jobs on the mistaken assumption that they lack the necessary qualifications, it would be more maladaptive to give autonomy to babies who do, in fact, lack the requisite capabilities.

## Research Evidence

Currently there is indirect evidence consistent with the hypothesis that impressions of people who vary in attractiveness or babyfaceness derive from an overgeneralization of reactions to those who are unfit or to babies. Psychological qualities attributed to unattractive people, such as low social skills, low intelligence, and poor health, do mirror the actual traits of those who are unfit, and facial qualities that mark unattractive people—asymmetry and nonaverageness—also mark those who are unfit. Similarly, psychological qualities attributed to baby-faced people, such as physical weakness, submissiveness, and naiveté, mirror the actual traits of babies, and facial qualities that mark baby-faced people—round face, large eyes, high eyebrows, small chin and nose bridge—also mark real babies (Enlow, 1990). However, the assumption that impressions of people who vary in attractiveness or babyfaceness can actually be predicted from their resemblance to the unfit or to babies has not been directly tested. The experiments reported in this article tested this assumption using connectionist models.

## The Utility of Connectionist Modeling for Testing the Overgeneralization Hypotheses

The essence of the anomalous-face and the baby-face overgeneralization hypotheses is that first impressions of people can derive from their facial resemblance to genetically anomalous individuals or babies. This type of similarity-based generalization is a natural property of connectionist models. Connectionist networks that have been trained to discriminate anomalous and normal adult faces will react to other faces according to their similarity to anomalous versus normal faces; networks trained to discriminate faces of babies and adults will react to other faces according to their similarity to babies versus adults. Network activation to the untrained faces captures the network's overgeneralization of veridical fitness or maturity information to those faces. In this article, the similarity of a face to an anomalous face or to a baby is operationalized as the extent to which it activates a network unit that has been trained to react to anomalous faces or babies. If the neural network's assessment of the physical similarity of faces to babies or those who are genetically unfit predicts impressions of their traits, this can provide support for the anomalous-face or the baby-face overgeneralization hypothesis.

An advantage of connectionist modeling for testing the overgeneralization hypotheses is that it reveals whether the physical similarity between two faces can in and of itself predict similar impressions of them. For example, if a neural network finds unattractive people physically similar to genetically unfit people, and this resemblance predicts trait impressions, this cannot be attributed to knowledge of overlapping social stereotypes about unattractive and unfit individuals. Rather, it can be due only to intrinsic similarities in the facial attributes known to the neural network. These attributes uniquely represent the face and are devoid of any social judgments. The feasibility of training a network to differentiate faces that vary in fitness or maturity is supported by previous research in which neural networks have been trained to recognize the identity of faces (Samal & Iyengar, 1992) and to differentiate faces that vary in sex (Golomb, Lawrence, & Sejnowski, 1991), race (O'Toole, Abdi, Deffenbacher, & Bartlett, 1991), and emotion (Bartlett, 2001; Cottrell, Dailey, Padgett, & Adolphs, 2001; Golomb et al., 1991; Lyons, Budynek, & Akamatsu, 1999; Mignault & Marley, 2001).

Given that connectionist models were used to test hypotheses derived from an ecological theory of perception (Gibson, 1966, 1979; McArthur & Baron, 1983), the relation between these two paradigms merits some discussion. Connectionist modeling is sometimes viewed as using internal representations, which may seem an anathema to the concept of *direct perception* and the eschewal of internal representation in Gibson's (1979) ecological theory of perception. However, the internal representations in connectionist models bear a striking similarity to one of Gibson's (1966) concepts. Specifically, the response of a neural network to a particular stimulus has been described with the metaphor of *resonance* (Smith, 1996), which is the term Gibson used to describe how the brain works. "Instead of postulating that the brain constructs information from the input of a sensory nerve, we can suppose that the centers of the nervous system, including the brain, resonate to information" (Gibson, 1966, p. 267). Consistent with this concept of resonance is the fact that there is no discrete representation of a schema in connectionist models. "Rather, schemata emerge at the moment they are needed from the interaction of large numbers of much simpler elements all working in concert with one another … and are created by the very environment that they are trying to interpret" (Rumelhart, Smolensky, et al., 1986, p. 20). Similarly, the ecological approach "assumes that the past history of one's interaction with the environment consistently retunes the perceptual apparatus on an online basis" (McArthur & Baron, 1983, p. 234). The resonance metaphor also reconciles what could be construed as an inconsistency between the overgeneralization hypotheses and the presumed veridicality of perception in Gibson's (1966, 1979) theory. As Shepard (1984) noted in his attempt at a rapprochement between Gibson's theory and internal representations, a resonant system is excited most by the pattern of energy to which it is tuned, but "it is also excited, though to a lesser degree, by a signal that is slightly different, weaker, or incomplete" (Shepard, 1984, p. 433). In addition to the shared concept of resonance, another synchrony between ecological theory and connectionist modeling is the responsiveness of the latter to higher order invariants. According to Gibson (1966, 1979), it is these invariants, rather than simple first-order stimulus qualities, that reveal the functionally significant properties of the external world. Applied to the domain of face perception, for example, the meaning of a face is conveyed not by simple metrics, such as the length of the nose or the width of the eyes, but rather by configural properties. Connectionist models are ideally suited to combining simple metric inputs into higher order, configural qualities that may differentiate two categories of faces. In sum, connectionist modeling does not rely on the sort of static internal representations that Gibson's ecological theory would reject, and Gibson's theorizing does suggest internal processes that are well instantiated in connectionist modeling.

## Overview of Experiments

A series of experiments tested the hypothesis that impressions of people can be predicted from the resemblance of their facial structures to adaptively significant facial qualities that mark unfit individuals or babies. There were four components to each experiment. First, in the *training phase,* facial metrics (discussed later) were provided as input to standard back-propagation neural networks that were trained to differentiate faces varying in fitness or maturity. There were two output units in each network, representing two levels of facial fitness (anomalous, normal) or two levels of facial maturity (baby, adult). In the second or *test phase* of each experiment, the trained network was tested on another set of faces that differed in the attribute on which the network was trained to establish that training was successful. In the third, *generalization phase,* the trained network was provided with input metrics from a new set of faces that did not vary in the adaptively significant attribute, and the extent to which the output units responded to each of these faces was determined. These three phases were repeated for 20 trials to establish a reliable index of network activation by each face. Finally, human judges' impressions of the faces were predicted from the activation of the output units to determine whether these impressions covaried with the network's overgeneralization of veridical fitness or maturity information to those faces. Because activation of one output unit in a pair was the mirror value of the other (i.e., 100 – value), prediction of human judges' impressions was made from activation of the anomalous- but not the normal-face output unit or from activation of the baby- but not the adult-face output unit.

## Experiment 1: Anomalous-Face Overgeneralization Effect

We tested the hypothesis that trait impressions of normal adult faces varying in attractiveness would be predicted from the extent to which they activate a neural network unit trained to recognize anomalous faces. Normalized deviation-from-average distances plus asymmetries were inputs to the neural network. Deviation distances were computed by subtracting each distance for a given face from the average value of that distance in a different set of young adult faces of the same sex. We used deviation-from-average distances, rather than raw distances, because judgments of facial attractiveness have been shown to be highly sensitive to the former (for a review see Zebrowitz & Rhodes, 2002) and also because we expected that the facial metrics of anomalous and normal faces would be differentiated more by their degree of deviation-from-average than by their raw values. This is because genetic vulnerability to environmental and developmental stressors has been theoretically and empirically linked to nonaverageness, as well as to asymmetry (Gangestad & Buss, 1993; Thornhill & Gangestad, 1993; Thornhill & Møller, 1997).[1]

### Faces

Training–test anomalous faces were 60 White young adults (30 men and 30 women) drawn from atlases depicting birth defects and syndromes characterized by facial deformities.[2] Training–test normal faces were 60 White young adults (30 women and 30 men) with a mean age of 19.2 years drawn from the Intergenerational Studies archive (IGS), a representative sample of 386 individuals who previously had been rated on a 7-point scale with endpoints labeled *unattractive/attractive* (see Zebrowitz, Olson, & Hoffman, 1993, for more details). These faces had been rated between the 40th and 60th percentiles in comparison with others of the same sex and age. Generalization faces were 80 White young adults (40 men and 40 women), with predominantly neutral expressions and a mean age of 18 years, drawn from the IGS archive. Forty faces had previously received attractiveness ratings in the top 20% of their sex and age and 40 had been rated in the bottom 20%. The large majority of faces had neutral expressions.

[1]The modeling was also conducted using raw distances as inputs both in this experiment and in Experiments 3 and 4. Although the results using raw distances tended to be somewhat weaker than those using deviations from average, they were generally highly comparable.

[2]A list of sources and a summary table of anomalies are available from the first author.

## Facial Ratings

Sixty-four undergraduate judges rated either male or female training–test or generalization faces, with 16 judges (8 men and 8 women) rating faces in each of these 4 groups. Faces were rated on 7-point scales: ***dominant****/submissive,* ***sociable****/unsociable, na-ive/****shrewd****, unhealthy/****healthy****,* ***intelligent****/unintelli-gent, physically weak/****physically strong****, cold/****warm****, unattractive/****attractive****, mature-faced/****baby-faced****.*[3] The data were coded so that higher scores represented higher levels of the traits in bold print. All faces were rated on one scale before proceeding to the next, and faces were shown for 6 sec during each rating. Two random orders of faces and scales were counterbalanced, with appearance ratings always at the end. Four different judges rated each face on a 7-point scale with endpoints labeled *no smile/big smile.*

## Actual Traits

Because subtle anomalies may be diagnostic of low intelligence and other traits in faces that appear relatively normal (Bell & Waldrop, 1982; Cummings et al., 1982; Krouse & Kauffman, 1982; Paulhus & Martin, 1986; Streissguth, Herman, & Smith, 1978; Thornhill & Møller, 1997; Waldrop & Halverson, 1972), we used measures of actual traits from the IGS archive to assess the possibility that the impressions predicted from network activation could be accurate rather than overgeneralizations. Measures of real health were available for all faces; IQ scores were available for 79 of the 80 faces; a measure of handgrip strength, averaged across right and left hands, was available for 38 faces; and *Q*-sort component scores (*assertive, outgoing, and warm*) were available for 49 faces. There was no measure in the archive that directly mapped onto perceived shrewdness (see Kalick et al., 1998, for a description of the health data; Zebrowitz et al., 2002, for a description of the IQ data; and Zebrowitz, Collins, et al., 1998, for a description of the *Q*-sort data).

