Cross- vs. within-racial judgments of attractiveness

IRA H. BERNSTEIN, TSAI-DING LIN, and PAMELA McCLELLAN
University of Texas, Arlington, Texas

Subjects from two pairs of ethnic groups (Chinese and White in Experiment 1, Black and White in Experiment 2) judged the attractiveness of faces in yearbook pictures of persons that belonged to their own or to the other ethnic group. This was to see whether: (1) a given group would perceive more variation in the attractiveness of faces belonging to its own vs. the other ethnic group, as suggested by the cross-racial literature, for example, Malpass and Kravitz (1969), and (2) the two groups would use the same or different rules to define attractiveness. There were essentially no differences in perceived variation for cross- vs. within-racial judgments, but there were differences in the criteria used to define attractiveness. As expected, Black and White aesthetic criteria were more like one another than were Chinese and White criteria. Discussion centered around reconciling these findings with the recognition literature.

Slightly more than a decade ago, Malpass and Kravitz (1969) reported that accuracy of cross-racial identifications is poorer than accuracy of within-racial identifications. In accord with stories familiar to all, the result is commonly taken to imply a greater apparent (perceived) homogeneity of the faces of unfamiliar group members. This finding has had enormous social ramifications, most obviously in legal settings (Loftus, 1979).

Various related question involve how judgments other than recognition may differ when made cross-racially. In particular, many theories of memory postulate that information about specific items, whether they be words or pictures, is encoded in the form of attributes (Bower, 1967; Underwood, 1969). A particular attribute which seems to play a major role in the perception of faces is attractiveness (Walster, Aronson, & Abrams, 1966). The specific question that one may pose is whether judgments of attractiveness are subject to influences that are analogous to those found in recognition. The term "analogous" is used because encoding of attributes like attractiveness necessarily involves a subjective element not present when recognition accuracy is considered.

There are four possible outcomes of special interest that might arise when members of two different

racial groups are asked to judge the attractiveness of members of one of these groups. One is that there simply are no judgmental differences among groups. This would imply two quite separate things about the judgments. The first is that the perceived variation in the stimuli (faces) cross- and within-racially is the same, in apparent contrast to the case when specific identifications are attempted. The second is that the physical attributes used to define attractiveness would also be the same. Thus, if one group preferred people with large eyes, so would the other.

One possible alternative outcome is that the two groups could perceive equal variation in the faces but could attend to different attributes. Thus, one group might pay primary attention to the eyes and secondary attention to the mouth and lips, whereas the other group might do the reverse. Another alternative is that there is greater apparent homogeneity of the faces of the unfamiliar group, but no difference in the attributes used to define attractiveness. That is, when people judge members of their own racial group, differences along the criterial dimensions are greater than when people judge a different racial group, but there are no systematic differences in the criterial dimensions (nose, eyes, etc.) that are chosen. Finally, differences may exist in both perceived variations and criterial dimensions used—people judging an unfamiliar group could perceive less variation among members of this group as well as use different dimensions to define attractiveness. Based upon the identification literature, either of the latter two outcomes would seem more probable, a priori, than the former two.

These four outcomes reflect two independent issues—whether or not there are differences in variation perceived for one's own vs. another racial

group, that is, whether or not there is a difference in perceived variation among faces, and whether or not the rules used to translate this variation among faces to a single dimension of judgment are the same for the two groups. One way to investigate these issues is simply to have members of race A and members of race B rate a common set of pictures belonging to race A.

Table 1 contains hypothetical data obtained from two members of race A (A_1 and A_2) and two members of race B (B_1 and B_2) in the form of the intercorrelations among them (interjudge reliabilities) over the ratings. Outcome A, *equal* perceived variation and the *same* judgmental rules, implies that the intercorrelations among the judges will be homogeneous within sampling error, that is, $r_{A_1A_2} = r_{A_1B_2} = r_{A_1B_2}$, etc. In contrast, outcome B, *equal* perceived variation but *different* rules, implies that both pairs of within-racial correlations would be the same but higher than the cross-racial correlations, which in turn would be homogeneous, that is, $r_{A_1A_2} = r_{B_1B_2} > r_{A_1B_1} = r_{A_1B_2}$, etc. Outcome C, *less* perceived variation cross-racially but the *same* judgmental rules, implies that the two judges of race A will corre-

late more highly with each other than will the two judges of race B with each other, but the cross-racial correlations would be limited solely by this disparity in within-racial correlations. Elementary psychometric considerations (Nunnally, 1978) can specify this limitation quantitatively. Finally, outcome D, *less* perceived variation and *different* judgmental rules, implies that the cross-racial correlations would be lower than can be accounted for by the disparity in within-racial correlations.

