# Longitudinal stability of facial attractiveness

JOHN B. PITTENGER
*University of Arkansas at Little Rock, Little Rock, Arkansas*

LEONARD S. MARK
*Miami University, Oxford, Ohio*

and

DOUGLAS F. JOHNSON
*University of Arkansas at Little Rock, Little Rock, Arkansas*

Profile facial photographs of 12 white males and 12 white females, each at six different ages, were rated on scales of perceived attractiveness, intelligence, and sociability. Each of the six age-level sets of photographs was rated by a different group of 18 male and 18 female white college students. Multivariate analysis of variance and correlational analyses showed a moderate degree of longitudinal stability of perceived facial attractiveness for the faces of most people in the stimulus sample. Some people, however, showed a progressive decline in perceived attractiveness. Additional analyses compared statistical properties of the three social judgments and tested for effects of sex of rater and stimulus person.

People have clear stereotypes of the characteristics and behaviors expected of individuals of particular degrees of attractiveness (Berscheid, 1980). These stereotypes appear to affect behavior toward a person: The influence of a person's facial attractiveness on his/her social interactions with others has been well-documented. In addition, the expectations of others can influence the individual's own behavior. Long-term effects of these expectations could be mediated by their influence on the person's self-concept: We may come to see ourselves as others see us. However, internalization of expectations based on attractiveness assumes a degree of stability over time sufficient for consistent expectations to be communicated to the individual (Berscheid & Walster, 1974).

Three experiments, one by Sussman, Mueser, Grau, and Yarnold (1983) and two by Adams (1977), have assessed the stability of facial attractiveness. In each experiment, longitudinal photographs of children's faces were presented to college students for attractiveness ratings. Correlational (Adams, 1977) and ANOVA analyses (Sussman et al., 1983) showed moderate, but not perfect, stability of attractiveness. The experiment reported here extends this earlier work by presenting faces of male and female children to male and female raters. Prior studies did not use the two sexes both as stimulus persons and

as raters. In addition, our stimulus set included a wider range of ages in the stimulus faces, and was therefore able to assess stability across nearly the full range of growth-produced changes in the face. Finally, we collected ratings of perceived intelligence and sociability, in addition to the ratings of attractiveness. This allowed comparison of the stability of attractiveness with that of other social perceptions.

## METHOD

### Stimuli

For each of 28 people, six black and white profile facial photographs were selected from the files of the Denver Child Research Council Longitudinal Growth Study (McCammon, 1970). Pictures were selected so that the faces in the series for each individual ranged from early childhood to the late teens, contained a fairly uniform progression of ages, showed no facial hair or glasses, and portrayed neutral, relaxed expressions. As discussed below, although the photographs of all 28 individuals were presented for ratings, the data for 4 of them were dropped from the analysis because of apparent influence of expressions or lighting on the ratings. The median ages of faces in the six age-level sets, based on the remaining 24 people (12 male and 12 female), are 3 years, 10.5 months; 6, 8.5; 8, 8.5; 11, 1.5; 13, 0; and 16, 7.5. The photographs were masked to hide as much clothing and hair as possible. The resultant pictures were rephotographed and made into slides.

### Stimulus Sets

Three random orders of the stimuli people were devised. Each random order was applied to each of the 6 age-level sets, thereby generating 18 different stimulus sets.

### Participants

Raters were 216 Caucasian introductory psychology students at the University of Arkansas at Little Rock who participated for course credit. Each stimulus set was rated by 6 males and 6 females.

### Rating Scales

Participants marked their ratings of the faces on forms printed with 28 numbered blocks of three scales, one for each dimension tested:

attractiveness, intelligence, and sociability. Each scale consisted of seven short lines with labels at each end. The labels at the left of the three scales were *not attractive*, *not intelligent*, and *not sociable*. On the right, the labels were *attractive*, *intelligent*, and *sociable*. Participants indicated their ratings on each dimension by marking the most appropriate of the seven lines.

## Procedure

Small groups of participants viewed the faces projected onto a movie screen. For the attractiveness judgments, the participants were instructed to mark the rightmost line of the scale for especially good-looking people and the leftmost for especially unattractive people. The intermediate lines were to be marked for people intermediate in attractiveness. The other two scales were to be used in a similar fashion. The instructions for intelligence judgments included the statement that although true intelligence is not generally visible in the face, people do vary in how intelligent they appear. Participants were asked to base their ratings on the appearance of intelligence. Sociability was defined for participants as an inclination to seek the company of others and to be friendly and agreeable toward them. Participants were shown the first 10 faces and asked to rate these faces without marking the forms. Participants then viewed the full set of faces and marked their ratings on the forms. Each face was presented for about 10 sec.

## RESULTS

Ratings were scored as integers from 1 (*least favorable*) to 7 (*most favorable*). Mean attractiveness ratings were computed for each stimulus person at each age level. The photographs of stimulus persons that showed extreme fluctuations in attractiveness were inspected for unusual expressions, artifacts in lighting, and so forth. In four cases such factors seemed to account for the fluctuations. All data from these stimulus persons were dropped from further analysis.