## Facial Metrics

Software developed for this study was used to mark a series of points on digitized images of each face viewed on a 21-in. PC monitor. Points were marked by two research assistants on a subset of 40 training and 40 generalization faces, with male and female faces equally represented. After establishing reliability, one judge marked the remaining faces, and those points were used in the final data analyses. A total of 64 facial points were marked (Figure 1). Twenty-seven simple facial distances (2 of which

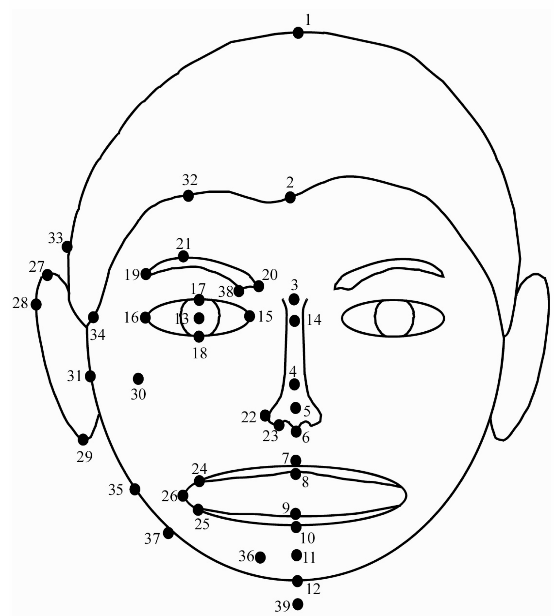


**Figure 1. Location of points. When identical points are marked on the right and left sides, only those on the person's right side are indicated.**

were normalization distances), 1 composite distance, and 3 asymmetry measures were computed from these points using a spreadsheet and automatic procedures written in Visual Basic (Excel, Microsoft). Figure 2 shows the locations of the simple distances on

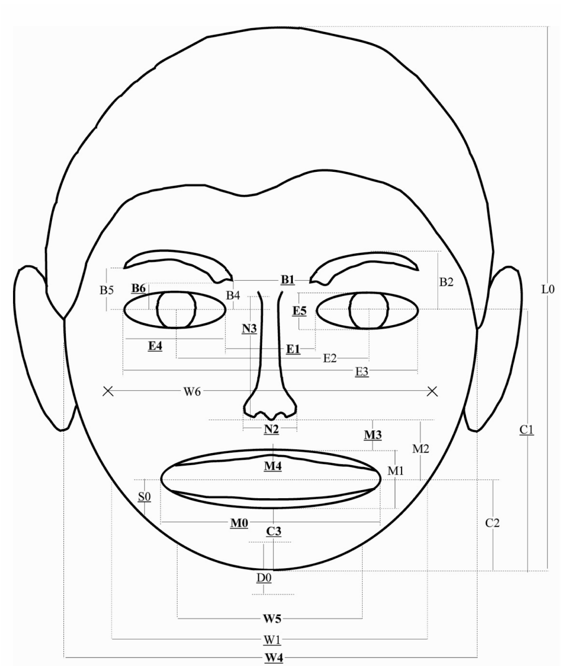


**Figure 2. Location of simple distances. Distances in bold, normalized by LO, were used as inputs in networks trained to differentiate anomalous from normal faces. Underlined distances, normalized by E2, were used as inputs in networks trained to differentiate baby from adult faces. DO indicates length of a double chin, if any, and SO indicates length of a jowl, if any.**

[3]Judges rated one additional trait, *traditional/untraditional,* which is excluded because ratings were unreliable.

the face. A composite variable of facial roundness was computed by determining the average of the radii of the circle connecting Facial Points 31 right, 35 right, and 12, and the circle connecting Facial Points 31 left, 35 left, and 12, with a smaller average radius signifying more roundness. An overall asymmetry measure was computed by summing all of the differences among the midpoints of six horizontal lines represented by the distances E1, E3, W4, N2, W1, and M0. The midpoint of each line was calculated with the formula: (right point + left point)/2 (Grammer & Thornhill, 1994). On a perfectly symmetrical face, all midpoints lie on the same vertical line, and the sum of all possible nonredundant midpoint differences is zero. A measure of eye iris asymmetry (wall-eyed or cross-eyed) was computed using *X* coordinates with the following formula: [(Point 16 right – Point 13 right) – (Point 16 left – Point 13 left)] + [(Point 15 right – Point 13 right) – (Point 15 left – Point 13 left)]. A measure of horizontal eye placement asymmetry (one eye closer to the middle of the face than the other) was computed using *X* coordinates with the following formula: (Point 14 – Point 16 right) – (Point 14 – Point 16 left). Deviation distances were calculated by subtracting each distance for a given face from the average value of that distance in another set of young adult White faces of the same sex (26 men, 25 women) with posed neutral expressions. Raw values, rather than deviations, were always used for the asymmetry measures.

To adjust for variations in distance from the camera, each facial metric was normalized by head length (LO). Another possible normalization metric, used in Experiment 2, was interpupil distance (E2). Because a common anomaly was unusually wide-set eyes, interpupil distance was an inappropriate basis for normalization in the anomalous-face overgeneralization experiments because it would distort differences between anomalous and normal faces in other facial metrics. Because a disadvantage of using head length for normalization is that there is no simple and reliable way to correct for facial angle toward or away from the camera, faces were selected only if they had a completely frontal presentation.

Non-normalization metrics that had achieved acceptable interjudge reliability were selected as inputs if they were not redundant with other inputs, as revealed by spatial overlap and intercorrelations. The metrics that met the selection criteria are the simple distances shown in bold type in Figure 2, facial roundness, and the three asymmetry measures described previously. Our use of normalized facial metrics as inputs to the connectionist network (Fellous, 1996, 1997; Kaiser & Wehrle, 1992) contrasts with previous research modeling face perception in which the inputs were facial images represented as an array of pixels (Bartlett, 2001; Cottrell et al., 2001; Golomb et al., 1991; Lyons, et al., 1999; Mignault & Marley, 2001; O'Toole et al., 1991; Samal & Iyengar, 1992). We used facial metrics rather than pixels as inputs for several reasons. First, an advantage of metric inputs is that they allow the use of theoretically grounded measures, such as facial roundness, asymmetry, and averageness. Second, using facial metrics avoids the dilemma of cropping the face to exclude irrelevant features (such as hair or earrings) or taking the whole face and including these irrelevant features. Third, metrics exclude irrelevant information that a pixel representation cannot, such as makeup. Fourth, metrics provide the possibility of identifying interpretable configurations of facial attributes that differentiate one group of faces from another. This process is less ambiguous than using pixel input. Fifth, the metric inputs representation is more economical than pixels in terms of amount of information. Despite these advantages, it should be acknowledged that the metric inputs do have disadvantages. First, the judges who evaluated the faces saw the pictures with hair and, in some cases, makeup or earrings. Therefore, a discrepancy between the judges' evaluations and the network's predictions could be attributed to these omitted variables. Second, the choice of feature points is somewhat limited, whereas pixel inputs simply include everything that is present in the selected window.

### Connectionist Modeling

The *total set* of faces used to train the network was composed of 120 faces (60 normal, 60 anomalous). Twenty trials (20 networks) were computed. At each trial, 40 normal faces and 40 anomalous faces were randomly selected from the total set to compose the *training set.* The rest composed the *test set.* At each trial, a different network was trained with the training set using different weights initialized at random values. Training was conducted using the Batch Gradient Descent with Momentum algorithm (trainingdm) in Matlab neural network software, version 6 (Neural Network Toolbox, version 4; The Mathworks, Natick, MA). The network was a standard back-propagation neural network with one input layer, one hidden layer, and one output layer. The nodes in the hidden layer were fed only by the input nodes, with each input feeding any or all of the hidden nodes. The output units (anomaly and normal) were fed only by the hidden nodes. The input weight matrices connecting the layers consisted of numbers between –1 and 1. The output units were rescaled into graded values ranging from 0% to 100% activation. All units were nonlinear and responded to their inputs according to a sigmoidal function. After training, the network was frozen and tested on the test set to establish successful learning. The network was then presented with inputs from a new set of faces, the generalization set, which included 80 normal adults (40 women and 40 men) varying in attractiveness. The average activation of the output unit to each of the generalization faces across the 20 trials was used to predict human judges' ratings of those faces.

Training was judged sufficient only if the network could correctly identify at least 90% of the 80 training faces and at least 75% of the 40 test faces, averaged across 20 trials. To achieve these criteria, four parameters were adjusted: number of hidden nodes, learning rate, number of training epochs, and error goal. If more than one combination of the foregoing parameters achieved the training criterion, then the combination that produced the greatest interclass separation of training and testing faces was selected. Further ties were resolved by giving priority to the combination that had the least number of hidden nodes. Preference also was given to parameters that yielded saturation of the generalization faces as close to 50% as possible. The saturation percentage indicated the percentage of generalization faces that activated an output node more than 90% or less than 10%. High- and low-saturation values were avoided because the former could indicate overtraining the network, and the latter could indicate incomplete learning. In all experiments, the training and generalization faces showed highly reliable levels of node activation across the 20 trials, mean $\alpha = .99$ for training faces and .91 for generalization faces.

## Results

### Reliability of Measures

Interjudge agreement was significant for the normalization distance (LO), $r = .74$, as well as the selected input metrics, mean $r = .84$. Male and female judges showed strong agreement in their trait and appearance ratings, mean $r = .88$, and reliability was assessed across judges of both sexes. The average alpha reliability coefficient across ratings was .91 for both training–test and generalization male faces and .87 for both training–test and generalization female faces. The alpha coefficient for smile ratings across all faces was .84. Data analyses used mean ratings for each face across judges.

### Network Training and Generalization

Training a network to differentiate anomalous from normal faces successfully met the criteria of at least 90% correct identification of the 40 training faces and at least 75% correct identification of the 20 test faces, averaged across 20 trials, with the best solution requiring 9 hidden nodes, 6,000 training epochs per trial, a .055 learning rate, and a .2 error goal. Across the 20 trials, activation of the anomaly output unit was significantly higher for anomalous faces ($M = 80.98$, $SD = 19.57$) than for normal faces ($M = 17.65$, $SD = 14.46$), $F(1, 118) = 406.39$, $p < .0001$. As expected, given that the generalization faces were all normal, they did not produce high activation of the anomaly unit ($M = 26.96$, $SD = 24.54$). Nevertheless, there was considerable variability in the extent to which these normal young adult faces activated the anomaly unit (range = 1.40–92.16).

### Impressions of Anomalous Versus Normal Faces

The anomalous-face overgeneralization effect provides a potential explanation for impressions of people on those traits that are judged to differentiate anomalous from normal faces. Anomalous faces were judged to be less attractive than normal faces, as expected, and they did not differ in rated babyfaceness. Anomalous faces also were perceived as less sociable, less warm, less strong, less healthy, less dominant, less shrewd, and less intelligent than normal faces. They were also judged to be smiling less. However, the differences in trait impressions held true with ratings of smiling controlled (Table 1).