Other outcomes are, of course, possible but are less readily explained. Moreover, no actual study would be limited to two pairs of raters, and the data obtained would probably be noisier. Differences in both perceived variation and judgmental criteria could exist between two members of the same race as well as between members of different races.

Despite these complications, the joint issues of possible cross-racial differences in perceived variation and judgmental criteria are readily explored. The particular approach that seems most direct and which we have chosen here is a form of factor analysis known as *confirmatory factor analysis* (Nunnally, 1978). In confirmatory factor analysis, one defines various sets of a priori factors and compares their ability to explain a given intercorrelation matrix, such as those portrayed in Table 1, with alternative models.

To be specific, it is readily verified that the four outcomes of Table 1 correspond to different factor structures. Outcome A requires but a single factor, on which the two races contribute (load) equally. Outcome B requires two correlated factors, which correspond to the two races, each explaining equal amounts of variance in the correlation matrix. Outcome C requires a single factor, as does outcome A, but one on which race A loads more strongly than race B. Finally, outcome D requires two factors, one for each race, such that the factor defined by race A explains more variance than does the factor defined by race B.

EXPERIMENT 1

Experiment 1 is primarily concerned with the attractiveness ratings made by White American and Chinese college students. Both White American and Chinese pictures were judged. Various models were evaluated along the general lines suggested in the previous paragraph, although, as will be noted, certain factor-analytic considerations complicate the testing slightly. In subsidiary analyses, differences in use of the rating scale regarding mean attractiveness and variability were also examined.

Method

Subjects. Thirty White American and 30 native Taiwanese[1] subjects, divided equally by sex, participated in the experiment. The Americans were undergraduates who received course credit.

Table 1
Four Different Hypothetical Intercorrelations Matrices
Generated by Two Judges of Race A (A_1 and A_2) and
Two Judges of Race B (B_1 and B_2) Who All Rate a
Common Set of Faces Belonging to Race A

	A_1	A_2	B_1	B_2
		A		
A_1	.8			
A_2	.8	.8		
B_1	.8	.8	.8	
B_2	.8	.8	.8	.8
		B		
A_1	.8			
A_2	.8	.8		
B_1	.5	.5	.8	
B_2	.5	.5*	.8	.8
		C		
A_1	.8			
A_2	.8	.8		
B_1	.7	.7	.6	
B_2	.7	.7	.6	.6
		D		
A_1	.8			
A_2	.8	.8		
B_1	.3	.3	.6	
B_2	.3	.3	.6	.6

Note—The off-diagonal elements of each matrix represent interjudge reliabilities; the diagonal elements represent communalities in the factor-analytic sense. (A) Races A and B use the same judgmental criteria in judging race A and perceive equal stimulus variation (i.e., are equally coherent). (B) Races A and B use different judgmental criteria in judging race A but perceive equal stimulus variation. (C) Races A and B use the same judgmental criteria in judging race A, but race A perceives more stimulus variation than race B. (D) Races A and B use different judgmental criteria in judging race A, and race A also perceives more stimulus variation than race B.

The Chinese subjects were undergraduate and graduate volunteers. None of the Chinese subjects had been in America for more than 2 years; all had come to the United States for the first time to study at the University of Texas at Arlington or North Texas State University.

