Journal of Personality and Social Psychology
1993, Vol. 64, No. 3, 453-466

0022-3514/93/$3.00

Stability of Babyfaceness and Attractiveness Across the Life Span

Leslie A. Zebrowitz, Karen Olson, and Karen Hoffman

Differential, structural, and absolute stability of babyfaceness and attractiveness at 5 ages were investigated. Attractiveness had differential stability across the life span. Babyfaceness had differential stability from childhood through the 30s for males and through adolescence for females. Consistent with sexual dimorphisms in facial maturation, males had less differential stability in babyfaceness from childhood to puberty than females. Structural stability of facial appearance, as reflected in the relationship between babyfaceness and attractiveness across the life span, was low, with these qualities positively related for females in childhood and for both sexes in their 30s and 50s but unrelated in puberty and adolescence. Absolute stability of babyfaceness and attractiveness was also low, with mean levels decreasing across the life span. Contrary to cultural stereotypes, age-related decreases in attractiveness were equal for male and female Ss.

Our impressions of people of all ages are strongly influenced by their facial appearance. More specifically, considerable research provides evidence for an attractiveness halo effect, whereby people who are physically attractive are perceived more positively on a variety of dimensions, and a babyface overgeneralization effect, whereby people with more neotenous facial structures are perceived to have more childlike traits (see Berry & McArthur, 1986; Berscheid & Walster, 1974; Bull & Rumsey, 1988; Eagly, Ashmore, Makhijani, & Longo, 1991, for reviews of pertinent research). These facial stereotypes have been shown to hold true for perceivers and targets of all ages (Adams & Huston, 1975; Johnson & Pittenger, 1984; Langlois, 1986; Montepare & Zebrowitz-McArthur, 1989; Zebrowitz & Montepare, 1992). They also have been shown to have significant social interaction consequences at all ages (e.g., Berkowitz & Frodi, 1979; Dion, 1972, 1974; Felson, 1980; Lerner & Lerner, 1977; Salvia, Algozzine, & Sheare, 1977; Zebrowitz, Kendall-Tackett, & Fafel, 1991; Zebrowitz & McDonald, 1991). These consequences span a wide variety of social interaction domains, including popularity, dating, and marriage; education and employment; and disciplinary treatment and the criminal justice system.

The profound influence of facial appearance on social perceptions and social interactions has led theorists to hypothesize that it will also exert a significant influence on personality development. More specifically, it has been proposed that facial appearance creates certain expectations in perceivers that in turn lead to social interactions that elicit the very behavior that had been expected. This elicited behavior can then influence an individual's self-concept such that similar behaviors are emitted even in the absence of specific social expectations. In this manner, people may acquire the stable behavioral dispositions (i.e., personality) that their facial appearance led others to expect (Adams, 1977; Berscheid & Walster, 1974; Langlois, 1986; Sorrell & Nowak, 1981). Although experimental research has demonstrated that facial appearance can indeed produce a short-term self-fulfilling prophecy effect (Snyder, Tanke, & Berscheid, 1977), long-term effects on personality development require a certain degree of stability in appearance. The purpose of the present study was to determine whether facial appearance does in fact show stability across time.

Differential stability has been defined as "the consistency of individual differences within a sample of individuals over time, the retention of an individual's *relative* placement within the group" (Caspi & Bem, 1990, p. 550). This type of stability is indexed by correlations between the appearance of a group of individuals at two points in time. If correlations are low—that is, if babyfaceness or attractiveness relative to one's peers changes markedly across the life span—then social expectations vis-à-vis one's peers would also change and appearance could not exert a consistent effect on behavior. Thus, the differential stability of appearance must be demonstrated if one is to argue for its long-term effects on individuals' behavior relative to those of the same age.

A second type of appearance stability that has implications for personality development is *absolute stability,* which is constancy over time in the absolute level of an attribute, such as attractiveness or babyfaceness (Caspi & Bem, 1990). This type of stability is typically indexed by mean differences in the appearance of a group of individuals at two points in time. If the difference is significant—that is, if the average level of attractiveness or babyfaceness changes markedly with age—this could yield age-related differences in social expectations and consequent behaviors among people in one age group versus

Leslie A. Zebrowitz, Karen Olson, and Karen Hoffman, Department of Psychology, Brandeis University.

This research was supported by National Institute of Mental Health Grant BSR 5 R01 MH42684 to Leslie A. Zebrowitz. Thanks are extended to Tom Boone, Regina Conte, Heather Coon, Rebecca Gilbert, Sean Killen, Sheryl Levy, Caleb McArthur, Lynn Steiner, and Liz Tighe for their help in collecting and coding the data and to Margie Lachman and Ron Spiro for their statistical assistance in the data analyses.

We express appreciation to the Institute of Human Development at the University of California, Berkeley, for providing the slides of the subjects' faces and to Barbara Burek for her invaluable assistance.

Correspondence concerning this article should be addressed to Leslie A. Zebrowitz, Department of Psychology, Brandeis University, Waltham, Massachusetts 02254-9110.

another. Thus, the absolute stability of appearance has implications for its long-term effects on the continuity of behavior at the group level.

A third type of appearance stability with implications for personality development is *structural stability,* which refers to stability in the relationships among various qualities across the life span (Caspi & Bem, 1990). This type of stability would be indexed by comparing the correlations between babyfaceness and attractiveness at two points in time. If the relationship between babyfaceness and attractiveness changes markedly with age, then the ultimate influence of either one of these qualities on social expectations and resultant behaviors may also change. For example, if babyfaceness and attractiveness are highly correlated in childhood, but not in adulthood, then the behaviors expected from babyfaced individuals may differ at these two age levels.

No studies have investigated the differential stability of babyfaceness, and although several have found differential stability of attractiveness, many of these studies were restricted to very narrow age spans, such as infancy (Langlois & Stephan, 1981) or childhood (Adams, 1977; Maruyama & Miller, 1981). Others have documented differential stability in attractiveness across somewhat broader age ranges, such as childhood to adolescence (Sussman, Mueser, Grau, & Yarnold, 1983; Pittenger, Mark, & Johnson, 1989), childhood to young adulthood (Alley, 1993), and young adulthood to middle age (Adams, 1977; Sparacino, 1980). However, the generalizability of these studies is limited either by the exclusion of faces of one sex (Sparacino, 1980; Sussman et al., 1983) or by a very small sample of faces (Adams, 1977; Alley, 1993; Pittenger et al., 1989). Another shortcoming of past stability studies is that only one photograph has been included at each age level. Unless photographs are carefully standardized, this method does not allow one to be certain that changes from one age to another in rated appearance reflect real changes in facial structure rather than transient factors associated with a particular photograph.

The absolute stability of babyfaceness, like its differential stability, has not been investigated. Whereas several studies have examined the absolute stability of attractiveness, this research is limited by the fact that it has used very small numbers of faces, often sampled cross-sectionally, and the results have been mixed. Two small-sample longitudinal investigations showed a linear decrease in attractiveness from childhood to maturity, particularly for male subjects (Alley, 1993; Pittenger et al., 1989). A small-sample, cross-sectional study, on the other hand, revealed a curvilinear relationship between age and attractiveness, with 17-year-olds rated higher than 7- or 30- to 50-year olds (Cross & Cross, 1971). Other cross-sectional studies have also shown an age-related decline in attractiveness at middle and older adulthood, particularly for women (Deutsch, Clark, & Zalenski, 1983; Milord, 1978; Walsh & Locke, 1980), but still others have shown either no decline (Johnson, 1985) or no consistent sex differences in the decline (Berman, O'Nan, & Floyd, 1981; Henss, 1989). Because of their cross-sectional nature, small sample sizes, or both, the foregoing studies do not provide unequivocal evidence that there will be changes in the attractiveness of the same individuals across the life span (see Alley, 1988, and Jackson, 1992, for pertinent reviews of the literature on the stability of attractiveness).

Only one study has investigated the structural stability of appearance, comparing the relationship between babyfaceness and attractiveness across the life span (Zebrowitz & Montepare, 1992). However, this was a cross-sectional investigation, and age-related differences in the correlation between these two appearance qualities could reflect differences among the sets of faces evaluated at each age rather than age-related changes in structural stability.

The present research extended past work on the stability of facial appearance in several respects. First, the stability of babyfaceness was examined in addition to the stability of facial attractiveness, and the degree of stability in these two qualities was compared. Second, stability was investigated across a broader time span ranging from childhood to older adulthood with five age levels represented: childhood, puberty, adolescence, the 30s, and the 50s. Third, multiple photographs were included at various age levels to ensure that appearance ratings reflected the targets' actual facial structure rather than idiosyncratic qualities of a particular photograph. Fourth, a large sample of faces was investigated. Finally, differential, structural, and absolute stability were all assessed. Differential and structural stability were examined in Study 1, where raters judged each face relative to others of the same age and sex. Absolute stability was examined in Study 2, where raters judged faces without regard to age or sex.