### Predicting Impressions of Young Adult Faces From Activation of the Anomaly Output Unit

Multiple-regression analyses determined whether impressions of normal young adult faces varying in attractiveness could be predicted from the extent to which they activated the anomaly output unit, controlling for sex of face (Table 2). Generalization faces eliciting greater activation of the anomaly unit were judged to be no different in smiling from those eliciting less activation and, as predicted, greater activation of the anomaly unit predicted ratings of lower attractiveness. Moreover, generalization faces that produced higher activation of the anomaly unit created impressions that paralleled those created by truly anomalous

**Table 1.** *Impressions of Anomalous and Normal Faces*

| | Training Face Group | | | | |
|---|---|---|---|---|---|
| | Anomalous | | Normal | | |
| Impression | *M* | *SD* | *M* | *SD* | F(1, 118) |
| Attractive | 2.10 | .67 | 4.15 | .62 | 303.37*** |
| Baby-faced | 3.72 | 1.15 | 3.61 | .83 | < 1 |
| Sociable | 3.20 | .78 | 4.49 | .86 | 75.81*** |
| Warm | 3.44 | .89 | 4.12 | .99 | 15.59*** |
| Healthy | 3.22 | .80 | 5.25 | .49 | 280.18*** |
| Physically strong | 3.35 | .88 | 4.74 | .53 | 108.06*** |
| Dominant | 3.56 | .91 | 4.49 | .71 | 38.44*** |
| Shrewd | 3.56 | .89 | 4.33 | .63 | 29.58*** |
| Intelligent | 3.13 | .74 | 4.60 | .47 | 168.15*** |

***$p < .001$

**Table 2.** *Impressions of Normal Adult Faces Predicted From Their Activation of Neural Network Output Units Trained to Respond to Inputs from Anomalous Faces*

| | Impression | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Predictor | Attractive β | Baby-faced β | Smiling β | Sociable β | Warm β | Healthy β | Strong β | Dominant β | Shrewd β | Intelligent β |
| Anomaly unit (controlling sex) | –.27* | .13 | –.14 | –.24* | –.25* | –.21+ | –.24* | –.15 | –.08 | –.21* |

+$p \leq .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

faces. More specifically, faces producing greater activation were perceived as less sociable, less warm, less strong, less healthy, and less intelligent than those producing less activation.

In a second regression analysis, we added actual trait scores, and we followed the procedure outlined by Baron and Kenny (1986) to determine whether they mediated the relation between anomaly unit activation and impressions. Actual health and IQ scores did not qualify as mediators of the corresponding impressions. We could not test for mediation of impressions of sociability, warmth, or strength by the corresponding actual traits because anomaly unit activation was no longer a significant predictor of these impressions in the reduced sample for which actual traits were available.[4]

## Discussion

Impressions of the traits of normal adults were predicted by the similarity of their appearance to that of people with genetic anomalies, as assessed by the extent to which a neural network confused their faces with anomalous ones. More specifically, the extent to which normal adult faces activated an output unit trained on metric inputs from anomalous faces predicted impressions of their attractiveness, sociability, warmth, health, strength, and intelligence—traits that also differentiated impressions of anomalous versus normal faces. The power of similarity to anomalous faces to predict trait impressions mirrors the attractiveness halo effect (Eagly et al., 1991; Feingold, 1992; Langlois et al., 2000; Zebrowitz, 1997). Whereas previous research has provided no satisfactory mechanism to explain impressions of faces that vary in attractiveness, these findings support the anomalous-face overgeneralization hypothesis by demonstrating that normal faces judged as unattractive physically resemble anomalous faces, and accurate impressions of unfit individuals as socially, physically, and cognitively deficient are overgeneralized to unattractive normal individuals. The finding that actual traits did not mediate the relation between anomaly unit activation and impressions of health or intelligence reinforces the argument that these impressions reflect an overgeneralization effect rather than accurate impressions of lower fitness in some faces that we had designated as normal.

## Experiment 2: Baby-Face Overgeneralization Hypothesis

We studied the extent to which impressions of young adults' faces varying in babyfaceness would be predicted from their activation of a neural network unit trained to respond to babies. The methodology was identical to that in Experiment 1 except as indicated in the following section.[5]

### Method

**Faces.** Training–test faces included 30 White infants, ranging in age from 5 to 9 months, drawn from previous research (Hidebrandt & Fitzgerald, 1974; Zebrowitz & Montepare, 1992) and 30 White adults (15 men and 15 women). Twenty-nine adult faces came from the IGS archive and previously had been rated in comparison with others of the same age and sex on a 7-point scale with endpoints labeled *baby faced/mature faced* (Zebrowitz, Olson, et al., 1993). They were in the top 30% in maturity ratings with a mean age of

[4]The only actual trait predicted by anomaly unit activation was IQ, ($\beta = -.32, p < .01$). However, IQ scores were unrelated to perceived intelligence ($\beta = .05, p = .65$), thus failing to satisfy the criteria for mediation. The latter finding is consistent with a lack of accuracy in judging IQ from faces at this age from the same data archive (Zebrowitz et al., 2002). The prediction of actual IQ scores from anomaly unit activation coupled with the failure of prediction from intelligence ratings indicates that the neural network proved sensitive to facial information that human observers missed. The additional finding that the prediction of perceived intelligence from anomaly unit activation did not change when controlling for actual IQ suggests that human judges' impressions were influenced by similarities to anomalous faces that were not diagnostic of actual intelligence.

[5]Unlike Experiment 1, there was no strong theoretical reason to expect actual traits to mediate the relation between network activation and impressions. Nevertheless, we did make an attempt to test for such mediation using measures of actual traits from the IGS archive. However, our efforts were unsuccessful because the effects of baby unit activation on impressions were not significant in the smaller subset of faces for which actual trait information was available.

17.4 years. Generalization faces included 80 White adults (40 women and 40 men), 73 of whom were selected from the IGS archive (mean age = 17.3 years). The exact ages of the remaining 7 faces were unknown, but all were high school or college students. Half of the faces of each sex had previously received facial maturity ratings in the top 30% compared with others of the same age and sex, and half had received ratings in the bottom 20%. The large majority of faces had neutral expressions.

**Facial ratings.** Eight judges of each sex rated one of two random orders of either male or female training–test faces (babies and mature-faced adults) or generalization faces (baby-faced and mature-faced adults). Three judges rated the degree of smiling of all faces.

**Facial metrics.** After reliability was established for points marked by two judges on a subset of 50 baby and adult faces, one judge marked the remaining faces, and those points were used in the final data analyses. Normalization adjusted for variations in distance from the camera and facial angle. All vertical distances were normalized by dividing them by the distance between the pupils (E2). All horizontal distances were computed by summing corresponding distances on the right and left side of the midline, with the right-side distances divided by the distance between the midline (Point 14) and the center of the right pupil, and the left-side distances divided by the distance between the midline and the center of the left pupil. This normalization method corrected for facial tilt away from the frontal plane. The inputs selected were 18 nonredundant facial metrics that achieved acceptable reliability, which included facial roundness plus the distances shown by underlined letters in Figure 2.

## Results

### Reliability of Measures

Interjudge agreement was significant for the normalization distance (E2), $r = .98$, as well as for the 18 input metrics, mean $r = .83$. Male and female judges showed strong agreement in their trait and appearance ratings, mean $r = .84$, and reliability was assessed across judges of both sexes. The average alpha reliability coefficient across ratings was .88 for both male and female training–test faces and .85 for both male and female generalization faces. The alpha coefficient for smile ratings across all faces was .74. Data analyses used mean ratings for each face across judges.

### Network Training and Generalization

Training a network to differentiate babies from adults was successful in meeting the criteria of at least 90% correct identification of the 40 training faces and at least 75% correct identification of the 20 test faces, averaged across 20 trials with the best solution requiring 4 hidden nodes, 3,000 training epochs per trial, a .03 learning rate, and a .2 error goal. Across the 20 trials, activation of the baby unit was significantly higher for babies ($M = 82.28$, $SD = 9.85$) than for adults ($M = 15.81$, $SD = 8.22$), $F(1, 59) = 805.00$, $p < .0001$. As expected, given that the generalization faces were all adults, they did not produce high activation of the baby unit ($M = 18.25$, $SD = 10.68$). Nevertheless, there was considerable variability in the extent to which these young adult faces activated the baby unit (range = 5.60 to 45.36).

### Impressions of Babies Versus Adults

The baby-face overgeneralization effect provides a potential explanation for impressions of adults on those traits that are judged to differentiate babies from adults. Babies were judged to be more attractive and more baby faced than young adults but no different in smiling. Babies were also perceived as warmer, physically weaker, more naive, more submissive, and less intelligent but no different in sociability or health (Table 3).

### Predicting Impressions of Young Adult Faces From Activation of the Baby Output Unit

Multiple regression analyses determined whether impressions of young adult faces varying in babyfaceness could be predicted from the extent to which they activated the baby output unit, controlling for sex of face (Table 4). As predicted, faces eliciting greater activation of the baby unit were perceived as

**Table 3.** *Impressions of Babies and Adults*

| | Training Face Group | | | | |
|---|---|---|---|---|---|
| | Babies | | Adults | | |
| Impression | *M* | *SD* | *M* | *SD* | F(1, 58) |
| Attractive | 4.38 | .84 | 3.54 | .94 | 13.35*** |
| Baby-faced | 5.37 | .68 | 2.80 | .70 | 207.10*** |
| Sociable | 4.40 | 1.00 | 3.96 | 1.37 | 2.10 |
| Warm | 4.60 | .72 | 3.52 | 1.05 | 21.48*** |
| Healthy | 4.65 | .90 | 4.68 | .92 | < 1 |
| Physically strong | 3.17 | .55 | 4.89 | .76 | 101.44*** |
| Dominant | 2.98 | .68 | 4.70 | .86 | 74.01*** |
| Shrewd | 2.76 | .66 | 4.75 | .62 | 145.78*** |
| Intelligent | 3.89 | .58 | 4.36 | .77 | 7.21** |

$^{**}p \le .01$. $^{***}p \le .001$.

**Table 4.** *Impressions of Young Adult Faces Predicted From Their Activation of a Neural Network Output Unit Trained to Respond to Inputs from Babies' Faces*

| | Impressions | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Predictor | Attractive β | Baby-face β | Smiling β | Sociable β | Warm β | Healthy β | Strong β | Dominant β | Shrewd β | Intelligent β |
| Baby unit (controlling sex) | .27* | .27* | –.05 | .19 | .26* | .17 | –.26* | –.30** | –.32** | .17 |

*$p \leq .05$. **$p \leq .01$.

more baby faced, warm, physically weak, naive, and submissive. Greater activation of the baby unit did not predict variations in smiling but did predict higher perceived attractiveness. However, the effects of baby unit activation on trait impressions did not weaken when attractiveness ratings were controlled, with the exception of impressions of warmth, which were reduced to a nonsignificant trend ($p = .12$).