Stimuli. Four sets of 100 pictures each were obtained from college yearbooks containing White male, White female, Chinese male, and Chinese female faces. The White pictures were obtained from a 1975 University of Texas at Arlington yearbook. The Chinese pictures were obtained from a 1974 National Taiwan University yearbook. The Chinese and White pictures differed in ways other than race. The White pictures were smaller (1×3 cm) than the Chinese pictures (3×4 cm) and represented a wider variety of camera angles, hair styles, facial expressions, and dress. We selected only those pictures of individuals who appeared to be in their middle to early twenties. The pictures were cut and pasted onto individual 3×5 index cards.

Procedure. The subjects were run individually or in small groups by coauthor T.D.L., who is a bilingual Taiwanese national. The subjects were given a sheet that explained the rating procedure in their preferred language; this was followed by oral instructions. They were then given the four sets of pictures at random. The subjects rated each face on a 7-point rating scale, with the low end as positive.

Results

Exploratory factor analyses. The results of exploratory factor analyses are necessary before the major hypotheses can be tested. Thus, principal-component factor analyses were performed on each of the four individual picture sets (White males, White females, Chinese males, and Chinese females). The analyses are exploratory because they do not directly test specific structures, such as those described above. These analyses do provide three results of direct relevance: (1) the maximum percentage of total variance explainable by a given number of factors (dimensions), (2) the number of factors that are "important" in the factor analytic sense of explaining as much variance as an average variable (judge), and (3) the percentage of total variance that the resulting set of important factors account for. In addition, these exploratory analyses furnish communality estimates or percentages of each judge's total variance that is shared with the remaining judges. These communality estimates are not directly relevant to this paper, but they do play an important indirect role in correcting the correlations between factors for the idiosyncracies (unique variance) of the judges.

The rows marked "One Principal Component" and "Two Principal Components" in Table 2 describe the maximum amount of variance that can be explained by one and two factors, respectively, and serve as a frame of reference for the theoretical models to be discussed. At most, one factor can account for 26.8% to 37.9% of the variance of the four sets and two factors can account for 34.5% to 45.9% of the variance. The number of factorially "important" dimensions (factors with eigenvalues ≥ 1.0) are 15, 14, 16, and 14 for White male, White female, Chinese male, and Chinese female picture sets, respectively. These account for 71.4%, 72.0%, 73.5%, and 74.0%, respectively, of the total variance in judgments. Thus, it is clear that judgments of attractiveness differ in numerous ways besides those that might reflect racial differences.

In addition to the principal component (unity diagonals) analyses reported, principal-axis solutions (communality estimates in the diagonals of the various matrices) were conducted as well, but the results were highly similar.

Confirmatory analyses, variance accounted for. There are several methods for testing factor solutions (Nunnally, 1978).[2] The oblique multiple group centroid approach was chosen over Joreskog's (1969) LISRL for its simplicity of programming and efficiency in use of computer resources. The computer program was written by the senior author. Details on the factoring method may be found in Harmon (1976, pp. 234-242). For applications to empirical problems, see Bernstein and Eveland (1982).

Several alternative models were examined. The simplest is to postulate one centroid and weigh each judge the same, thereby ignoring race differences.

Table 2
Proportions of Variance Accepted for by Alternative Models in Experiment 1

| Model | Type of Picture | | | |
	White Males	White Females	Chinese Males	Chinese Females
One Principal Component	.268	.320	.301	.379
One Centroid	.267	.319	.300	.378
Two Principal Components	.345	.387	.351	.459
White + Chinese	.338	.383	.339	.456
White Group Factor	.229	.282	.389	.343
Chinese Group Factor	.235	.284	.280	.323
Random	.294	.336	.323	.401
Random Factor 1	.258	.311	.287	.366
Random Factor 2	.254	.312	.292	.366

Next, one may postulate one factor for White raters and a second for Chinese raters. This will be termed the White + Chinese model. A third, random, model consists of two factors formed by assigning subjects randomly to factors regardless of race and sex, subject to the constraint that no more than eight members of each sex × race combination be assigned to a factor. We shall be reporting the first of several such random models tested. The variance among these models was used as an inferential device. Differences discussed below all greatly exceed the variance in results of alternative random-factor models. This model is necessary because two-factor models will generally explain sample data better than a single-factor model, as the former tend to capitalize upon chance more.