Study 1

It was predicted that there would be significant differential stability in babyfaceness and attractiveness from childhood through older adulthood and that the degree of such stability would vary as a function of the age levels under consideration as well as the sex of the faces. Lower differential stability was expected between childhood and puberty than between puberty and adolescence because of individual differences in maturational changes in the face. These changes may include development of the brow ridges and frontal sinuses, which produces a forward growth of the forehead; growth of the soft tissues of the nose and the nasal bone, which brings the point of the nose further forward and downward relative to the rest of the face; and elongation and projection of the jaw (Tanner, 1962). Because pubertal changes in the face are typically greater for boys than for girls, who retain a more neotenous facial structure into adulthood (Gray, 1985; Tanner, 1962), less differential stability in appearance was expected between childhood and puberty for male subjects than for female subjects. Like puberty, older adulthood was expected to reflect relatively low differential stability in appearance because of individual differences in facial changes associated with aging. In particular, bone loss may produce a less angular jaw line, and loss of elasticity in the skin may produce wrinkling, pouches, hooding of the eyes, and a double chin. In the absence of evidence that these changes are more common for one sex than for the other, less differential stability in appearance was predicted for both male and female subjects between the 30s and the 50s than between adolescence and the 30s.

Structural stability, as reflected in age-related changes in the correlation between babyfaceness and attractiveness, was difficult to predict from past research. In infants of both sexes,

attractiveness is positively related to babyfaceness (Hildebrandt & Fitzgerald, 1979; Ritter, Casey, & Langlois, 1990; Sternglanz, Gray, & Murakami, 1977; Zebrowitz & Montepare, 1992). During childhood and puberty, research has revealed no significant correlation between these qualities for targets of either sex (Zebrowitz & Montepare, 1992). In adulthood, the data are mixed, with some studies reporting a positive relationship for men but not women (Zebrowitz & Montepare, 1992), others reporting a positive relationship for women but not men (Keating, 1985; McArthur & Apatow, 1983–1984), and still others reporting that the most attractive individuals are those with a mixture of mature and babyish features, which suggests no correlation between babyfaceness and attractiveness (Cunningham, 1986, 1990). Although these findings clearly suggested that the correlation between babyfaceness and attractiveness would vary across the life span, thus revealing low structural stability, the precise nature of the correlations could not be predicted with any confidence.

Method

Subjects

The subjects for this study were people whose pictures were included in the Intergenerational Studies of Development and Aging, a combination of three longitudinal studies begun between 1928 and 1933 and archived at the University of California Berkeley Institute of Human Development (IHD). Three samples were used: Berkeley Guidance, Berkeley Growth, and Oakland Growth, hereafter referred to, respectively, as Guidance, Berkeley, and Oakland. The original 248 people in the Guidance sample were born between 1928–1929 in Berkeley, California. Family educational status was above the average for the general U.S. population (Eichorn, 1981). Most were from White, Protestant families, and 60% came from middle-class homes (Caspi, Elder, & Bem, 1987). The Berkeley sample consisted of 61 healthy, White, full-term infants born in Berkeley between September 1928 and May 1929. Over the next 3 years, an additional 13 infants were added to augment the sample. Most of the families were middle class and Protestant (Eichorn, 1981). The first measures for the original 212 subjects in the Oakland sample were taken in 1932, when the subjects ranged in age from 10 to 12 years. Therefore, they were, on average, 7.75 years older than either the Guidance or Berkeley subjects. All of the families were White, a reasonable representation of the population attending Oakland schools at that time (Eichorn, 1981).

Slides of subjects' faces were prepared from black-and-white photographs contained in the archives, and they were included in this study if pictures were available for at least two age levels between childhood and later adulthood. Each slide had an identification number and the age of the person printed along the lower edge. The faces had neutral expressions, except for those at later adulthood, which often had smiling expressions.

The people in the slides were White, ranged in age from 8 to 62, and represented five age levels: childhood ($M = 10.15$, $SD = 0.46$), puberty ($M = 14.55$, $SD = 0.42$ for boys; $M = 12.55$, $SD = 0.43$ for girls), adolescence ($M = 17.49$, $SD = 0.58$), the 30s ($M = 31.20$, $SD = 1.70$), and the 50s ($M = 55.58$, $SD = 3.56$). The different puberty ages selected for boys and girls reflect normative data indicating that the adolescent growth spurt in girls is typically completed by age 13, whereas the growth spurt in boys is typically completed by age 15 (Tanner, 1962). A sixth age level, early childhood ($M = 4.60$ years, $SD = 0.49$), was also investigated for a small group of subjects whose pictures were available at this age. Although the sample size was too small to include this age level in the main data analyses reported later, stability from early childhood to childhood is reported for the interested reader.

The maximum number of slides available for each subject in the Guidance sample was 11: 3 at childhood, 3 at puberty, 3 at adolescence, and 1 each in the 30s and in the 50s. The Oakland sample had a maximum of either 9 slides for female, or 10 for male subjects, consisting of 1 slide for female, or 2 slides for male subjects at childhood, 3 at puberty, 3 at adolescence, and 1 each in the 30s and in the 50s. (The age of the latter slide was actually in the 60s for most Oakland subjects.) The Berkeley sample had a maximum of 10 slides: 3 at childhood, 3 at puberty, 3 at adolescence, and 1 in the 50s. A subset of this sample also had 2 slides available at a sixth age level, early childhood.

Originally, there was a total of 379 subjects rated, 185 males and 194 females. However, a subset of subjects (103 males and 104 females), who had slides at a minimum of four age levels and were not missing more than five ratings in all, was selected for this report. All of these subjects were rated at childhood, puberty, adolescence, and either during their 30s or during their 50s or both. Fifty-eight male and 64 female subjects were drawn from the Guidance sample, 33 male and 26 female subjects were from the Oakland sample, and 12 male and 14 female subjects were from the Berkeley sample.

Judges

One hundred ninety-two undergraduate students enrolled in introductory psychology classes rated one of three samples of faces for $5 and partial credit toward a course requirement. For the Berkeley sample, 16 male and 16 female subjects rated either male or female faces on both attractiveness and babyfaceness, completing one rating for all slides before rating the other characteristic, with order of the two ratings counterbalanced across judges. For the Guidance and Oakland samples, approximately equal numbers of male and female subjects rated male or female faces on either attractiveness or babyfaceness. In general, ratings were done in groups of eight.

Dependent Measures

Each slide was rated for both facial attractiveness and facial babyfaceness using 7-point scales. The endpoints of the attractiveness scales were labeled *unattractive* and *attractive.* The endpoints of the babyface scales were labeled *babyfaced* and *maturefaced.* Each judge was given a packet consisting of a set of scales. The slide identification number was listed to the left of each scale, and each age group began on a new page. The age group of the people in the slides was written in the corner of each page.

Procedure

The judges were told that the purpose of the study was to examine whether judgments of appearance were related to personality traits. More specifically, they were told that they would rate a series of male (or female) faces of various ages on attractiveness (or babyfaceness, or both) so that these appearance qualities could be correlated with personality measures that had been previously taken on the individuals they would be rating. Judges were instructed to rate each face relative to others of the same age and were quickly shown a few slides at the beginning of each age group to ensure they were familiar with some faces at the new age before beginning their ratings.

Each slide was shown for approximately 6 s, and the total time of a session varied from 1–2 hr, depending on which sample was being rated. The slides were shown in either a forward or backward order. In the forward order, slides were presented in ascending order according to age. In the backward order, the slides were reversed within each age group and shown beginning with the oldest age group and ending with

the youngest. Judges rating the Guidance sample saw either 712 female or 610 male slides. Oakland judges rated 443 female or 465 male slides. Judges rating slides from the Berkeley sample saw 280 female slides or 489 male slides. Although most of these slides were from the Berkeley sample, there were also slides from the other two samples, some of which had not been previously rated and others of which were rerated because of missing data. All together, 2,847 slides were rated. This study reports on a subset of 1,816 of these slides, representing those subjects who had slides at a minimum of four age levels.