## Discussion

Impressions of the personality traits of adults were predicted by the similarity between their appearance and that of babies, as assessed by the extent to which a neural network confused their faces with those of babies. The extent to which adult faces activated an output unit trained on inputs from the faces of babies predicted impressions of their babyfaceness, attractiveness, warmth, physical weakness, submissiveness, and naiveté. These were the very traits that differentiated impressions of babies versus adults and that have repeatedly been shown to differentiate baby-faced from mature-faced adults (Montepare & Zebrowitz, 1998; Zebrowitz, 1997). Moreover, impressions of traits that did not differentiate real babies from adults—sociability and health—were not higher for the generalization faces that produced higher activation of the baby unit. Although these impressions of baby-faced versus mature-faced adults have been demonstrated in previous research, the methodology employed in that research left open the possibility that the perceptions could derive from processes other than a baby-face overgeneralization effect. For example, semantic associations to wide-set eyes connote honesty independently of babyfaceness (Zebrowitz et al., 1996), and round faces are found in cultural icons that connote warmth, such as Santa Claus and the smiley face. Thus, trait impressions of large-eyed, round-faced adults could be explained by such cultural metaphors rather than by their resemblance to babies. Although these findings do not rule out the possibility of such contributions to trait impressions, they do provide unequivocal evidence that adults judged as baby faced physically resemble babies and that the degree of adults' physical resemblance to babies in and of itself makes a significant contribution to impressions of their warmth, physical strength, dominance, and shrewdness, thereby providing strong support for the baby-face overgeneralization hypothesis.

## Predicting Group Stereotypes From Face Overgeneralization Effects

Experiments 3 and 4 investigated whether stereotypes of common social categories could be explained by the extent to which faces in the category resembled babies' or anomalous faces. We first identified the stereotypes elicited by faces varying in age or weight, and we then tested for mediation of older person and weight stereotypes by resemblance to anomalous faces or babies' faces. The faces, input distances, and method employed in training and testing of the networks for these experiments were identical to that employed in Experiments 1 and 2. All that differed was the set of generalization faces. Facial distances from generalization faces varying in age or weight were entered into networks trained on anomalous versus normal faces or babies versus adults. In all cases, training successfully met the criteria of at least 90% correct identification of the training faces and at least 75% correct identification of the test faces.[6]

## Experiment 3: Elderly Stereotypes

A common picture emerging from research on elderly stereotypes is one of physical, cognitive, and interpersonal deficiency (for reviews, see Montepare & Zebrowitz, 2002; Nelson, 2001). In contrast to previous assumptions that such stereotypes have been spawned by youth-oriented Western cultures, recent

[6]An additional experiment determined that a neural network output unit trained to respond to faces of babies generalized that response to women's faces more than to men's faces, and the stereotypic impression of women as less dominant than men lost significance when controlling the extent to which the network confused faces of each sex with those of babies. Although this experiment is not reported in detail due to marginally significant results, it is consistent with previous evidence that reversing the link between sex and facial maturity in schematic faces reversed gender stereotypes on the power dimension (Friedman & Zebrowitz, 1992).

research has suggested a more universal origin. In particular, it appears that individuals in many Eastern cultures also endorse negative elderly stereotypes, with positive views of aging reflecting culturally mandated "oughts" rather than actual beliefs (Best & Williams, 1999; Koyano, 1989; Noesjirwan, Gault, & Crawford, 1983; Sharma, 1971). Consistent with the deficiencies attributed to elderly adults, recent work also has shown that elderly stereotypes are similar to stereotypes of people labeled as retarded or disabled (Cuddy & Fiske, 2002). Although research also has demonstrated some positive stereotypes of older adults (Andreoletti, Maurice, & Walen, 2001; Brewer, Dull, & Lui, 1981; Hummert, Garstka, Shaner, & Strahm, 1994; for reviews, see Cuddy & Fiske, 2002; Kite & Wagner, 2002; Montepare & Zebrowitz, 2002), these positive images are more likely to be associated with younger looking elderly adults (Hummert, 1994; Hummert, Garstka, & Shaner, 1997), perhaps because the older looking faces are perceived as less attractive, a perception manifested from early childhood through older adulthood (Downs & Walz, 1981; Henss, 1991; Johnson & Pittenger, 1984; Kogan, Stephens, & Shelton, 1961; Korthase & Trenholme, 1983; Zebrowitz, Olson, et al., 1993). The similarity of elderly stereotypes to stereotypes of people who are labeled as retarded or disabled, the cultural generality of these stereotypes, and the moderating effects of appearance all suggest that negative stereotypes of elderly adults may derive at least in part from an anomalous-face overgeneralization effect.

Although an anomalous-face overgeneralization effect may contribute to elderly stereotypes, so may a baby-face overgeneralization effect. Although it might seem like an oxymoron to call older adults baby faced, older faces are in fact more similar to babies in some ways than are young faces. Age-associated bone loss causes elderly people to have small jaws, double chins, and jowls, just as babies do. Moreover, research has shown that although some older person stereotypes parallel impressions of unattractive people, others parallel impressions of babies. Indeed, the characterization of eldelyr stereotypes as "doddering but dear" (Cuddy & Fiske, 2002) captures the incompetence and warmth that characterize impressions of babies. Thus, at least some stereotypes of elderly adults may be the by-product of a universal mechanism for recognizing infant faces.

## Method

**Faces.** The 80 generalization faces were 40 elderly (20 women and 20 men) and 40 young (20 men and 20 women) White adults. Most elderly faces were those of individuals living at a senior citizens' home. The ages were known for 26 of them ($M$ = 78.65, $SD$ = 6.47). Most young faces were taken from the IGS data archive. The ages were known for 38 of them ($M$ = 17.28, $SD$ = .51). Sixteen judges (8 men and 8 women) rated faces of elderly and young men or women.[7]

## Results

### Reliability of Measures

Reliability of the input metrics was established for a subset of 31 generalization faces. Interjudge agreement was significant for the normalization distances of interpupil distance (E2), $r$ = .99, and head length (LO), $r$ = .85, as well as for the input metrics, mean $r$ = .82.[8] Male and female judges agreed in their trait and appearance ratings, mean $r$ = .79, and reliability was assessed across judges of both sexes. The average alpha reliability coefficient across ratings was .87 for male faces and .86 for female faces. The alpha coefficient for smile ratings across all faces was .81. Data analyses used mean ratings for each face across all judges.

### Network Generalization

Mean activation was lower for the baby than the adult unit (i.e., < 50%) for both elderly and young adult faces, as would be expected given that the generalization faces were all adults (Table 5). Nevertheless, as predicted, inputs from elderly faces activated the baby unit more than did inputs from younger faces, and this held true when controlling for sex and smiling, which differed by age as described later, $F(1, 76) = 11.65$, $p$ = .001. Although elderly and young generalization faces were all normal adults, mean activation of the anomaly unit by inputs from elderly faces was *higher* than acti-

[7]When training the network and analyzing the overgeneralization effects, a different set of young faces was substituted for those originally rated with the older faces. Young faces in the original group were chosen to test the baby-face overgeneralization hypothesis, making sure that they did not appear as training faces in the baby-face network. We subsequently decided to investigate whether age stereotypes also could be explained by an anomalous-face overgeneralization effect. However, many of the young faces in the original generalization set could not be used because they served as normal training faces in the anomalous-face network. Because it was desirable to use the same young generalization faces in both networks, we substituted a different group of young faces that did not overlap with training faces in either one. Reliabilities for trait ratings are reported for the young and old faces that were rated together. Reliabilities for the young faces used as replacements were also high, as reported in Experiment 2, where they were also included among the generalization faces.

[8]Reliabilities of input metrics in Experiments 3 and 4 are reported for distances normed by head length (LO), which was used when testing the anomalous-face overgeneralization hypothesis. Reliabilities were somewhat higher for distances normed by interpupil distance (E2), which was used when testing the baby-face overgeneralization hypothesis because there was higher interjudge agreement for that norm.

**Table 5.** *Activation of the Anomaly Output Unit and the Baby Output Unit by Elderly and Young Faces*

| | Generalization Face Group | | | | | |
|---|---|---|---|---|---|---|
| | Elderly | | | Young | | |
| Output Unit Activation | *M* | *SD* | Range | *M* | *SD* | Range |
| Anomaly unit | 66.90 | 19.68 | 23.07 – 96.21 | 27.43 | 24.44 | 3.36 – 91.36 |
| Baby unit | 28.58 | 11.79 | 10.52 – 57.70 | 18.85 | 11.84 | 4.94 – 45.45 |

*Note:* Means are adjusted to control for sex of face and smiling.

vation of the normal unit (i.e., > 50%) with sex and smiling controlled, $F(1, 76) = 51.38$, $p < .001$.

### Predicting Elder Stereotypes From Activation of the Baby Output Unit or the Anomaly Output Unit

Despite our efforts to select faces that were equated in smiling, elderly faces were rated higher ($M = 2.32$, $SD = 1.01$) than were young faces ($M = 1.52$, $SD = .52$), as noted previously, $F(1, 77) = 19.97$, $p < .001$. Therefore, age, sex, and smiling were entered in the first step of the regression analyses to determine what traits were stereotypically attributed more to elderly faces, with sex and smiling controlled. Table 6 provides mean trait ratings of elderly and young adult faces, adjusted for smiling and sex, and Table 7 provides results of the regression analyses. Elderly faces were judged as less attractive, as predicted. Interestingly, despite their greater physical resemblance to the faces of babies, elderly faces also were judged as less baby faced than young faces. This probably reflects an influence on judged babyfaceness of youthfulness—nearness in age to babies—because judges were not explicitly instructed that babyfaceness was not the same thing as age, as they were in earlier research (e.g., Zebrowitz & Montepare, 1992). Elderly faces also were perceived as less sociable, less warm, less healthy, less strong, and shrewder.

**Table 6.** *Impressions of Elderly and Young Faces*

| | Generalization Face Group | | | | |
|---|---|---|---|---|---|
| | Elderly | | Young | | |
| Impression | *M* | *SD* | *M* | *SD* | *F*(1, 76) |
| Attractive | 2.70 | .81 | 3.67 | .85 | 26.39*** |
| Baby-faced | 2.40 | .61 | 3.86 | 1.26 | 34.86*** |
| Sociable | 3.51 | 1.07 | 4.25 | .93 | 19.71*** |
| Warm | 3.67 | 1.07 | 4.08 | .96 | 5.68* |
| Healthy | 3.17 | .75 | 4.48 | .72 | 58.75*** |
| Physically strong | 3.02 | .83 | 4.22 | 1.12 | 23.37*** |
| Dominant | 4.04 | 1.00 | 3.86 | .94 | < 1 |
| Shrewd | 4.60 | .80 | 3.99 | .83 | 9.38** |
| Intelligent | 4.29 | .75 | 4.26 | .61 | <1 |

*Note:* Means are adjusted to control for sex of face and smiling.
$^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

In a second block of the regression analyses, we added either activation of the anomaly unit (Block 2a) or activation of the baby unit (Block 2b), and we followed the procedure outlined by Baron and Kenny (1986) to determine whether either of these variables mediated age stereotypes. Anomaly unit activation satisfied the criteria for mediation of impressions of elderly faces as less sociable and warm, and the Sobel tests for mediation were significant, $z$s = 3.17 and 3.59, $p$s < .01 and < .001, respectively. Although elderly faces continued to be perceived as less attractive and less healthy when anomaly unit activation was controlled, there was significant partial mediation by similarity to anomalous faces for perceived attractiveness, $z = 2.60$, $p < .01$, and marginally significant partial mediation for perceived health, $z = 1.68$, $p = .09$. Also, an impression of elderly faces as more intelligent than younger ones emerged when activation of the anomaly unit was controlled, suggesting that resemblance of elderly faces to anomalous ones significantly suppressed impressions of their higher intelligence. On the other hand, stereotypes of elderly faces as weaker and shrewder were not mediated by their resemblance to anomalous faces. However, baby unit activation satisfied the criteria for mediation of impressions of elderly faces as physically weaker, and the Sobel test was marginally significant, $z = 1.68$, $p < .10$. Baby unit activation did not satisfy the criteria for mediation of any other age stereotypes.