## Longitudinal Stability

**Multivariate analysis of variance.** A multivariate analysis of variance, using the three social judgments as dependent variables, showed significant effects of age level and stimulus person, and of their interaction [Wilk's criterion $F$s are, respectively, $F(15,574.60) = 3.18$, $p < .001$; $F(69,14424.34) = 17.44$, $p < .001$; and $F(345,14482.54) = 4.32, p < .001$]. For each judgment, there was a significant ($p < .001$) main effect of stimulus person and interaction between stimulus person and age level. A main effect of age level was found only for ratings of attractiveness ($p < .001$). Table 1 presents the mean ratings by age level and judgment.

The significant stimulus person × age level interactions suggest a type of instability in social perception: The temporal pattern of change in social ratings varies among individuals. The interaction, however, is weak in that it accounts, in the analysis of the attractiveness ratings, for only 15.7% of the variance. Age level and stimulus person accounted for 11.3% and 16.6%, respectively.

Inspection of the mean attractiveness ratings for individual stimulus persons plotted against age in years reveals the nature of the age level effect and the age level × stimulus person interaction. About one-third of the plots show a fairly consistent decrease in rated mean attractiveness with age, whereas about one-third to one-half

#### Table 1
#### Mean Ratings for the Social Judgments by Age Level

| Judgment | Age Level | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Attractiveness | 3.70 | 3.85 | 3.52 | 3.40 | 3.30 | 3.22 |
| Intelligence | 4.11 | 4.28 | 4.06 | 3.96 | 4.17 | 4.11 |
| Sociability | 3.98 | 4.21 | 4.06 | 3.91 | 4.02 | 3.96 |

Note—Each mean is based on 864 data points (36 raters × 24 pictures).

#### Table 2
#### Mean Correlations Between Ratings at Different Age Levels

| Judgment | Step Size | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Attractiveness | .53 | .43 | .41 | .26 | .19 |
| Intelligence | .51 | .35 | .49 | .31 | .24 |
| Sociability | .26 | .34 | .43 | .24 | .32 |

Note—See text for the definition of *step size*.

appeared to be fairly constant in rated attractiveness. For several individuals, attractiveness ratings are sawtooth functions of age. No clear cases of increasing, U-shaped, or inverted U-shaped patterns are evident.

**Correlational analyses.** Correlational techniques provided another assessment of longitudinal stability. For each stimulus person, age level, and social judgment, mean ratings were computed over raters. Then, for each social judgment, product-moment correlations were computed between every pair of age levels over stimulus persons. Table 2 presents the mean correlations by step size for each social judgment. *Step size* refers to the difference in the two age levels used in the correlation. For example, correlations between age levels 1 and 2 and between age levels 2 and 3 each have a step size of 1, whereas those between levels 1 and 3 and between levels 2 and 4 have step sizes of 2.

The coefficients are positive and moderately large. Also, there is a consistent decrease in the magnitude of the coefficients for attractiveness ratings as a function of increasing step size. This pattern might have been expected on the grounds that step sizes are correlated with age differences in the faces and, thus, the degree of change in facial structure produced by growth. Perceived sociability and intelligence, however, show no indication of an effect of step size.

Judgments on the three dimensions were highly correlated with each other. At each age level and for each pair of dimensions, a correlation coefficient was computed over the individual stimulus persons' mean ratings. The resulting coefficients were all large and positive, ranging from .72 to .93.

## Agreement Among Participants

**Cronbach alphas.** The amount of agreement among raters on social perceptions was assessed by calculating Cronbach alpha statistics for the ratings on each social dimension at each age level. All coefficients were greater than +.85, indicating a high amount of agreement among

raters. There also were no apparent differences in the amount of interrater agreement among the three judgments across age level.

**Standard deviations.** At each age level and for each judgment, standard deviations were computed for each stimulus person over raters. These numbers provide another index of how well raters agreed about the appearance of a particular face on a particular social dimension, with low standard deviations indicating high agreement. The mean standard deviations in Table 3 show that attractiveness ratings were more variable in the middle age levels than at either the lower or higher levels. Also, attractiveness ratings were least variable and sociability ratings were most variable.

In summary, although the existence of differences in the amount of agreement among age levels and judgments is problematical, all three judgments can be made very reliably at all age levels.

## Effects of Sex of Stimulus Persons and Participants

The effect of sex of participants was tested as a main effect in a separate analysis for each dimension. Effects of stimulus persons were tested via three sets of contrasts made on male versus female stimulus persons: one over male participants, one over females, and one with the sexes combined (see Table 4).

Female participants gave significantly higher ratings of attractiveness [$F(1,204) = 4.80, p < .05$] and intelligence [$F(1,204) = 4.61, p < .05$], but not of sociability [$F(1,204) = 3.38, p > .05$], than did males. None of the contrasts between male and female stimulus persons, either for the means collapsed over sex of participants or separately for male and female participants, were significant at the .05 level. Finally, none of the age level × sex of participants interactions were significant.