Results

Overview

Three sets of analyses were undertaken. The first used structural equation modeling to examine the differential stability of babyfaceness and attractiveness for men and women across the life span. The purpose of this analysis was not to find the best model for maximizing the prediction of appearance but rather to assess the differential stability of appearance between successive age levels. The advantage of structural equation modeling for this purpose is that it provided a measure of stability unattenuated by error. A disadvantage is that planned comparisons to assess age and sex differences in appearance stability are very cumbersome, requiring the creation of multiple models. Consequently, a second set of analyses computed Pearson correlation coefficients between appearance ratings at successive age levels, and *z*- and *t*-test planned comparisons were performed on these coefficients. The third set of analyses examined the structural stability of appearance as reflected in the correlations between attractiveness and babyfaceness across the life span.

Reliability and Validity of Ratings

The average correlations between the ratings of male and female judges were quite high. For attractiveness, the average correlation was .79 for male faces and .82 for female faces. For babyfaceness, the average correlation was .75 for male faces and .77 for female faces. Therefore, reliability was assessed after combining ratings from both the male and female judges.

Reliability coefficients were calculated for each age group within each sex and ranged from .71 to .96. The average alpha coefficient for attractiveness ratings was .87 for male faces and .90 for female faces. The average alpha coefficient for babyfaceness ratings was .85 for male faces and .85 for female faces. Given these high reliabilities, mean ratings of attractiveness and mean ratings of babyfaceness were calculated from the ratings of all judges for each face at each age and used in subsequent data analyses.

The fact that judges rated photographs of the same subjects at various ages raises the possibility that findings of stability could reflect these repeated measures rather than actual stability in appearance. However, there are three arguments against this claim. First, the judges were required to rate too many slides too quickly to be able to make a conscious attempt to be consistent. Any carryover effects would have been weakened by the large number of intervening slides. Second, because many subjects were missing pictures at various ages, the actual set of subjects being rated varied somewhat from age to age with consequent variations in the faces that immediately preceded and followed a given subject. Third, certain groups of slides were procured after the others in that sample had already been rated, and these slides were rated by judges who had not seen the same subjects at other age levels. This resulted in 10 instances in which correlations between attractiveness or babyfaceness at adjacent ages could be compared when the two ratings were made by the same judges versus different judges. In no case did ratings by the same judges yield higher stability than ratings by different judges. The average correlation between adjacent ages was .51 when both ratings were made by the same individuals and .53 when ratings of slides at the same adjacent ages were made by two sets of individuals.

Further evidence of the validity of the ratings obtained in the present study is provided by a comparison of attractiveness ratings for late adolescent girls with prettiness ratings of the same female faces obtained in the 1960s when 14 judges (7 men and 7 women) drawn from the staff at the IHD rated only this age group (B. Burek, personal communication, September 1991). The significant correlation between these two sets of ratings, $r(64) = .50$, demonstrates that judgments of the relative attractiveness of subjects in the present study may be generalized to another method, where no carryover effects are possible, to another historical time, where standards of beauty may have been different, and to another sample of judges.

Missing Data

Missing data in childhood, puberty, and adolescence were replaced with regression estimates of their values based on faces rated within the same age level, an appropriate procedure when data are assumed to be missing at random even if not observed at random (Little & Rubin, 1987). Because there was only one face within each age level in the 30s and 50s, regression estimates for missing data could not be used, and mean values for each of these ages were used instead. Missing data were replaced rather than handled by listwise deletion because there were no cases with data across all 11 observation times. Pairwise deletion was rejected because this procedure would make it impossible to interpret differences in stability at different age levels. For example, with pairwise deletion, a finding that attractiveness is more stable from adolescence to the 30s than from the 30s to the 50s could reflect either the presence of different people in the sample at the two transitions or a larger sample size at the earlier transition rather than reflect greater stability of attractiveness in young than in older adulthood.

The percentage of missing data is shown in Table 1. It should be noted that although there is sometimes a sizable percentage of missing data at one specific age (e.g., Childhood 1) every subject had at least one photo at childhood, at puberty, and at adolescence, and all subjects had a photo either in the 30s, 50s, or both. In most cases, missing data can be attributed to none of the subjects' in a sample being photographed at that age. Specifically, slides of Berkeley subjects were not available at age 30, Oakland female slides were available at only one childhood age, and Oakland male slides were available at only two of the three childhood ages. Because the unavailability of slides could not explain the high percentage of missing data at the last two adolescent ages, it was possible that the subjects who were retained

Table 1
Percentage of Missing Data for Faces at Each Age for Both Attractiveness and Babyfaceness Ratings

	Childhood			Puberty			Adolescence				
Sex of face	1	2	3	1	2	3	1	2	3	30s	50s
Male	47	26	5	16	17	9	11	30	37	28	17
Female	47	40	5	9	10	3	12	27	34	32	20

Note. $n = 103$ for male faces and $n = 104$ for female faces. All subjects had at least one photo at childhood, puberty, and adolescence as well as a photo at 30, 50, or both.

at these ages differed in appearance from those who were lost. If so, this could conceivably inflate the stability of appearance from puberty to adolescence. To investigate this possibility, the correlation between mean appearance ratings at puberty and Adolescence 1, when most subjects were retained, were compared for those subjects who were also retained at Adolescence 2 or 3 versus those who were lost. The results revealed that stability of appearance from puberty to Adolescence 1 was comparable for the two subgroups. The average correlation of male and female subjects' babyfaceness and attractiveness from puberty to Adolescence 1 was .65 for those subjects who were retained at Adolescence 2 or 3 and .56 for those who were lost. The high percentage of missing values at the last two adolescent ages was consequently assumed to be unrelated to the characteristics assessed in this study. Nevertheless, all of the analyses reported herein were corroborated by analyses on a smaller subset of subjects who had no missing data.

Differential Stability

Structural equation modeling. Ratings of attractiveness and babyfaceness of male and female subjects were examined in four separate analyses using the Lisrel 7 program for structural equation modeling (Jöreskog & Sörbom, 1989). Appearance at five age levels was treated as the latent variable, with three indicators at childhood, puberty, and adolescence. In the 30s and in the 50s, there was only one indicator at each age level. Therefore, the path from the latent to the indicator variable was fixed at 1.0 and the error fixed at [(1 − reliability) × variance] for that indicator (Hayduk, 1987). This model is illustrated in Figure 1. Input for each analysis was an 11 × 11 covariance matrix. Correlations and standard deviations are reported in Tables 2 through 5. The results are summarized in Table 6.

Statistics related to the fit of the models are also presented in Table 6. All chi-squares were nonsignificant, indicating that the model provided a good fit to the data. Although the moderate goodness-of-fit statistics suggested that the model might have been improved by including other variables, it should be recalled that the goal of these analyses was not to maximize the prediction of facial appearance but rather to assess its stability between successive age levels. As shown in Table 6, the standardized coefficients reflected significant stability across the entire age span of this study for the attractiveness of both male and female subjects. Babyfaceness, on the other hand, showed stability only from childhood through the 30s for male subjects and from childhood through adolescence for female subjects. To extend the findings further down the age scale, Pearson correlation coefficients were calculated between mean appearance ratings at early childhood and childhood for the small sample of subjects with pictures at early childhood. The results revealed that appearance was stable down to this early age for female subjects, $rs(10) = .62$ and .65 for attractiveness and babyfaceness, respectively (both $ps < .05$). However, due in part to the small sample size, these stability effects failed to attain statistical significance for male subjects, $rs(10) = .48$ and .31, for attractiveness and babyfaceness, respectively (both $ps > .05$).

Planned comparisons. To assess sex and age differences in the stability of appearance, planned comparisons were performed on Pearson correlation coefficients between subjects' mean appearance ratings at each age level as computed from the same data that was input into the Lisrel analyses. As shown in Table 7, these correlations showed the same pattern of results as the path coefficients from the Lisrel analysis, although the path coefficients are all slightly higher, by virtue of being corrected for attentuation due to unreliability. Comparisons between the coefficients for male and female subjects were performed by converting them to z scores, whereas t-test comparisons were performed within sex to compare the stability coefficients for different pairs of age levels (Cohen & Cohen, 1983).

Comparisons between the correlation coefficients for male subjects at various age levels supported the prediction of lower stability from childhood to puberty than from puberty to adolescence for babyfaceness, albeit not for attractiveness, $ts(100) = 3.18$ and 1.03, $ps < .01$ and $> .10$. Also as predicted, male subjects showed higher stability from adolescence to the 30s than from the 30s to the 50s for babyfaceness, but they did not show this effect for attractiveness, $ts(100) = 3.56$ and 1.24, $ps < .01$ and $> .10$, respectively.