## Discussion

Neural network output units trained to respond to anomalous faces or babies' faces each generalized their responses more to faces of elderly than young adults. Moreover, stereotypic impressions of elderly faces as unsociable, cold, unattractive, and unhealthy were at least partially mediated by their resemblance to anomalous faces. In addition, impressions of elderly faces as more intelligent emerged when controlling their resemblance to anomalous faces, suggesting a suppressor effect. These results are consistent with the hypothesis that an anomalous-face overgeneralization effect

**Table 7.** *Age Stereotypes Controlling for Activation of Neural Network Output Units Trained to Respond to Inputs from Anomalous Faces or Babies' Faces*

| | Impression | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Predictors | Attractive β | Baby-Faced β | Sociable β | Warm β | Healthy β | Strong β | Dominant β | Shrewd β | Intelligent β |
| Block 1 (controlling sex, smile) | | | | | | | | | |
| Age | –.56*** | –.62*** | –.37*** | –.20* | –.74*** | –.54*** | .10 | .37** | .02 |
| Block 2a (controlling sex, smile) | | | | | | | | | |
| 2a. Age | –.30* | –.74*** | –.17+ | .03 | –.58*** | –.50*** | .11 | .31* | .34* |
| 2a. Anomaly unit | –.36** | .16 | –.29** | –.32*** | –.23* | –.05 | –.02 | .08 | –.45*** |
| Block 2b (controlling sex, smile) | | | | | | | | | |
| 2b. Age | –.62*** | –.68*** | –.41*** | –.23** | –.72*** | –.45*** | .18 | .43*** | .00 |
| 2b. Baby unit | .16 | .16 | .11 | .08 | –.07 | –.23* | –.22+ | –.17 | .06 |

*Note:* Positive betas signify higher values for elderly faces. Blocks 2a –2c report the results of 3 separate regression analyses in which different mediators of older stereotypes were tested.
+$p \leq .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

diminishes impressions of the cognitive and interpersonal competence of elderly adults. However, resemblance to anomalous faces did not mediate stereotypes of elderly faces as less strong and more shrewd than young ones. The former impression, which was the only elderly stereotype that paralleled impressions of babies, was partially mediated by elderly faces having greater resemblance to babies' faces. This influence of resemblance to babies was clearly tacit because elderly adults were perceived as less, not more, baby faced than young adults. Finally, impressions of shrewdness, as well as those impressions that were only partially mediated by similarity to baby or anomalous faces, must reflect other mechanisms, such as social knowledge about changes in sagacity, strength, and health across the life span.

## Experiment 4: Overweight Stereotypes

There is a strong tendency to evaluate overweight individuals more negatively on a wide range of traits (Crandall, 1994; Ryckman, Robbins, Kaczor, & Gold, 1989). There also is evidence that some weight stereotypes may be produced by differences in attractiveness (Rothblum, Miller, & Garbutt, 1988), which suggests that an anomalous-face overgeneralization effect may contribute to overweight stereotypes. Research also has reported a tendency to rate overweight people relatively high on certain childlike traits, such as agreeable and dependent (Sleet, 1969; see DeJong & Kleck, 1986, for a review). Because some facial characteristics that differentiate babies from adults also differentiate overweight from normal-weight faces, such as double chins and jowls, it is also possible that a baby-face overgeneralization effect contributes to overweight stereotypes. However, despite parallels in stereotypes and appearance that support these overgeneralization hypotheses, there is also reason to believe that overweight stereotypes may reflect cultural influences rather than being the by-product of a universal mechanism for recognizing anomalous or infant faces. In particular, unlike stereotypes of unattractive or older adults, negative stereotypes of overweight people have varied across time and cultures, and there is evidence that they are related to particular social ideologies (Commons & Wilson, 1999; Crandall & Martinez, 1996; Cunningham, Roberts, Wu, Barbee, & Druen, 1995; Iwawaki & Lerner, 1974). If overweight stereotypes are in fact driven by culturally specific influences, then they should not be diminished when controlling the resemblance of faces that vary in weight to anomalous faces or babies' faces.

### Method

**Faces.** Generalization faces were 40 overweight (20 men and 20 women) and 40 normal-weight (20 women and 20 men) White young adults. Thirty-eight normal-weight (mean age = 17.3 years) and 15 overweight (mean age = 16.9 years) faces were selected from the IGS archive. Height and weight information, available for all but 7 normal-weight faces, was used to compute body mass index (BMI). Average BMI values were significantly higher for overweight faces ($M = 26.98$, $SD = 3.70$) than for normal-weight faces ($M = 20.87$, $SD = 2.33$), $F(1, 44) = 46.66$, $p < .0001$, and the mean value for overweight individuals exceeded the conventional criterion for determining overweight for individuals of this age, which is .25, the 85th percentile (Kuczmarski, Flegal, Campbell, & Johnson, 1994; Najjar & Rowland, 1987). The remaining 25 overweight faces were selected from photos posted on Web sites for overweight people. Judges rated one of two random orders of faces of overweight and normal-weight men or women. After rating the faces on all other scales, judges rated them

**Table 8.** *Activation of the Anomaly Output Unit and the Baby Output Unit by Overweight and Normal-Weight Generalization Faces*

| | Generalization Face Group | | | | | |
|---|---|---|---|---|---|---|
| | Overweight | | | Normal-Weight | | |
| **Output Unit Activation** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** |
| Anomaly unit | 31.50 | 24.76 | 1.91 – 80.71 | 28.27 | 22.50 | 3.88 – 94.12 |
| Baby unit | 28.58 | 9.72 | 12.73 – 52.19 | 21.30 | 9.00 | 10.38 – 45.72 |

*Note:* Means are adjusted to control for sex of face and smiling.

on a 7-point scale with endpoints labeled *not at all overweight/very overweight*.[9]

## Results

### Reliability of Measures

Reliability of the input metrics was established for a subset of 31 generalization faces. Interjudge agreement was significant for the normalization distances of interpupil distance (E2), $r = .99$, and head length (LO), $r = .72$. With the exception of chin width (W5), the other facial metrics also achieved significant interjudge agreement, mean $r = .83$. Despite low reliability, chin width was included as an input when testing the anomalous-face overgeneralization hypothesis to parallel the inputs used in Experiment 1, where it was reliable. Male and female judges agreed in their trait and appearance ratings, mean $r = .79$, and reliability was assessed across judges of both sexes. The average alpha reliability coefficient across ratings was .88 for male faces and .80 for female faces. Overweight ratings also showed high reliability with alphas of .98 and .97 for male and female faces, respectively, as did smile ratings, $\alpha = .93$. Data analyses used mean ratings for each face across judges.

### Network Generalization

Mean activation of the anomaly unit and the baby unit by the generalization faces was lower than activation of the normal face and adult face units (i.e., < 50%), as would be expected given that the generalization faces were all normal adults (Table 8). Nevertheless, as predicted, overweight faces activated the baby unit more than normal-weight faces did, and this held true when controlling sex and smiling, which differed by weight, $F(1, 76) = 7.84$, $p < .01$. Overweight and normal-weight faces did not differ significantly in anomaly unit activation, $F < 1$.

### Predicting Overweight Stereotypes From Activation of the Baby Output Unit or the Anomaly Output Unit

Despite our efforts to select faces equated in smiling, overweight faces were rated higher ($M = 3.42$, $SD = 1.87$) than normal-weight faces ($M = 1.52$, $SD = .52$), as noted previously, $F(1, 78) = 38.44$, $p < .001$. Therefore, weight, sex, and smiling were entered in the first block of the regression analyses to determine what traits were stereotypically attributed more to overweight individuals, with sex and smiling controlled. Table 9 provides mean trait ratings of overweight and normal-weight faces, adjusted for smiling and sex, and Table 10 provides results of the regression analyses. Overweight faces were perceived as less attractive, less sociable, less healthy, and less intelligent.

In a second block of the regression analyses, we entered either activation of the anomaly unit (Block 2a) or activation of the baby unit (Block 2b), and we followed the procedure outlined by Baron and Kenny (1986) to determine whether either of these variables

**Table 9.** *Impressions of Overweight and Normal-Weight Faces*

| | Generalization Face Group | | | | |
|---|---|---|---|---|---|
| | Overweight | | Normal-Weight | | |
| **Impression** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***F*(1, 76)** |
| Attractive | 2.93 | .75 | 3.54 | .85 | 8.02** |
| Baby-faced | 4.09 | .99 | 3.88 | 1.26 | <1 |
| Sociable | 3.71 | 1.02 | 4.35 | .93 | 11.71*** |
| Warm | 3.94 | 1.03 | 4.20 | .96 | 2.04 |
| Healthy | 3.66 | .78 | 4.36 | .72 | 11.66*** |
| Physically strong | 4.08 | .69 | 4.13 | 1.12 | <1 |
| Dominant | 3.69 | .65 | 3.96 | .94 | 1.56 |
| Shrewd | 3.84 | .68 | 4.04 | .83 | <1 |
| Intelligent | 3.67 | .63 | 4.35 | .61 | 18.89*** |

*Note:* Means are adjusted to control for sex of face and smiling.
$^{**}p \le .01$; $^{***}p \le .001$.

[9]Overweight faces were rated as significantly more overweight (M = 4.95) than normal-weight faces ($M = 2.11$), $F(1, 78) = 243.20$, $p < .0001$. Although judges rated overweight faces that were randomly interspersed with normal-weight faces, those particular normal-weight faces were replaced with a different group when calculating BMI differences, training the network, and analyzing overgeneralization effects. A different set of normal-weight faces was substituted for those originally rated with the overweight faces for the same reasons as in Experiment 3 (see Footnote 7). The normal-weight generalization faces in Experiment 4 were identical to the young generalization faces in Experiment 3.