## DISCUSSION

### Longitudinal Stability of Attractiveness

The main effect of stimulus person on the attractiveness ratings indicates a degree of stability in at least some of the stimulus people; however, the stimulus person × age level interaction indicates a form of instabil-

#### Table 3
#### Mean Standard Deviations of Ratings of Stimulus Persons

| Judgment | Age Level | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Attractiveness | 1.23 | 1.33 | 1.51 | 1.37 | 1.37 | 1.27 |
| Intelligence | 1.43 | 1.33 | 1.50 | 1.46 | 1.39 | 1.40 |
| Sociability | 1.53 | 1.47 | 1.56 | 1.51 | 1.45 | 1.46 |

#### Table 4
#### Mean Ratings by Sex of Participant

| Judgment | Sex of Participant | |
|---|---|---|
| | Male | Female |
| Attractiveness | 3.41 | 3.59 |
| Intelligence | 4.02 | 4.21 |
| Sociability | 3.95 | 4.10 |

ity. It is impossible to be certain of the source of this interaction: Both real physical changes in facial structure produced by growth and transient factors such as expressions are captured in the photographs and can be expected to influence the ratings. There are, however, some reasons to suppose that both factors influenced the data reported here.

Since the continuous processes of craniofacial growth would be expected to lead to smooth and rather gradual change in attractiveness, the zigzag patterns found for the ratings of some individuals are likely to be the result of transient factors. However, there are enough relatively smooth plots in the present sample to suggest that some individuals do show strongly constant attractiveness and that others show a progressive loss of attractiveness. This mix of smooth and irregular progessions of means is also evident in the results reported by Sussman et al. (1983).

The correlations among ratings of individuals across age levels are positive and moderately strong, indicating a reasonable degree of stability in attractiveness (see also Adams, 1977). The correlations reported here are somewhat lower than those reported by Adams, a result that is consistent with the fact that our sample of faces involved a greater age span than did Adams's sets.

Overall, our findings are in agreement with those of Adams (1977) and Sussman et al. (1983). They extend the previous results in that a wider range of ages of stimulus faces was used and both males and females served as raters and were represented in the stimulus faces.

### Other Aspects of Age and Attractiveness

Since growth progressively changes facial structure, we expected that correlations of attractiveness ratings would decline as the age levels used in the calculation became more widely separated. This prediction was confirmed.

Standard deviations for individual stimulus people, computed across raters, tended to be higher in the middle years. This finding may have reflected differences in the standards applied to people of different ages. As a result, there might be some variation among raters as to which standards are applicable to people in late childhood. This is consistent with the observed pattern of standard deviations. However, since Cronbach alphas show no temporal pattern, firm conclusions cannot be made.

### Comparisons Among the Social Judgments

The high correlations among the three judgments might have been produced by the methods used in this study: All three judgments of a face were made at the same time. If judgments along each dimension had been made in different presentations, the correlations might have been considerably lower. Thus, these results cannot be used to assess the separability of the dimensions. However, in everyday life, we make many simultaneous judgments of others. Therefore, even if judgments on different social dimensions can be separated under some experimental methods, high correlations may be common outside of the laboratory.

The standard deviations presented in Table 4 suggest that agreement among participant's ratings is highest for attractiveness, lower for intelligence, and lowest for sociability. This pattern is consistent with the authors' experiences in judging the three dimensions and with participants' comments made after the ratings were collected. Specifically, although it is easy and natural to judge the attractiveness of a face, judgments of apparent intelligence and sociability generally seem rather forced and difficult.

### Conclusions

We draw three major conclusions from the present results. First, the attractiveness of many individuals is moderately stable over time and could support incorporation of other people's expectations into their self-images. For others, however, attractiveness appears to decline. This would lead to changing expectations, and possibly to difficulty in forming a stable self-image. Second, at least when made simultaneously, ratings along the three dimensions are highly correlated.

Finally, progress toward better understanding of longitudinal aspects of social perception may be limited by weaknesses in available stimulus materials. We selected photographs from the archives of a large collection, which, for the purpose of documenting physical growth, were well controlled. Although we tried to exercise the utmost care in our stimulus selection and preparation, we were unable to fully exclude

artifacts due to lighting and facial expressions from our original stimulus set. Definitive answers to questions about longitudinal social perception require stimulus materials that are well controlled with respect to perceptual variables. Neither family photographs nor those in longitudinal growth studies attempt to meet this need. In addition, it may not be possible to achieve an emotional neutral expression in static photographs. It may be that people's actual appearance can be captured dynamically only on film or videotape.

## REFERENCES

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

BERSCHEID, E. (1980). An overview of the psychological effects of physical attractiveness. In G. W. Lucker, K. A. Ribbens, & J. A. McNamara, Jr. (Eds.), *Psychological aspects of facial form* (pp. 1-23). Ann Arbor, MI: Center for Human Development.

BERSCHEID, E., & WALSTER, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 155-215). New York: Academic Press.

McCAMMON, R. W. (1970). *Human growth and development*. Springfield, IL: Thomas.

SUSSMAN, S., MUESER, K. T., GRAU, B. W., & YARNOLD, P. R. (1983). Stability of females' facial attractiveness during childhood. *Journal of Personality & Social Psychology*, **44**, 1231-1233.