Comparisons between the correlation coefficients for female subjects at various age levels revealed that, contrary to prediction, the stability in babyfaceness and attractiveness from childhood to puberty was not significantly lower than from puberty to adolescence, $ts(101) = 1.28$ and 1.30, respectively (both $ps > .10$). Also contrary to prediction, the stability of babyfaceness and attractiveness was not higher from adolescence to the 30s than from the 30s to the 50s, $t < 1$ and $t(101) = 1.41$, $p > .10$, respectively.[1]

Planned comparisons between the correlation coefficients for male and female subjects supported the prediction of less differential stability in babyfaceness from childhood to puberty among male than female subjects, $z = 2.71$, $p < .01$. However, there was no significant sex difference in the differential stability of attractiveness at this age ($z = 1.52$, $p > .10$).

Although no predictions had been made concerning the degree of differential stability in babyfaceness versus attractiveness, post hoc comparisons were made between the stability coefficients for these two appearance qualities at each age transition. None of the differences were significant for male subjects (all $ps > .10$), indicating equal stability of babyfaceness and attractiveness across the age span investigated. For female subjects, attractiveness was more stable than babyfaceness at

[1] Additional, unpredicted findings were higher stability in babyfaceness and attractiveness from puberty to adolescence than from adolescence to the 30s for both male and female subjects (all $ps < .01$).

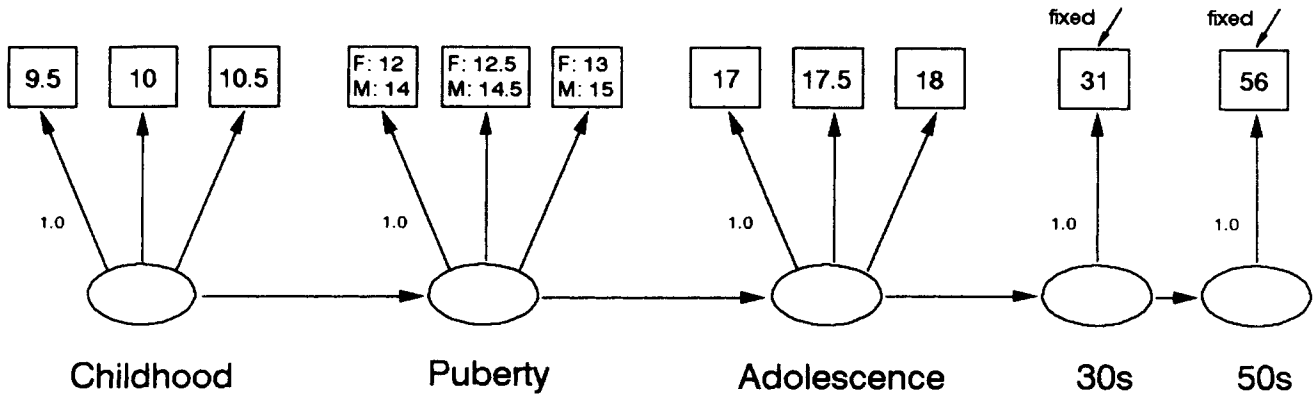

Figure 1. Structural equation model. (Four separate analyses were performed for attractiveness and babyfaceness ratings of male (M) and female (F) faces. Measurement error was fixed at [(1 − reliability) × (variance)] for single indicator variables. Numbers within observed variable boxes represent approximate ages of the faces.)

the transition from adolescence to the 30s ($z = 2.41$, $p < .05$), whereas the two appearance qualities were equally stable at all other age transitions (all $ps > .10$).

Effect of replacing missing data. Because replacing missing data with the group mean at the 30s and 50s age levels would reduce the variability of the appearance measures, it is possible that the lower differential stability of appearance at these ages reflected the procedure for handling missing data rather than a true effect of age. To assess this possibility, correlations were computed between mean appearance ratings at one age level with the next level for that subset of subjects who had at least one slide at every level. A subject's appearance rating at each age level was the average of all ratings available at that age. As shown in Table 7, these correlations replicated the magnitudes and pattern obtained when missing data were replaced as described earlier. Thus, the procedure for handling missing data did not affect any of the conclusions regarding the differential stability of appearance.

Effect of variations in facial expression. Because faces in the 50s often had smiling expressions, in contrast with the predominantly neutral expressions at other ages, it is possible that the lower differential stability at this age reflected variations in facial expression rather than a true effect of age. To assess this possibility, the correlations reported in Table 7 for the sample with no missing data were compared with partial correlations, controlling for changes in a subject's facial expression from one age to the next. (The facial expression ratings, which were collected for Study 2, are described in more detail in Study 2.) No effects were lost or gained in these partial correlations, which were practically identical to the zero orders. Thus, age-related changes in facial expression did not affect the differential stability of appearance.

Structural Stability

Correlations between the mean babyfaceness and attractiveness at each age level revealed that the relationship varied with

Table 2
Correlation Matrices and Standard Deviations for LISREL Input: Attractiveness Ratings of Male Faces (n = 103)

Age level	1	2	3	4	5	6	7	8	9	10	11
1. Ch1	—										
2. Ch2	.59	—									
3. Ch3	.64	.58	—								
4. Pub1	.44	.54	.46	—							
5. Pub2	.47	.49	.52	.78	—						
6. Pub3	.37	.50	.42	.75	.73	—					
7. Adol1	.31	.43	.41	.50	.57	.60	—				
8. Adol2	.34	.42	.40	.58	.63	.58	.71	—			
9. Adol3	.33	.41	.42	.50	.57	.56	.67	.87	—		
10. Ad30	.15	.18	.09	.21	.35	.24	.32	.37	.40	—	
11. Ad50	.14	.28	.15	.09	.11	.11	.16	.24	.18	.26	—
SD	.61	.66	.77	.71	.73	.80	.69	.67	.76	.69	.57

Note. Ch = childhood (ranged from 9.5 to 10.5 years); Pub = puberty (ranged from 12 to 13 years for girls and from 14 to 15 years for boys); Adol = adolescence (ranged from 17 to 18 years); Ad30 = adulthood from 30 to 39 years; Ad50 = adulthood from 52 to 60 years.

Table 3
Correlation Matrices and Standard Deviations for LISREL Input: Attractiveness Ratings of Female Faces (n = 104)

Age level	1	2	3	4	5	6	7	8	9	10	11
1. Ch1	—										
2. Ch2	.87	—									
3. Ch3	.87	.83	—								
4. Pub1	.67	.69	.70	—							
5. Pub2	.60	.59	.62	.72	—						
6. Pub3	.52	.50	.54	.58	.67	—					
7. Adol1	.48	.43	.45	.49	.56	.49	—				
8. Adol2	.46	.40	.44	.50	.58	.54	.85	—			
9. Adol3	.45	.42	.43	.48	.59	.50	.79	.83	—		
10. Ad30	.35	.30	.29	.28	.29	.34	.27	.34	.45	—	
11. Ad50	.15	.13	.12	.11	.22	.11	.21	.21	.19	.22	—
SD	.82	.83	.88	.86	.83	.90	.81	.87	.86	.68	.71

Note. Ch = childhood (ranged from 9.5 to 10.5 years); Pub = puberty (ranged from 12 to 13 years for girls and from 14 to 15 years for boys); Adol = adolescence (ranged from 17 to 18 years); Ad30 = adulthood from 30 to 39 years; Ad50 = adulthood from 52 to 60 years.

subjects' age and sex. As shown in Table 8, significant positive correlations were obtained for female subjects at childhood and for both male and female subjects in their 30s and 50s. At puberty and adolescence, on the other hand, these two facial characteristics were unrelated for subjects of both sexes. To extend the findings further down the age scale, Pearson correlation coefficients were calculated between mean babyfaceness and attractiveness at early childhood for subjects with pictures at that age. The results revealed positive correlations for both male and female subjects, $rs(10) = .76$ and .80, respectively (both $ps < .10$).

Another set of correlations was computed to assess whether the results of the structural stability analyses were distorted by the procedure for handling missing data or by the fact that the same raters had judged both the attractiveness and the babyfaceness of some subjects. Specifically, correlations were computed between mean attractiveness and babyfaceness at each age level for that subset of subjects who had a slide at every age level and whose babyfaceness and attractiveness were judged by different raters. A subject's appearance rating at each age was the average of all ratings available at that age. One cell in the design was empty in this analysis because there were no male subjects in the 50s meeting the criteria for inclusion. As shown in Table 8, the correlations that could be computed replicated the pattern obtained when missing data were replaced and when subjects with babyfaceness and attractiveness ratings from the same judge were included. Thus, the positive relationships between these two appearance qualities cannot be attributed to the procedures for replacing missing data or to carryover effects between ratings of babyfaceness and attractiveness.