**Table 10.** *Weight Stereotypes Controlling for Activation of Neural Network Nodes Trained to Respond to Inputs From Anomalous Faces or Babies' Faces or Controlling for Attractiveness*

| | Impression | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Predictors** | **Attractive** β | **Babyface** β | **Smiling** β | **Sociable** β | **Warm** β | **Healthy** β | **Strong** β | **Dominant** β | **Shrewd** β | **Intelligent** β |
| Block 1(controlling sex, smile) | | | | | | | | | | |
| Weight | –.37*** | .09 | — | –.33*** | –.12 | –.43*** | –.03 | –.17 | –.13 | –.54*** |
| Block 2a (controlling sex, smile) | | | | | | | | | | |
| 2a. Weight | –.39** | .10 | — | –.35*** | –.14 | –.45*** | –.03 | –.17 | –.12 | –.55*** |
| 2a. Anomaly unit | –.34*** | .14 | — | –.22** | –.17** | –.20** | –.01 | –.01 | .05 | –.22** |
| Block 2b (controlling sex, smile) | | | | | | | | | | |
| 2b. Weight | .38** | .04 | — | –.33** | –.13 | –.38** | .08 | –.10 | –.09 | –.59*** |
| 2b. Baby unit | .02 | .15 | — | –.01 | .01 | –.15 | –.30** | –.19 | –.01 | –.14 |
| Block 2c (controlling sex, smile) | | | | | | | | | | |
| 2c. Weight | — | .03 | — | –.17* | –.05 | –.16+ | .05 | –.13 | –.07 | –.33** |
| 2c. Attractiveness | — | –.17 | — | .43*** | .21** | .73*** | .22+ | .10 | .16 | .57*** |

*Note:* Positive betas signify higher values for overweight faces. Blocks 2a-2c report the results of 3 separate regression analyses in which different mediators of weight stereotypes were tested.
+$p \leq .10$; *$p \leq .05$; **$p \leq .01$; ***$p \leq .001$.

mediated weight stereotypes. Neither one met the criteria for mediation. On the other hand, when perceived attractiveness was entered (Block 2c), we found that it did satisfy the criteria. The stereotype of overweight faces as unhealthy was reduced to marginal significance, and the Sobel test was significant, $z = 2.72$, $p < .01$. The magnitude of the stereotypes of overweight faces as unsociable and unintelligent was also reduced, showing significant partial mediation by attractiveness, respective *z*s = 2.56 and 2.57, both *p*s < .01.

## Discussion

Resemblance to babies' or anomalous faces failed to predict stereotypes of overweight adults. Although overweight faces were confused more with babies by the neural network, as predicted, these faces were not rated as more baby faced, indicating that the neural network detected a resemblance that human judges did not. In addition, the traits attributed to overweight faces did not include the warmth, submissiveness, naïveté, and weakness associated with babies and baby-faced adults. On the other hand, impressions of overweight faces as less sociable, healthy, and intelligent did parallel impressions of anomalous faces. Yet, these overweight stereotypes were not mediated by the extent to which normal and overweight faces activated the anomalous-face unit, and overweight faces were not confused more with anomalous ones by the neural network.

The results of Experiment 4 indicate that physical resemblance to a baby's face, which can be detected by a neural network, is not a sufficient condition for the perception of babyfaceness and impressions of childlike traits. The network may be more sensitive than humans to facial similarities between babies and overweight faces, or it may weight the similarities and dissimilarities differently from humans, or both factors may be operating. The results of Experiment 4 also indicate that physical resemblance to an anomalous face is not a necessary condition for the perception of unattractiveness and halo-effect trait impressions. It thus appears that perceived unattractiveness is not isomorphic with similarity to anomalous faces. Rather, there seem to be various forms of perceived unattractiveness, one of which is associated with overweight faces and another that is associated with anomalous, normal unattractive, or elderly faces. This finding demonstrates the discriminant validity of the anomaly unit activation. If activation were merely a proxy for rated attractiveness, then overweight faces should activate these units more than those of normal weight. On the other hand, if activation reflects a more specific resemblance to unfit faces that can explain only universal face stereotypes, then it is less surprising that overweight faces did not produce higher activation. Indeed, evidence for the cultural and historical variability in reactions to overweight suggests that the lower attractiveness of overweight faces and the accompanying negative stereotypes are socially driven rather than reflecting some universal mechanism such as the face overgeneralization effects (Commons & Wilson, 1999; Crandall & Martinez, 1996; Cunningham et al., 1995; Iwawaki & Lerner, 1974).

## General Discussion

Despite aphorisms, such as "Don't judge a book by its cover," "Pretty is as pretty does," and "Beauty is only skin deep," the proclivity for reading faces prevails across the lifespan, across cultures, and across historical eras (Zebrowitz, 1997). Appearance qualities, such as attractiveness and babyfaceness, have a profound effect on first impressions as well as on outcomes in important life arenas, despite evidence that contradicts the baby-face stereotype (Montepare & Zebrowitz, 1998; Zebrowitz & Lee, 1999), and only weak evidence for any accuracy to the attractiveness stereotype (Feingold, 1992; Langlois et al., 2000; Zebrowitz et al., 2002). Some universal process with strong adaptive value seems necessary to explain the ubiquity of reading faces. The face overgeneralization hypotheses provide an explanation for this phenomenon, and the results of these experiments offer strong supporting evidence.

Consistent with the anomalous-face overgeneralization hypothesis, a neural network, which knows only facial metrics and has no social knowledge, responds to unattractive or elderly faces similarly to anomalous ones. Moreover, human judges' impressions of the sociability, warmth, health, and intelligence of young adult faces that vary in attractiveness were predicted from similarities in the network's response to these faces and anomalous ones. Similarly, stereotypes of the sociability, warmth, health, and intelligence of elderly versus young adult faces were mediated in part by similarities in the network's response to these faces and anomalous ones. Consistent with the baby-face overgeneralization hypothesis, a neural network unit trained to respond to babies generalized that response more to baby-faced than to mature-faced adult faces and more to faces of elderly than young adults. Moreover, stereotypes of the strength of elderly versus young adult faces and impressions of the warmth, strength, dominance, and shrewdness of young adult faces that vary in babyfaceness were predicted from similarities in the network's response to these faces and those of babies.

The evidence that both stereotypes of elderly adults and the attractiveness halo effect in impressions of normal young adults were predicted by the extent to which their facial metrics resembled those of anomalous faces is consistent with other evidence that the cognitive decision making in attractiveness ratings may be best simulated by a strategy of "avoid the worst trait" rather than "use the best trait" (Grammer et al., 2002). It is also consistent with evidence for a greater impact of negative than positive attributes on impression formation, social attraction and sexual desire, a greater consensus in identifying rejected individuals than popular ones, and stronger neurological responses to negative than to positive stimuli (Baumeister et al., 2001; Kanouse & Hanson, 1972; Rozin & Royzman, 2001). Finally, this evidence is consistent with the ecological theory assumption that social perception serves an adaptive function (Gibson, 1979; McArthur & Baron, 1983). In the case of face perception, overgeneralization effects derive from the adaptive value of detecting and responding appropriately to particular facial qualities, including those that reveal fitness. The evolutionary importance of appropriate responses to unfit individuals is presumed to produce a strong preparedness to make such responses that is overgeneralized to individuals whose appearance merely resembles those who are unfit. Although overgeneralization may not seem particularly adaptive, one can argue that the errors that result from responding to unattractive people as if they are unfit are less maladaptive than failing to respond appropriately to those who are indeed unfit.[10]

Neither a baby-face nor an anomalous-face overgeneralization effect predicted stereotyped impressions of overweight adult faces. Even though the network unit trained to respond to babies was activated more by overweight than normal-weight faces, overweight stereotypes were not mediated by this greater activation. And, even though some overweight stereotypes paralleled impressions of anomalous and ordinary unattractive faces, overweight faces did not produce greater activation of a network unit trained to respond to anomalous faces. Yet, stereotypes of overweight versus normal-weight faces were mediated by variations in their perceived attractiveness. The finding that attractiveness judgments could be independent of similarity to anomalous faces is consistent with evidence that judgments of attractiveness can be influenced by many factors, including social learning and cultural values (Zebrowitz & Rhodes, 2002). It appears that stereotypes of overweight individuals are grounded more in the socially constructed meaning of

[10]The overgeneralization hypotheses and these findings can be related to the concept of *psychological essentialism,* which holds that people assume things that look alike tend to share deeper similarities (cf. Medin, 1989; Medin & Ortony, 1989). They also are consistent with the argument that perceptual systems have evolved such that this essentialism is often correct inasmuch as similarities in surface structure often reflect underlying functional similarities (Medin, 1989; Medin & Ortony, 1989; Shepard, 1984). Applying the concept of psychological essentialism to the domain of social perception, Yzerbyt and his colleagues (e.g., Yzerbyt, Rocher, & Schadron, 1997) suggested that group stereotypes are strongly related to perceivers' beliefs about the underlying essence shared by all group members. Our findings show that certain social stereotypes can reflect correct assumptions about the inherent essence of babies or those who are genetically unfit. However, this paradigm differs from that espoused by Yzerbyt and colleagues in at least two ways. First, it views stereotyping as a continuum rather than a categorical effect. Babyfaceness and attractiveness are continuous rather than categorical variables and stereotyped impressions vary with the degree of a person's resemblance to babies' or anomalous faces. Similarly, elderly faces that show less resemblance to anomalous faces will be perceived less stereotypically even if they are categorized as older persons. Second, within this formulation, essentialist assumptions are viewed as tacit mediators of stereotypes rather than as explicit components. Indeed, the greater resemblance of elderly faces to babies' faces predicted impressions of their lesser strength even though they were not explicitly judged as more babyfaced. Even when perceived babyfaceness parallels trait impressions, people are typically unaware that their impressions are driven by the babyishness of facial features (e.g., McArthur & Apatow, 1983).

their appearance than in their resemblance to adaptively significant categories of faces (Crandall & Martinez, 1996).

The failure to predict impressions of overweight faces from their resemblance to anomalous ones sharpens the meaning of the network activation in Experiments 1 and 3. If the network were merely an attractiveness detector rather than an anomaly detector, then one would expect overweight generalization faces to produce greater activation than normal-weight faces because the former were rated as significantly less attractive. Thus, the network cannot be construed as trained to detect any unattractive faces; rather it detects resemblance to anomalous faces. Overweight faces do not bear such a resemblance, and trait impressions must be explained by other mechanisms. On the other hand, ordinary unattractive faces and elderly faces do bear such a resemblance, and consistent with the anomalous-face overgeneralization hypothesis, this predicts impressions of their traits.