Table 2 contains the variances explained by the individual factors of the various models and by the set of factors within each model taken jointly. These results were obtained with unities in the diagonals of the correlation matrix, that is, by performing component analyses. The results were substantially the same using a common-factor approach (communality estimates in the diagonals).

The two main findings are that in every picture set: (1) the White + Chinese model accounts for nearly as much variance as the first two principal components and 1.6% to 5.1% more than the random model, implying that this model accounts for the maximum variance possible in two dimensions, and (2) the separate White and Chinese factors within this model account for essentially the same amount of variance. Jointly, these findings for race are consistent with outcome B of Table 1. Note that there is no tendency for the White factor to account for more variance than the Chinese factor with White pictures or vice versa for the Chinese pictures.

There are two minor findings commonly obtained in multivariate applications. First, the centroid and one principal component model differ only slightly. This means that taking individual differences other than race into account makes very little difference. Second, the random model's two factors account for more variance than either single-factor model, illustrating how multiple factors capitalize upon chance.

As may be seen, all conclusions held across the four picture sets. The *equivalence* of results does not mean that the sets produced *identical* results. More variance was explained within the female and Chinese sets relative to the male and White sets, respectively, which implies that there was a greater consensus among the former.

Confirmatory factor analyses, factor correlations. Table 3 contains the correlations between the two factors of the White vs. Chinese model (top) and the random model (below). "Raw" correlations are derived from the component analyses presented above, whereas "corrected" correlations are derived from common-factor solutions and attempt to remove the attenuation produced by the unique portion of the variation in a judge's ratings. The main point is that while the correlations between White and Chinese factors are high, with the raw correlations ranging from .638 to .801 across picture sets, they are much lower than the correlations between randomly generated factors. This is another way of saying that the two groups' judgmental criteria were separable.

There was not a consistent difference in magnitude of correlation between White and Chinese or male and female pictures.

Sex differences. Following this examination of racial differences, we performed parallel factor analyses to look for sex differences. In fact, differences between males and females fit outcome A of Table 1 in terms of perceiving equal variation and using the same criteria to define attractiveness.

Rating means and standard deviations. The rating means and standard deviations for individual subjects, derived within each of the four groups of judges from the pictures within each of the four sets, were analyzed within a two between-subjects variables (race of judge and sex of judge) and two within-subjects variables (race of picture and sex of picture) mixed model. As indicated above, race of

Table 3
Factor Correlations for Alternative Models: Raw (Component) and Corrected for Attenuation (Common Factor) in Experiment 1

Model	Type of Picture			
	White Males	White Females	Chinese Males	Chinese Females
	White vs. Chinese			
Raw	.638	.715	.801	.695
Corrected	.655	.734	.824	.709
	Random			
Raw	.904	.948	.922	.927
Corrected	.938	.976	.950	.948

Table 4
Cell Means (C) of Individual Subject Means (S) and Standard Deviations (SD) in Experiment 1

Judge's Sex	Judge's Race		Race and Sex of Picture							
			White				Chinese			
			Male		Female		Male		Female	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Male	White	C	4.32	.75	3.60	.64	4.67	.84	4.46	.55
		S	1.27	.29	1.42	.30	1.21	.25	1.42	.33
Female	White	C	4.46	.73	4.04	.64	4.86	1.06	4.48	.84
		S	1.45	.23	1.38	.24	1.23	.47	1.44	.34
Male	Chinese	C	4.14	.50	3.79	.50	4.36	.48	4.01	.41
		S	1.05	.25	1.07	.20	.98	.24	1.06	.22
Female	Chinese	C	4.54	.80	4.14	.84	4.63	.82	4.20	.80
		S	1.09	.28	1.12	.30	.98	.23	1.00	.24

Note— Each C mean is the mean of 15 subjects' individual means; each S mean is the mean of 15 subjects' individual standard deviations.

picture is confounded with pose and size. Even though these data are only tangentially relevant to the issues raised in the introduction (since the process of computing correlations eliminates both mean and variability differences in the use of the rating scale across both pictures and judges), they are of relevance to the more general question as to how people of one ethnic group view members of their own and other ethnic groups.