Finally, the babyfaceness–attractiveness correlations for subjects with no missing data were compared with partial correlations, controlling for facial expression. No effects were lost or gained in these partial correlations, which replicated the mag-

Table 4
Correlation Matrices and Standard Deviations for LISREL Input: Babyfaceness Ratings of Male Faces (n = 103)

Age level	1	2	3	4	5	6	7	8	9	10	11
1. Ch1	—										
2. Ch2	.76	—									
3. Ch3	.75	.59	—								
4. Pub1	.50	.41	.56	—							
5. Pub2	.49	.41	.54	.81	—						
6. Pub3	.51	.41	.56	.85	.82	—					
7. Adol1	.29	.35	.36	.63	.62	.64	—				
8. Adol2	.39	.38	.43	.70	.64	.72	.71	—			
9. Adol3	.49	.47	.50	.69	.69	.69	.67	.86	—		
10. Ad30	.22	.24	.21	.27	.24	.35	.42	.38	.39	—	
11. Ad50	.25	.27	.17	.22	.23	.20	.32	.22	.25	.06	—
SD	.60	.64	.68	.84	.82	.84	.72	.75	.70	.71	.62

Note. Ch = childhood (ranged from 9.5 to 10.5 years); Pub = puberty (ranged from 12 to 13 years for girls and from 14 to 15 years for boys); Adol = adolescence (ranged from 17 to 18 years); Ad30 = adulthood from 30 to 39 years; Ad50 = adulthood from 52 to 60 years.

Table 5
Correlation Matrices and Standard Deviations for LISREL Input: Babyfaceness Ratings of Female Faces (n = 104)

Age level	1	2	3	4	5	6	7	8	9	10	11
1. Ch1	—										
2. Ch2	.81	—									
3. Ch3	.83	.77	—								
4. Pub1	.67	.65	.66	—							
5. Pub2	.67	.63	.68	.81	—						
6. Pub3	.66	.66	.68	.69	.71	—					
7. Adol1	.50	.47	.57	.57	.70	.64	—				
8. Adol2	.60	.57	.64	.55	.67	.56	.78	—			
9. Adol3	.55	.52	.62	.52	.68	.57	.81	.83	—		
10. Ad30	.18	.17	.19	.04	.05	.04	.04	.04	.09	—	
11. Ad50	.03	.03	.01	.06	.11	.10	.16	.10	.21	.06	—
SD	.68	.85	.87	.90	.92	.97	.92	.81	.84	.62	.68

Note. Ch = childhood (ranged from 9.5 to 10.5 years); Pub = puberty (ranged from 12 to 13 years for girls and from 14 to 15 years for boys); Adol = adolescence (ranged from 17 to 18 years); Ad30 = adulthood from 30 to 39 years; Ad50 = adulthood from 52 to 60 years.

nitude and pattern obtained in the zero orders. Thus, variations in the relationship of attractiveness and babyfaceness across the life span were not mediated by age-related variations in facial expression.

Discussion

The prediction of significant differential stability in facial appearance was confirmed for the age span from approximately 10 years to 56 years for the attractiveness of male and female subjects, 10 years to 31 years for the babyfaceness of male subjects, and 10 years to 17 years for the babyfaceness of female subjects. There was also evidence that differential stability in both of these appearance qualities begins as early as 4 years of age. Thus, individuals who are attractive relative to their peers in childhood tend to remain relatively attractive into their late 50s, whereas the relative babyfaceness of individuals in their 50s cannot be predicted from their standing at earlier ages. Analyses controlling for facial expression revealed that this lack of stability cannot be attributed to more positive facial expressions in the 50s than at other ages. Although the lack of stability in babyfaceness for subjects in their 50s was not predicted, this finding is consistent with the prediction that differential stability would be lower at this age level than at others because of individual differences in the facial changes associated with aging. It appears that such changes eliminate the differential stability of babyfaceness but not that of attractiveness.

The finding that babyfaceness was stable through the 30s for male subjects but only through late adolescence for female subjects was unanticipated. One possible explanation is that, in the transition from late adolescence to the 30s, women are more likely than men to begin to show the soft tissue changes that come with aging and that can influence babyfaceness, such as wrinkling and hooded (smaller looking) eyes. If more women than men show such changes, this would yield less stability in women's babyfaceness relative to their peers. This explanation is consistent with the widespread belief that women's faces show the effects of aging earlier than men's do. It is also consistent with the finding that babyfaceness lacked stability for both sexes in the transition to the 50s, when the soft tissue effects of aging should be evident for both sexes. Interestingly, however, if

Table 6
Differential Stability of Appearance as Reflected in the Results of the LISREL Analysis

	Standardized path coefficients between latent variables[a]							
Facial appearance	Chld →	Pub →	Adol →	30s → 50s	χ^2 (42)	GFI	AGFI
Attractive							
Male faces	.70	.72	.43	.30	47.51[b]	.927	.885
Female faces	.79	.70	.41	.25	45.37[b]	.927	.886
Babyface							
Male faces	.61	.82	.47	.10[b]	54.30[b]	.916	.868
Female faces	.85	.79	.08[b]	.08[b]	56.78[b]	.914	.864

Note. n = 103 for males faces and n = 104 for female faces. Chld = childhood; Pub = puberty; Adol = adolescence; GFI = Goodness of Fit Index, AGFI = Adjusted Goodness of Fit Index.
[a] All path coefficients are significant except where noted. [b] Not significant.

Table 7
Differential Stability of Appearance as Reflected in Pearson Correlation Coefficients Between Successive Age Levels

Facial appearance	Pearson correlation coefficients between age levels			
	Chld → Pub	Pub → Adol	Adol → 30s	30s → 50s
Attractive				
Male faces				
Correlation 1	.60	.67	.40	.26
Correlation 2	.55	.64	.46	.34
Female faces				
Correlation 1	.72	.64	.38	.22
Correlation 2	.78	.72	.56	.38
Babyface				
Male faces				
Correlation 1	.58	.77	.44	.06[b]
Correlation 2	.57	.75	.51	.09[b]
Female faces				
Correlation 1	.78	.72	.06[b]	.06[b]
Correlation 2	.75	.74	.09[b]	.09[b]

Note. $n = 103$ for male faces and $n = 104$ for female faces for Correlation 1. $n = 56$ for male faces and $n = 50$ for female faces for Correlation 2; these were calculated on a subset of faces with no missing data. Chld = childhood; Pub = puberty; Adol = adolescence.
[a] All correlations are significant at $p < .05$ or better except where noted. [b] Not significant.

there were such sex differences in facial aging in the 30s, they did not yield sex differences in the differential stability of attractiveness.

The prediction that appearance would show less stability from childhood to puberty than between other adjacent age levels received mixed support. As predicted, male subjects showed less differential stability in babyfaceness from childhood to puberty than from puberty to adolescence. However, female subjects showed equally high stability in babyfaceness at these two age transitions. Although it had been expected that female subjects would show a significant decrease in the stability of babyfaceness at puberty, the fact that they showed higher stability than male subjects did was predicted and is consistent with sexual dimorphisms in the facial changes associated with maturation.[2] Neither male nor female subjects showed less stability in attractiveness from childhood to puberty than from puberty to adolescence. It thus appears that the facial changes at puberty affect the babyfaceness of males relative to their peers but have no effect on the relative attractiveness of either sex.

Changes in the correlations between babyfaceness and attractiveness revealed a lack of structural stability in facial appearance across the life span. These appearance qualities were positively correlated for both sexes in early childhood as well as in the 30s and 50s, whereas they were uncorrelated during puberty and late adolescence. The positive correlations in the 30s and 50s are consistent with evidence that postchildhood individuals who look young are judged to be more attractive than those who have a less youthful facial appearance (Sorrell & Nowak, 1981). However, the overall pattern of correlations deviates from a cross-sectional life span investigation that obtained positive correlations only for infants of both sexes and for young adult men (Zebrowitz & Montepare, 1992). Two conclusions may be drawn from these results. First, the relationship between babyfaceness and attractiveness varies with the particular faces sampled at a given age. Second, in a random sample of faces, such as were used in the present study, the impact of babyfaceness on attractiveness varies across the life span. A babyface increases attractiveness in early childhood and in adulthood, whereas it does not reap this benefit during puberty and adolescence.

Study 2

Study 2 examined the absolute stability of attractiveness and babyfaceness for male and female faces across the life span as reflected in age-related changes in the mean levels of these facial qualities. As noted earlier, the findings of past research on the absolute stability of attractiveness have been mixed. In view of these results, no firm predictions could be made. However, it was anticipated that attractiveness would show a steady decline with age and that this effect might be stronger for female faces than for male faces. No overall difference in the attractiveness of male and female faces was expected. Although research has found more consensual, more differentiated, and more extreme attractiveness ratings for female than for male faces, this work found no differences in the mean levels of attractiveness (Schulman & Hoskins, 1986).