An interesting question is whether the impressions predicted from network activation could be construed as accurate perceptions rather than overgeneralizations. In particular, the network that was trained to differentiate anomalous from normal faces may have subsequently detected subtle yet diagnostic anomalies in some of the generalization faces that we classified as normal. Consistent with this possibility is the fact that some faces that may be categorized as normal by human judges manifest minor physical anomalies that are markers of psychological traits, including learning disabilities, low IQ, and hostile, competitive personalities (Bell & Waldrop, 1982; Cummings et al., 1982; Krouse & Kauffman, 1982; Paulhus & Martin, 1986; Streissguth, Herman, & Smith, 1978; Thornhill & Møller, 1997; Waldrop & Halverson, 1972). If this were true for the generalization faces used in Experiment 1, then the impressions predicted from network activation could be accurate rather than overgeneralizations. However, our results do not support this possibility. Although we found no evidence that impressions of health and intelligence were accurate, the accuracy of other impressions remains an open question because we were unable to conduct satisfactory tests due to missing data for the actual scores. Also, it is certainly possible that there is some truth to the impressions of elderly adults as less healthy than young adults, an impression that was partially mediated by anomaly unit activation in Experiment 3. However, there is no obvious reason to believe that elderly adults are less warm and sociable—impressions that are predicted by the overgeneralization hypothesis.

Some additional caveats to these findings should be noted. First, although activation of the neural network output units yielded significant prediction of trait impressions in Experiments 1 to 3, the effect sizes were modest. This suggests that factors other than resemblance to anomalous faces or babies' faces make a significant contribution to trait impressions, or that limitations of the facial distance inputs or neural network architecture attenuated the predictive power of network activation, or it could be a combination of both. Second, our results rely on a particular choice of network architecture (three layers, feed-forward net), learning mechanism (back-propagation), and knowledge representation (facial metrics information). These choices could be thought of as an implementation of "innateness" in that they intrinsically influence the results we have obtained (Elman et al., 1996). A different architecture with feedback connections, a non-Hebbian learning paradigm, and the inclusion of facial texture information may (or may not) change the results we presented here. Using different techniques, it may be possible to relax these choices and let the network decide what architecture, learning mechanisms, and knowledge representation are the best for the classification tasks at hand. However, it is not clear that such a general approach is possible and, even if it were, whether such an approach would not be just a more complex algorithm suffering from the same shortcomings as the simpler one we have used. We provide here a starting point from which future studies can be conducted.

A third caveat regarding these findings is that although they are consistent with the overgeneralization hypotheses, they do not prove that overgeneralization actually does occur. It is true that the network overgeneralizes responses from anomalous faces to unattractive faces and elderly faces and that these network responses predict human judges' impressions of those faces. Similarly, it is true that the network overgeneralizes responses from babies' faces to baby-faced adults' and elderly faces and that these network responses predict human judges' impressions. However, the connectionist models are formal simulations that may or may not capture any neural reality. Indeed, we do not presume that the neural network architecture maps actual neural information processing schemes, which are unknown at this point. Nevertheless, it would be instructive to apply functional magnetic resonance imaging (fMRI) techniques to determine whether the faces that were shown to be similar in these simulations also show distinctive similarities in actual neural activation patterns. It also would be useful to investigate what particular facial measurements and higher order, configural qualities contribute to a network's confusion of unattractive or elderly faces with anomalous ones and the confusion of baby-faced adults with babies. Although a search for neural bases of the overgeneralization effects may seem antithetical to Gibson's (1979) explicit renunciation of internal representations, it should be noted that his emphasis on higher order invariants was prescient, and it stimulated research that has identified neurons tuned to such information (Nakayama, 1994).

Although these experiments focused on the anomalous-face and the baby-face overgeneralization hypotheses, there are other face overgeneralization hypotheses that merit investigation (Zebrowitz, 1996, 1997). One of these is the facial-identity overgeneralization hypothesis. The evolutionary and social importance of differentiating known individuals from strangers and being wary of the latter may have produced a tendency for responses to strangers to vary as a function of their facial resemblance to known individuals. Indeed, some research has shown that reactions to people do depend on their facial resemblance to known others (Hill, Lewicki, Czyzewska, & Schuller, 1990; Lewicki, 1985; Secord & Jourard, 1956; cf. Andersen & Berk, 1998). In addition to explaining idiosyncratic impressions, identity overgeneralization may also contribute to negative impressions of other race faces, which show less resemblance to known individuals. The emotional-face overgeneralization hypothesis is also worthy of study. The evolutionary and social importance of detecting emotion in the face may have produced a strong preparedness to respond to the facial qualities revealing particular emotions that is overgeneralized to individuals whose facial structure merely resembles a particular emotion. Thus, people may be perceived to have those psychological traits associated with the emotional expressions that their facial features resemble, reflecting the adaptive value of responding to emotional expressions, such as avoiding an angry person and approaching a happy one (Knutson, 1996). Hopefully the face overgeneralization hypothesis and our findings will stimulate additional research designed to unmask the face, so we can understand why facial structure conveys so many qualities despite our best efforts to ignore it.

## References

Andersen, S. M., & Berk, M. S. (1998). The social-cognitive model of transference: Experiencing past relationships in the present. *Psychological Science, 7,* 109–115.

Andreoletti, C., Maurice, J. K., & Walen, H. (2001, February). *Gender, race, and age: Compound stereotypes across the lifespan.* Poster session presented at the annual meeting of the Society for Personality and Social Psychology, San Antonio, TX.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bartlett, M. S. (2001). *Face image analysis by unsupervised learning.* Boston: Kluwer Academic.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323–370.

Bell, R. Q., & Waldrop, M. F. (1982). Temperament and minor physical anomalies. In R. Porter & G. M. Collins (Eds.), *Temperamental differences in infants and young children: CIBA Symposium No. 89* (pp. 206–220). London: Pitman.

Berry, D. S. (2000). Attractiveness, attraction, and sexual selection: Evolutionary perspectives on the form and function of physical attractiveness. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 273–342). San Diego, CA: Academic.

Best, D. L., & Williams, J. E. (1999). Anticipation of aging: A cross-cultural examination of young adults' views of growing old. In J. Pandey, D. Sinha, & D. P. B. Bhawuk (Eds.), *Asian contributions to cross-cultural psychology* (pp. 274–288). Thousand Oaks, CA: Sage Publications.

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology, 41,* 656–670.

Campbell, M., Geller, B., Small, A. M., Petti, T. A., & Ferris, S. H. (1978). Minor physical anomalies in young psychotic children. *American Journal of Psychiatry, 135,* 573–575.

Commons, M., & Wilson, M. (1999, October). *Fat prejudice as a determinant of African American and White men's preferences in a potential romantic partner.* Paper presented at the meeting of the Society of Experimental Social Psychology, St. Louis, MO.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? The view from UCSD. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 285–345). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology, 66,* 882–894.

Crandall, C. S., & Martinez, R. (1996). Culture, ideology, and antifat attitudes. *Personality and Social Psychology Bulletin, 22,* 1165–1176.

Cuddy, A. J. C., & Fiske, S. T. (2002). Doddering but dear: Process, content, and function in stereotyping of older persons. In T. D. Nelson (Ed.), *Ageism: Stereotyping and prejudice against older persons* (pp. 3–26). Cambridge, MA: MIT Press.

Cummings, C. D., Flynn, D., & Preus, M. (1982). Increased morphological variants in children with learning disabilities. *Journal of Autism and Developmental Disorders, 12,* 373–383.

Cunningham, M. R., Barbee, A. P., & Philhower, C. L. (2002). Dimensions of facial physical attractiveness: The intersection of biology and culture. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 193–238). Westport, CT: Ablex.

Cunningham, M. R., Roberts, A. R., Wu, C., Barbee, A. P., & Druen, P. B. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology, 68,* 261–279.

DeJong, W., & Kleck, R. E. (1986). The social psychological effects of over-weight. In C. P. Herman, M. P. Zanna, & E. T. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario Symposium* (Vol. 3, pp. 65–87). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dion, K. K. (1973). Young children's stereotyping of facial attractiveness. *Developmental Psychology, 9,* 183–188.

Dion, K. K. (2002). Cultural perspectives on facial attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 239–260). Westport, CT: Ablex.

Downs, A. C., & Walz, P. J. (1981). Sex differences in preschoolers' perceptions of young, middle-aged, and elderly adults. *Journal of Psychology, 109*, 119–122.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). "What is beautiful is good, but … ": A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109–128.

Elman, J. L., Bates, E. A., Johnson M. H., Karminoff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A* *connectionist perspective on development.* Cambridge, MA: MIT Press.

Endler, J. A., & Basolo, A. L. (1998). Sensory ecology, receiver biases and sexual selection. *Trends in Ecology & Evolution, 13,* 415–420.

Enlow, D. H. (1990). *Facial growth* (3rd ed.). Philadelphia: Harcourt Brace.

Enquist, M., & Arak, A. (1994, November 10). Symmetry, beauty and evolution. *Nature, 372,* 169–172.

Enquist, M., Ghirlanda, S., Lundqvist, D., & Wachtmeister, C. A. (2002). An ethological theory of attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 127–151). Westport, CT: Ablex.

Enquist, M., & Johnstone, R. A. (1997). Generalization and the evolution of symmetry preferences. *Proceedings of the Royal Society of London: Series B, 264,* 1345–1348.

Feingold, A. (1992). Good looking people are not what we think. *Psychological Bulletin, 111,* 304–341.

Fellous, J. M. (1996). *A neural code for face representation: From V1 receptive fields to IT "face cells."* Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Fellous, J. M. (1997). Gender discrimination and prediction on the basis of facial metric information. *Vision Research, 37,* 1961–1973.

Friedman, H., & Zebrowitz, L. A. (1992). The contribution of typical sex differences in facial maturity to sex-role stereotypes. *Personality and Social Psychology Bulletin, 18,* 430–438.

Gangestad, S. W., & Buss, D. M. (1993). Pathogen prevalence and human mate preferences. *Ethology and Sociobiology, 14,* 89–96.

Gibson, J. J. (1966). *The senses considered as perceptual system.* Boston: Houghton Mifflin Company.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1991). Sexnet: A neural network identifies sex from human faces. In R. P. Lippman, J. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems.* (Vol. 3, pp. 572–577). San Mateo, CA: Kaufmann.

Grammer, K., Fink, B., Juette, A., Ronzal, G., & Thornhill, R. (2002). Female faces and bodies: N-dimensional feature space and attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 91–125). Westport, CT: Ablex.

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology, 108,* 233–242.

Guy, J. D., Majorski, L. V., Wallace, C. J., & Guy, M. P. (1983). The incidence of minor physical anomalies in adult male schizophrenics. *Schizophrenia Bulletin, 9,* 571–582.

Halberstadt, J., & Rhodes, G. (2000). The attractiveness of non-face averageness: Implications for an evolutionary explanation of the attractiveness of average faces. *Psychological Science, 11,* 285–289.

Hazan, C., & Diamond, L. M. (2000). The place of attachment in human mating. *Review of General Psychology, 4,* 186–204.

Henss, R. (1991). Perceiving age and attractiveness in facial photographs. *Journal of Applied Social Psychology, 21,* 933–946.

Hildebrandt, K. A., & Fitzgerald, H. E. (1979). Facial feature determinants of perceived infant attractiveness. *Infant Behavior and Development, 2*, 329–339.