In the absence of past research on the absolute stability of babyfaceness across the life span, predictions could be derived only from known age-related structural changes in the face. The significant structural changes during the years of physical maturation should yield linear decreases in babyfaceness from childhood to puberty, particularly for male faces. The more limited structural changes between puberty and the early 30s should yield stability in babyfaceness at this time. Predictions regarding stability from the 30s through the 50s are less straightforward. Some facial changes in older adulthood, such as less angular jaws and double chins, contribute to a more neotenous appearance and may yield an increase in babyfaceness ratings at this age. However, other facial changes, such as hooding of the eyes and wrinkling of the skin, contribute to a less neotenous appearance and may yield a decrease in babyfaceness ratings at this age. On the basis of the sexual dimorphisms in facial appearance that develop at puberty, it was expected that female subjects would be more babyfaced than male subjects from puberty through the 50s with no sex differences in facial maturity in childhood.

To investigate absolute stability in appearance across age and gender, several methodological changes were made in the procedure used in Study 1. First, faces of various ages were presented in a random order, raters were not told the ages of the faces, and they were not instructed to judge each face relative to

[2] It should be noted that this sex difference cannot be attributed to the greater number of years between childhood and puberty for boys. Male subjects also showed less stability than female subjects between childhood and adolescence, when the same number of years had elapsed for both sexes (see Tables 4 and 5).

Table 8
Correlations Between Attractiveness and Babyfaceness at Each Age Level

Sex of face	Chld	Pub	Adol	30s	50s
Male					
Correlation 1	−.06	−.11	−.10	.31*	.45**
Correlation 2	−.14	−.25	−.12	.60**	—
Female					
Correlation 1	.22*	.08	−.09	.46**	.47**
Correlation 2	.15	.03	−.14	.66**	.35*

Note. $n = 103$ for male faces and $n = 104$ for female faces for Correlation 1. Correlation 2 data were calculated on a smaller subset of faces for which there were no missing data and whose babyfaceness and attractiveness were judged by different raters. Sample sizes for these correlations ranged from 14 to 56. Chld = childhood; Pub = puberty; Adol = adolescence.
* $p < .05$. ** $p < .001$.

others of the same age. These procedures were adopted so that judgments would not be influenced by a tendency to rate individuals as attractive or babyfaced simply because they were high on this quality relative to their peers rather than in absolute terms.

Second, raters were told that they should think of babyfaceness as the extent to which the face in some way resembles a baby's face and that judging babyfaceness isn't necessarily the same as judging age. This procedure was used because past research has shown that when faces of varying ages are judged on babyfaceness without such instructions, babyfaceness ratings are indistinguishable from age ratings (Friedman, 1991). The intent of the instructions was to unconstrain the relationship between babyfaceness ratings and age so that if a relationship were obtained it could not be due to demand characteristics.

A third methodological change from Study 1 was that each rater judged both male and female faces. This procedure was adopted so that ratings would reflect implicit comparisons between the sexes, thus providing a more sensitive test of the hypothesis that there is a double standard of aging.

A fourth difference from Study 1 was that only those subjects who had a slide at each of the five age levels were rated, and only one slide was used at each age level. This procedure was adopted so that raters could judge all faces in one experimental session. Finally, ratings of facial expression were obtained for use in the data analyses to control for the possibility that changes in the absolute levels of babyfaceness or attractiveness reflected systematic age-related changes in expression (Hildebrandt, 1983; McGinley, Blau, & Takai, 1984; Mueser, Grau, Sussman, & Rosen, 1984; Schulman & Hoskins, 1986).

Method

Judges

Sixteen male and sixteen female undergraduates enrolled in introductory psychology classes rated slides for $5 and partial credit toward a course requirement. Half of the subjects of each sex rated attractiveness, and half rated babyfaceness. Ratings were generally done in groups of 8.

Subjects

The 56 male and 50 female subjects for this study were the subset of subjects from Study 1 who had slides available at each of the five age levels. The subjects included 37 males and 36 females from the Berkeley Guidance sample and 19 males and 13 females from the Oakland Growth sample. One female subject was subsequently dropped from the analyses because of missing ratings. The age of the subject, which appeared on each slide in Study 1, was occluded in Study 2. Also unlike Study 1, each age level was represented by only one slide for each subject. The age selected within those levels with multiple ages was the one with the largest number of subjects available—Childhood 3, Puberty 3, and Adolescence 1. Subjects selected in this manner ranged in age from 9.5 to 62 and represented five age levels: childhood ($M = 10.58$, $SD = 0.22$), puberty ($M = 14.98$, $SD = 0.27$ for boys; $M = 13.06$, $SD = 0.15$ for girls), adolescence ($M = 17.05$, $SD = 0.35$), the 30s ($M = 30.87$, $SD = 1.47$), and the 50s ($M = 56.8$, $SD = 3.73$).

Dependent Measures

Each slide was rated for both attractiveness and babyfaceness using the same 7-point scales used in Study 1. Ratings of facial expression were also obtained for use in the data analyses as a control variable. Specifically, two raters (one man and one woman) rated the amount of smiling displayed in each slide on 5-point scales with the endpoints labeled *no smile* and *big smile.* Two raters (one man and one woman) rated the facial expression displayed in each slide on 5-point scales with the endpoints *sad* and *happy* and the midpoint *neutral.*

Procedure

Judges were told that the purpose of the study was to examine whether judgments of appearance were related to personality traits, and they were asked to rate 530 slides of male and female faces of various ages on attractiveness or babyfaceness. As in Study 1, each slide was shown for approximately 6 s, and the slides were shown in one of two orders. Order 1 presented faces randomly with respect to sex and age, and Order 2 was the reverse of Order 1. Unlike Study 1, raters were not shown the ages of the faces, nor were they instructed to judge each face relative to others of the same age. Raters judging babyfaceness were told that they should think of babyfaceness as the extent to which the face in some way resembles a baby's face and that judging babyfaceness isn't necessarily the same as judging age.

Results

Reliability of Ratings

Attractiveness and babyfaceness. The correlations between the mean ratings of each face by male and female judges were quite high. For attractiveness, the correlation was .75 for all male faces and .77 for all female faces. For babyfaceness, the correlation was .91 for male faces and .88 for female faces. Therefore, reliability was assessed after combining ratings from both the male and female judges.

Reliability coefficients for attractiveness and babyfaceness were calculated separately for faces of each age and sex. The alpha coefficient for attractiveness ratings averaged across age groups was .82 for male faces and .87 for female faces. The alpha coefficient for babyfaceness ratings averaged across age groups was .73 for both male and female faces. Given these reliabilities, mean ratings of attractiveness and mean ratings of

babyfaceness were calculated from the ratings of all judges for each face at each age and used in subsequent data analyses.

Facial expression. Both expression ratings showed high interrater agreement (.76 for happy–sad and .79 for degree of smiling), and these ratings were consequently averaged across raters. Because happy–sad and smiling ratings were positively related within each age level, with an average correlation of .82, these ratings were summed to create a positive facial expression composite for use as a covariate in subsequent analyses.[3]

Absolute Stability

As a first step in determining whether it was necessary to use facial expression as a control variable in the analyses assessing the stability of babyfaceness and attractiveness, correlations were computed between the positive facial expression composite and babyfaceness and attractiveness ratings. For male subjects, there was a positive relationship between babyfaceness and a positive facial expression at all ages, an effect that was significant in adolescence, the 30s, and 50s (rs = .37, .48, and .50, respectively; all ps < .01) and marginal in childhood and puberty (both rs = .24, ps < .10). For female subjects, there was a positive relationship between babyfaceness and a positive facial expression in childhood, the 30s, and 50s (rs = .42, .45, and .49, respectively; all ps < .01), but not in puberty or adolescence (ps > .25). Male attractiveness was augmented by a positive facial expression in childhood (r = .26, p = .05), but not at other ages (all ps > .15), and female attractiveness was augmented by a positive facial expression in adolescence (r = .24, p < .10) and the 30s and 50s (rs = .33 and .30, both ps < .05), but not in childhood or puberty (ps > .25).

Having found that facial expression is significantly correlated with babyfaceness and attractiveness, a 2 × 5 (Subject Sex × Age Level) analysis of variance was performed on the positive facial expression composite to determine whether facial expression showed systematic variations across subject age or sex. Subject sex served as a between-groups variable in this analysis, and subject age level was a repeated measure. The means and standard deviations for facial expression are reported in Table 9. The results yielded no effect for sex (F < 1). However, there was a significant main effect for age, $F(4, 412) = 20.57$, $p < .001$, and a significant Sex × Age interaction, $F(4, 412) = 4.17$, $p < .01$. These effects revealed a tendency for facial expression to be less positive in the 30s than at other ages for male subjects and to be more positive in the 50s than at other ages for both male and female subjects, an effect that reflected less standardized poses in the adulthood photographs. Given these age- and sex-related variations in the positivity of facial expression and the correlations between expression and appearance ratings, covariance analyses, controlling for expression, were used to examine the absolute stability of babyfaceness and attractiveness.

Two 2 × 5 (Subject Sex × Age Level) analyses of covariance (ANCOVAs) were performed on attractiveness and babyfaceness ratings, with facial expression as the covariate. The adjusted means and standard deviations for attractiveness and babyfaceness are reported in Table 9.[4]

Attractiveness did not vary significantly with subject sex, $F(1, 102) = 2.36$, $p = .13$, but it did vary with age level, $F(4, 411) = 39.78$, $p < .001$. Planned comparisons between adjacent age levels revealed a steady decline in attractiveness across the life span, with a significant decrease from childhood to puberty, $t(104) = 2.66$, $p < .01$; a marginally significant decrease from puberty to adolescence, $t(104) = 1.84$, $p < .10$; and significant decreases from adolescence to the 30s as well as from the 30s to the 50s, ts(104) = 6.76 and 3.89 (both ps < .01). The interaction between sex and age level was not significant, $F(4, 411) = 1.45$, $p = .22$, and the means provided no hint of a tendency for attractiveness to decline more with age for women than for men.

Babyfaceness also did not vary with subject sex, $F(1, 102) = 1.14$, $p = .29$, but varied with age, $F(4, 411) = 854.76$, $p < .001$. Planned comparisons between adjacent age levels revealed that, as predicted, babyfaceness decreased significantly from childhood to puberty, $t(104) = 18.39$, $p < .001$, and from the 30s to the 50s, $t(104) = 14.17$, $p < .001$. Contrary to expectation, babyfaceness also decreased between puberty and adolescence and between adolescence and the 30s, ts(104) = 12.92 and 26.34, ps < .001. In addition to these age-related decreases in babyfaceness, which held true both for both male and female subjects, there was a significant Sex × Age interaction, $F(4, 411) = 5.55$, $p < .001$. This effect reflected a tendency for male subjects to be more babyfaced than female subjects in childhood, $t(103) = 4.25$, $p < .001$, whereas female subjects were more babyfaced in puberty, the 30s, and the 50s, ts(103) = 2.59, 2.77, and 4.44, respective ps < .05, .01, and .001.

Discussion

Attractiveness does not show absolute stability across the life span, as reflected in significant decreases in attractiveness with increasing age. Contrary to the popular notion that there is a double standard of aging for women versus men, this decrease in attractiveness was not more pronounced for women. Because the only research that has documented more deleterious effects of aging on women's attractiveness has used a cross-sectional design, past findings (e.g., Deutsch et al., 1983; Milord, 1978) may simply reflect the particular samples of men and women who were rated at younger versus older ages. Although the present data clearly show that aging does not diminish the attractiveness of women more than that of men, it should be noted that the same absolute decrease in attractiveness may have more adverse social consequences for women than it has for men because of the higher valuation placed on womens' appearance (e.g., Buss, 1989; Kenrick, Trost, & Keefe, 1992).

Like attractiveness, babyfaceness did not show absolute stability. Rather there were significant decreases in mean levels of babyfaceness across the life span. As predicted, babyfaceness was lower at puberty than in childhood. Contrary to prediction, an age-related decrease in babyfaceness was also observed from puberty to adolescence and from adolescence to the 30s. Finally, the finding that people were less babyfaced in their 50s than in their 30s suggests that facial signs of aging that augment

[3] This composite was also used in Study 1 for the differential and structural stability analyses partialing out the effects of facial expression.

[4] The unadjusted means showed the same changes across age in attractiveness and babyfaceness as the adjusted means.

Table 9
Facial Expression, Attractiveness, and Babyfaceness as a Function of Sex and Age

	Childhood		Puberty		Adolescence		30s		50s	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Positive expression										
Male faces	3.12	1.16	3.39	0.97	3.11	0.87	2.57	0.48	3.52	0.87
Female faces	3.22	0.90	2.89	0.69	2.98	0.63	2.79	0.63	3.78	0.91
Attractiveness[a]										
Male faces	3.67	0.63	3.56	0.51	3.52	0.55	3.09	0.47	2.84	0.44
Female faces	3.54	0.80	3.38	0.78	3.24	0.67	3.03	0.59	2.91	0.51
Babyfaceness[a]										
Male faces	5.53	0.52	4.61	0.51	4.16	0.49	3.03	0.46	2.42	0.47
Female faces	5.30	0.58	4.75	0.49	4.16	0.42	3.18	0.51	2.66	0.54

Note. $n = 56$ for male faces and $n = 49$ for female faces.
[a] Ratings are adjusted means from covariance analysis controlling for facial expression.

babyfaceness, such as less angular jaws and double chins, were outweighed by those that reduce babyfaceness, such as hooding of the eyes and wrinkling of the skin. Because the balance between these two sets of changes may shift as people grow older, it would be interesting to compare the babyfaceness of people when they are in their 50s versus their 70s or 80s.

In addition to an age-related linear decline in babyfaceness for both male and female subjects, there was a Sex × Age interaction. Ten-and-one-half-year-old girls were more mature faced than boys of the same age, a finding that may reflect the fact that the girls are closer to puberty. The finding that female subjects were more babyfaced than male subjects during puberty, the 30s, and the 50s is consistent with morphological differences between the mature male and female faces. The unexpected absence of a sex difference at adolescence could reflect a tendency for the adolescent judges to rate faces of their peers relative to other faces of the same sex rather than to others of both sexes.

A positive facial expression increased the babyfaceness of male subjects at all ages and female subjects in childhood, the 30s, and the 50s. The popular view of children as happy-go-lucky may underlie this unanticipated association of facial immaturity with the expression of positive affect. A positive expression also increased the attractiveness of female subjects from adolescence through the 50s, but increased the attractiveness of male subjects only in childhood. This finding is consistent with other evidence that smiling affects the attractiveness of women more than men (Schulman & Hoskins, 1986). Despite these relationships between facial expression and appearance ratings and the variations in expression across age and sex, the observed age and sex differences in babyfaceness and attractiveness were not mediated by differences in facial expression.

Conclusions

The present findings have important implications for theories that hold that physical appearance can influence personality development by virtue of eliciting stable social expectations. The finding that both attractiveness and babyfaceness show differential stability across a wide age range indicates that a necessary condition for such influence is met—there is stability of appearance vis-à-vis one's peers. Moreover, evidence for convergence in appearance ratings made 30 years apart by very different samples of judges indicates that reactions to an individual's appearance will be consistent across perceivers as well as across time. Although appearance showed significant differential stability, appearance at one age level was by no means a perfect predictor of appearance at the next. Thus, some people who were high in attractiveness or babyfaceness at one age were low in these qualities at the next and vice versa. An interesting question for future research is whether individual differences in the degree of appearance stability predict differences in personality stability.

Not only were there individual differences in the stability of appearance at a given age transition, but also the average level of differential stability varied across transitions. Specifically, the facial changes associated with puberty and with aging diminished the differential stability of babyfaceness at these times, albeit not that of attractiveness. The destabilizing effects of facial maturation at puberty were observed for male subjects, whereas female subjects seemed to show destabilizing effects of aging in the 30s, and both sexes showed it in the 50s. An implication of these findings is that babyfaceness has a greater potential to influence personality development from childhood to puberty for female than for male subjects, whereas it has a greater potential to influence personality development from adolescence to the 30s for male than for female subjects.

Despite declines in the differential stability of male subjects' babyfaceness at puberty and in the 50s, the stability of this facial quality never dipped significantly below that of attractiveness, which suggests that babyfaceness and attractiveness have the potential for equally strong effects on mens' personality development. For women, on the other hand, attractiveness may have a stronger impact than babyfaceness on personality development from adolescence to the 30s, when attractiveness showed significantly higher stability.