Hill, T., Lewicki, M., Czyzewska, M., & Schuller, G. (1990). The role of learned inferential encoding rules in the perception of faces: Effects of nonconscious self-perpetuation of a bias. *Journal of Experimental Social Psychology, 26,* 350–371.

Hummert, M. L. (1994). Physiognomic cues and activation of stereotypes of the elderly in interaction. *International Journal of Aging and Human Development, 39,* 5–20.

Hummert, M. L., Garstka, T. A., & Shaner, J. L. (1997). Stereotyping of older adults: The role of target facial cues and perceiver characteristics. *Psychology and Aging, 12,* 107–114.

Hummert, M. L., Garstka, T. A., Shaner, J. L., & Strahm, S. (1994). Stereotypes of the elderly held by young, middle-aged, and elderly adults, *Journal of Gerontology, 49*(5), 240–249.

Iwawaki, S., & Lerner, R. M. (1974). Cross-cultural analyses of body-behavior relations: I. A comparison of body build stereotypes of Japanese and American males and females. *Psychologica, 17,* 75–81.

Johnson, D. F., & Pittenger, J. B. (1984). Attribution, the attractiveness stereotype, and the elderly. *Developmental Psychology, 20,* 1168–1172.

Johnstone, R. A. (1994, November 10). Female preference for symmetrical males as a by-product of mate recognition. *Nature, 372,* 172–175.

Kaiser, S., & Wehrle, T. (1992). Automated coding of facial behavior in human-computer interactions with FACS. *Journal of Nonverbal Behavior, 16,* 67–84.

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., & Johnson, R. M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science, 9,* 8–13.

Kanouse, D. E., & Hanson, L. R. (1972). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 47–62). Morristown, NJ: General Learning Press.

Keating, C. F. (2002). Charismatic faces: Social status cues put face appeal in context. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 153–192). Westport, CT: Ablex.

Keating, C. F., & Bai, D. L. (1986). Children's attributions of social dominance from facial cues. *Child Development, 57,* 1269–1276.

Kite, M. E., & Wagner, L. S. (2002). Attitudes toward older adults. In T. D. Nelson (Ed.), *Ageism: Stereotyping and prejudice against older persons* (pp. 129–161). Cambridge, MA: MIT Press.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20*, 165–182.

Kogan, N., Stephens, J. W., & Shelton, F. C. (1961). Age differences: A developmental study of discriminability and affective response. *Journal of Abnormal and Social Psychology, 62,* 221–230.

Korthase, K. M., & Trenholme, I. (1983). Children's perceptions of age and physical attractiveness. *Perceptual and Motor Skills, 56,* 895–900.

Koyano, W. (1989). Japanese attitudes toward the elderly: A review of research findings. *Journal of Cross-Cultural Gerontology, 4*, 335–345.

Kramer, S., Zebrowitz, L. A., San Giovanni, J. P., & Sherak, B. (1995). Infants' preferences for attractiveness and babyfaceness. In B. G. Bardy, R. J. Bootsma, & Y. Guiard (Eds.), *Studies in perception and action III* (pp. 389–392). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Krouse, J. P., & Kauffman, J. M. (1982). Minor physical anomalies in exceptional children: A review and critique of research. *Journal of Abnormal Child Psychology, 10,* 247–264.

Kuczmarski, R. J., Flegal, K. M., Campbell, S. M., & Johnson, C. L. (1994). Increasing prevalence of over-weight among U.S. adults: The national health and nutrition examination surveys 1960–91. *Journal of the American Medical Association, 272,* 205–211.

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 127*, 187–208.

Langlois, J. H., Kalakanis, L. E., Rubenstein, A. J., Larson, A. D., Hallam, M. J., & Smoot, M. T. (2000). Maxims and myths of beauty: A meta-analytic and theoretical review. *Psychological Bulletin, 126,* 390–423.

Langlois, J. H., Ritter, J. M., Roggman, L. A., & Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology, 27,* 79–84.

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

Levy, B. R. (1999). The inner self of the Japanese elderly: A defense against negative stereotypes of aging. *International Journal of Aging & Human Development,* 48, 131–144.

Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology, 48,* 563–574.

Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21,* 1357–1362.

McArthur, L. Z., & Baron, R. A. (1983). An ecological approach to social perception. *Psychological Review, 90,* 215–238.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469–1481.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadouo & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

Mignault, A., & Marley, A. A. J. (2001). *Holistic connectionist models of the perception of facial expressions of emotion.* Unpublished manuscript.

Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 93–163). New York: Academic.

Montepare, J. M., & Zebrowitz, L. A. (2002). A social-developmental view of ageism. In T. D. Nelson (Ed.), *Ageism: Stereotyping and prejudice against older persons* (pp. 77–125). Cambridge, MA: MIT Press.

Montepare, J. M., & Zebrowitz-McArthur, L. A. (1989). Children's perceptions of babyfaced adults. *Perceptual and Motor Skills, 69,* 467–472.

Najjar, M. F., & Rowland, M. (1987). *Anthropometric reference data and prevalence of over-weight: United States, 1976–1980* (Vital & Health Statistics, Series 11, No. 238). Hyattsville, MD: U.S. Department of Health and Human Services.

Nakayama, K. (1994). James J. Gibson—An appreciation. *Pyschological Review, 101,* 329–335.

Nelson, T. D. (2001). Ageism: A critical review of current research and theory. In T. Nelson (Ed.), *The psychology of prejudice. Boston: Allyn & Bacon.*

McArthur, L. Z., & Apatow, K. (1983–84). Impressions of babyfaced adults. *Social Cognition*, *2*, 315–342.

Neuberg, S. L., Smith, D. M., & Asher, T. (2000). Why people stigmatize: Toward a biocultural framework. In T. Heatherton, R. Kleck, J. G. Hull, & M. Hebl (Eds.), *The social psychology of stigma* (pp. 31–61). New York: Guilford.

Noesjirwan, J., Gault, U., & Crawford, J. (1983). Beliefs about memory in the aged. *Journal of Cross-Cultural Psychology, 14*, 455–468.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Bartlett, J. C. (1991). Classifying faces by race and sex using an auto-associative memory trained for recognition. In K. J. Hammond & D. Gentner (Eds.), *Proceedings of the 13th Annual Conference of the Cognitive Sciences Society* (pp. 847–851). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Toole, A. J., Price, T., Vetter, T., Bartlett, J. C., & Blanz, V. (1999). 3D shape and 2D surface textures of human faces: The role of "averages" in attractiveness and age. *Image and Vision Computing, 18,* 9–19.

Paulhus, D. L., & Martin, C. L. (1986). Predicting adult temperament from minor physical anomalies. *Journal of Personality and Social Psychology, 50,* 1235–1239.

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behaviour, 20,* 295–307.

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review, 5,* 659–669.

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration and facial attractiveness. *Psychological Science, 7,* 105–110.

Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution and Human Behavior, 22,* 31–46.

Rothblum, E., Miller, C. T., & Garbutt, B. (1988). Stereotypes of obese female job applicants. *International Journal of Eating Disorders, 7,* 277–283.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5,* 296–320.

Rubenstein, A. J., Langlois, J. H., & Roggman, L. A. (2002). What makes a face attractive and why: The role of facial averageness in defining facial beauty. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 1–33). Westport, CT: Ablex.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J.L. McClelland, D.E. Rumelhart & the PDP Research Group (Eds.). *Parallel distributed processing (*Vol 2, pp. 7-57). Cambridge, MA: MIT Press.

Ryckman, R. M., Robbins, M. A., Kaczor, L. M., & Gold, J. A. (1989). Male and female raters' stereotyping of male and female physiques. *Personality and Social Psychology Bulletin, 15,* 244–251.

Samal, A., & Iyengar, P. A., (1992). Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition, 25*, 65–77.

Secord, P., & Jourard, S. M. (1956). Mother-concepts and judgments of young women's faces. *Journal of Abnormal and Social Psychology, 52,* 246–250.

Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior, 20,* 71–76.

Sharma, K. L. (1971). A cross cultural comparison of stereotypes towards older persons. *Indian Journal of Social Work, 32,* 315–320.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91,* 417–447.

Sleet, D. A. (1969). Physique and social image. *Perceptual and Motor Skills, 28,* 295–299.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology*, *70*, 893–912.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35,* 656–666.

Streissguth, A. P., Herman, C. S., & Smith, D. W. (1978). Intelligence, behavior, and dysmorphogenesis in the fetal alcohol syndrome: A report on 20 patients. *The Journal of Pediatrics, 92,* 363–367.

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

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3,* 452–460.

Todd, J. T., Mark, L. S., Shaw, R. E., & Pittenger, J. B. (1980). The perception of human growth. *Scientific American, 24,* 106–114.

Waldrop, M. F., & Halverson, C. F. (1972). Minor physical anomalies: Their incidence and relation to behavior in a normal and a deviant sample. In R. C. Smart & M. S. Smart (Eds.), *Readings in child development and relationships* (pp. 146–155). New York: Macmillan.

Yzerbyt, V., Rocher, S., & Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The social psychology of stereotyping and group life* (pp. 20–50). Malden, MA: Blackwell Publishers.

Zebrowitz, L. A. (1990). *Social perception.* Pacific Grove, CA: Brooks/Cole.

Zebrowitz, L. A. (1996). Physical appearance as a basis of stereotyping. In N. MacRae, M. Hewstone, & C. Stangor (Eds.), *Foundations of stereotypes and stereotyping* (pp. 79–120). New York: Guilford Press.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview.

Zebrowitz, L. A., Andreoletti, C., Collins, M. A., Lee, S. Y., & Blumenthal, J. (1998). Bright, bad, babyfaced boys: Appearance stereotypes do not always yield self-fulfilling prophecy effects. *Journal of Personality and Social Psychology, 75,* 1300–1320.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review, 1,* 204–223.

Zebrowitz, L. A., Collins, M. A., & Dutta, R. (1998). Appearance and personality across the lifespan. *Personality and Social Psychology Bulletin, 24,* 736–749.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin, 28,* 238–249.

Zebrowitz, L. A., & Lee, S. Y. (1999). Appearance, stereotype incongruent behavior, and social relationships. *Personality and Social Psychology Bulletin, 25,* 569–584.

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced males and females across the lifespan. *Developmental Psychology, 28,* 1143–1152.

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individuated impressions of other racial groups. *Journal of Personality and Social Psychology, 65,* 85–101.

Zebrowitz, L. A., Olson, K., & Hoffman, K. (1993). Stability of babyfaceness and attractiveness across the lifespan. *Journal of Personality and Social Psychology, 64,* 453–466.

Zebrowitz, L. A., & Rhodes, G. (2002). Nature let a hundred flowers bloom: The multiple ways and wherefores of attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 261–293). Westport, CT: Ablex.

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the lifespan. *Personality and Social Psychology Bulletin, 22,* 1258–1269.