The structural and absolute stability findings also bear on theories that posit an impact of appearance on personality development. First, any effects of attractiveness and babyfaceness on personality development are likely to be independent, because the correlations between these two appearance qualities were at most moderate and often nonsignificant. Second, the

documented differences in the absolute levels of babyfaceness and attractiveness across the life span could conceivably contribute to differences in behavior among various age groups by virtue of fostering different social expectations and interactions. Third, the greater babyfaceness of women as compared with men could contribute to sex-differentiated behavior by eliciting different social reactions to men and women (cf. Friedman & Zebrowitz, 1992).

References

Adams, G. R. (1977). Physical attractiveness research: Toward a developmental social psychology of beauty. *Human Development, 20,* 217–239.

Adams, G. R., & Huston, T. L. (1975). Social perceptions of the middle-aged varying in physical attractiveness. *Developmental Psychology, 11,* 658–675.

Alley, T. R. (1988). The effects of growth and aging on facial aesthetics. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces* (pp. 51–62). Hillsdale, NJ: Erlbaum.

Alley, T. R. (1993). The developmental stability of facial attractiveness: New longitudinal data and a review. *Merrill-Palmer Quarterly, 39,* 265–278.

Berkowitz, L., & Frodi, A. (1979). Reactions to a child's mistakes as affected by her/his looks and speech. *Social Psychology Quarterly, 42,* 420–425.

Berman, P. W., O'Nan, B. A., & Floyd, N. (1981). The double standard of aging and the social situation: Judgments of attractiveness of the middle-aged woman. *Sex Roles, 7,* 87–96.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin, 100,* 3–18.

Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 157–215). San Diego, CA: Academic Press.

Bull, E., & Rumsey, N. (1988). *The social psychology of facial appearance.* New York: Springer-Verlag.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12,* 1–49.

Caspi, A., & Bem, D. (1990). Personality continuity and change across the life course. In L. A. Pervin (Ed.), *Handbook of personality theory and research* (pp. 549–575). New York: Guilford Press.

Caspi, A., Elder, G. H., & Bem, D. J. (1987). Moving against the world: Life-course patterns of explosive children. *Developmental Psychology, 23,* 308–313.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression-correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cross, J. F., & Cross, J. (1971). Age, sex, race, and the perception of facial beauty. *Developmental Psychology, 5,* 433–439.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology, 50,* 925–935.

Cunningham, M. R. (1990). What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology, 59,* 61–72.

Deutsch, F. M., Clark, M. E., & Zalenski, C. M. (1983, April). *Is there a double standard of aging?* Paper presented at the 54th Annual Meeting of the Eastern Psychological Association, Philadelphia.

Dion, K. K. (1972). Physical attractiveness and evaluation of children's transgressions. *Journal of Personality and Social Psychology, 24,* 207–213.

Dion, K. K. (1974). Children's physical attractiveness and sex as determinants of adult punitiveness. *Developmental Psychology, 10,* 772–778.

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.

Eichorn, D. H. (1981). Samples and procedures. In D. H. Eichorn, J. A. Clausen, N. Haan, M. P. Honzik, & P. H. Mussen (Eds.), *Present and past in middle life* (pp. 33–51). San Diego, CA: Academic Press.

Felson, R. B. (1980). Physical attractiveness, grades, and teachers' attributions of ability. *Representative Research in Social Psychology, 11,* 64–71.

Friedman, H. (1991). *Effects of facial maturity and attractiveness on occupational attainment.* Unpublished manuscript, Brandeis University, Waltham, MA.

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

Gray, H. (1985). *Anatomy of the human body* (20th ed.). Philadelphia: Lea & Febiger.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore: Johns Hopkins University Press.

Henss, R. (1989). *Perceiving age (and attractiveness) in facial photographs.* Unpublished manuscript, Arbeiten der Fachrichtung Psychologie, Universitat des Saarlandes, Saarbrücken, Germany.

Hildebrandt, K. A. (1983). Effect of facial expression variations on ratings of infants' physical attractiveness. *Developmental Psychology, 19,* 414–417.

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

Jackson, L. A. (1992). *Physical appearance and gender: Sociobiological and sociocultural perspectives.* Albany, New York: SUNY Press.

Johnson, D. F. (1985). Appearance and the elderly. In J. A. Graham & A. M. Kligman (Eds.), The psychology of cosmetic treatments (pp. 152–160). New York: Praeger.

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

Jöreskog, K., & Sörbom, D. (1989). *LISREL VII. Analysis of linear structural relationships by the method of maximum likelihood, instrumental variables, and least squares methods.* Mooresville, IN: Scientific Software.

Keating, C. (1985). Gender and the physiognomy of dominance and attractiveness. *Social Psychology Quarterly, 48,* 61–70.

Kenrick, D. T., Trost, M. R., & Keefe, R. C. (1992, July). *Integrating evolutionary and social exchange perspectives on relationships.* Paper presented at the Joint Meeting of the European Association of Experimental Social Psychology and the Society for Experimental Social Psychology. Leuven/Louvain-la-Neuve, Belgium.

Langlois, J. H. (1986). From the eye of the beholder to behavioral reality: The development of social behaviors and social relations as a function of physical attractiveness. In C. P. Herman, M. P. Zanna, & E. T. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario Symposium* (Vol. 3, pp. 23–51). Hillsdale, NJ: Erlbaum.

Langlois, J. H., & Stephan, C. W. (1981). Beauty and the beast: The role of physical attractiveness in the development of peer relations and social behavior. In S. S. Brehm, S. M. Kassin, & F. X. Gibbons (Eds.), *Developmental social psychology* (pp. 152–168). New York: Oxford University Press.

Lerner, R. M., & Lerner, J. V. (1977). Effects of age, sex, and physical attractiveness on child-peer relations, academic performance, and elementary school adjustment. *Developmental Psychology, 13,* 585–590.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Maruyama, G., & Miller, N. (1981). Physical attractiveness and personality. In B. Maher (Ed.), *Progress in experimental personality research* (Vol. 10, pp. 203–280). San Diego, CA: Academic Press.

McArthur, L. Z., & Apatow, K. (1983–1984). Impressions of babyfaced adults. *Social Cognition, 2,* 315–342.

McGinley, H., Blau, G. L., & Takai, M. (1984). Attraction effects of smiling and body position: A cultural comparison. *Perceptual and Motor Skills, 58,* 915–922.

Milord, J. T. (1978). Aesthetic aspects of faces: A (somewhat) phenomenological analysis using multidimensional scaling methods. *Journal of Personality and Social Psychology, 36,* 205–216.

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

Mueser, K. T., Grau, B. W., Sussman, S., & Rosen, A. J. (1984). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology, 46,* 469–478.

Pittenger, J. B., Mark, L. S., & Johnson, D. F. (1989). Longitudinal stability of facial attractiveness. *Bulletin of the Psychonomic Society, 27,* 171–174.

Ritter, J. M., Casey, R. J., & Langlois, J. H. (1990). Adults' responses to infants varying in appearance of age and attractiveness. *Child Development, 62,* 68–82.

Salvia, J., Algozzine, R., & Sheare, J. B. (1977). Attractiveness and school achievement. *Journal of School Psychology, 15,* 60–66.

Schulman, G. I., & Hoskins, M. (1986). Perceiving the male versus the female face. *Psychology of Women Quarterly, 10,* 141–154.

Snyder, M., Tanke, E., & 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.

Sorrell, G. T., & Nowak, C. A. (1981). The role of physical attractiveness as a contributor to individual development. In R. M. Lerner & N. S. Busch-Rossnagel (Eds.), *Individuals as producers of their development: A life span perspective* (pp. 389–446). San Diego, CA: Academic Press.

Sparacino, J. (1980). Physical attractiveness and occupational prestige among male college graduates. *Psychological Reports, 47,* 1275–1280.

Sternglanz, S. H., Gray, J. L., & Murakami, M. (1977). Adult preferences for infantile facial features. An ethological approach. *Animal Behavior, 19,* 233–250.

Sussman, S., Mueser, K. T., Grau, B. W., & Yarnold, P. R. (1983). Stability of females' facial attractiveness during childhood. *Journal of Personality and Social Psychology, 44,* 1231–1233.

Tanner, J. M. (1962). *Growth at adolescence.* Oxford, England: Blackwell Scientific.

Walsh, R. P., & Locke, C. (1980, November). *Perceptions of womens' physical attractiveness across the life cycle.* Paper presented at the 33rd Annual Meeting of the Gerontological Society, San Diego, CA.

Zebrowitz, L. A., Kendall-Tackett, K., & Fafel, J. (1991). The influence of children's facial maturity on parental expectations and punishments. *Journal of Experimental Child Psychology, 52,* 221–238.

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' babyfacedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior, 15,* 603–623.

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

Received November 18, 1991
Revision received September 24, 1992
Accepted September 30, 1992 ■