Table 4 contains the cell means of the individual subject means and standard deviations, along with the standard deviations of these statistics.

Analysis of variance (ANOVA) of the means revealed three significant effects: (1) race of picture, $F = 30.28$, $p < .001$, MSe = .2167, (2) sex of picture, $F = 46.57$, $p < .001$, MSe = .2118, and (3) race of picture × race of judge, $F = 9.19$, $p < .01$, MSe = .2167 (all df = 1 and 56). Pictures of Whites (mean = 4.12) and females (mean = 4.09) were rated more favorably (lower) than pictures of Chinese (mean = 4.45) and males (mean = 4.50). White judges rated White pictures much more favorably than Chinese pictures (means = 4.10 and 4.61), but Chinese judges rated the two picture races nearly the same (means = 4.15 and 4.30).

ANOVAs of the standard deviations revealed five significant effects: (1) race of judge, $F = 23.96$, $p < .001$, MSe = .2372, (2) race of picture, $F = 11.31$, $p < .001$, MSe = .0224, (3) sex of picture, $F = 9.85$, $p < .01$, MSe = .0233, (4) race of picture × sex of picture, $F = 6.16$, $p < .02$, MSe = .0232, and (5) the four-way interaction of race of judge, sex of judge, race of picture, and sex of picture, $F = 4.02$, $p < .05$, MSe = .0232 (all df = 1 and 56). White judges ($\bar{s} = 1.23$, where $\bar{s}$ denotes the average of individual standard deviations) spread their ratings more than did Chinese judges ($\bar{s} = 1.17$), pictures of Whites ($\bar{s} = 1.23$) evoked more varied ratings than did pic-

tures of Chinese ($\bar{s} = 1.17$), and pictures of females ($\bar{s} = 1.24$) evoked more varied ratings than did pictures of males ($\bar{s} = 1.16$). However, the ratings of White female and White male pictures varied about the same ($\bar{s} = 1.25$ and 1.21), but pictures of Chinese males less evoked varied ratings than pictures of Chinese females ($\bar{s} = 1.23$ and 1.10). The four-way interaction is probably spurious; an ANOVA of the variances of the ratings, that is, a square transformation, revealed this effect to be nonsignificant.

Thus, although White raters tend to spread their ratings on the scale more than they do Chinese raters, their ratings are no more systematic, as seen in the factor analyses. Likewise, although there were differences in variability of the ratings to the two sets of male faces which were in the direction expected from the greater similarity of pose and dress of the Chinese, they did not translate into differences in correlations, since the correlations in the Chinese sets are actually higher than in the White sets.

EXPERIMENT 2

Experiment 2 employed the identical design but involved White and Black American subjects and pictures. Given the results of Experiment 1 and the greater familiarity that Blacks and Whites have with each other compared with that of Americans and Chinese, it was expected that the resulting Black and White factors would be more highly correlated than the Chinese and White factors were in Experiment 1.

Method

Subjects. Thirty White and 30 Black American subjects, divided equally by sex, participated in the experiment. The Whites and most of the Blacks were undergraduates receiving course credit, but some of the Blacks were recruited as volunteers.

Stimuli. Four sets of 100 pictures each were obtained from college yearbooks containing White male, White female, Black male, and Black female faces. The White pictures were obtained from a 1975 University of Texas at Arlington yearbook. The Black pictures were obtained from the same source and from the 1980 yearbook of Bishop College, a primarily Black institution in Dallas. We selected only those pictures of individuals who appeared to be in their early to middle twenties and who did not wear glasses. The pictures were 1 × 3 cm and were pasted onto individual 3 × 5 cards.

Procedure. The subjects were run individually or in small groups by coauthor T.D.L., a native Taiwanese who also had run Experiment 1. The same procedure was followed.

Results

Exploratory factor analyses. The preliminary principal component analyses on the four data sets indicated that one factor can account for 27.5% to 45.4% of the data and two factors can account for 32.8% to 50.2%. The precise values may be found in Table 5, which parallels Table 2. In essence, these results indicate that the degree of consensus was higher in this study than in Experiment 1, especially for the Black females. Given the greater familiarity that the subjects had with one another's faces, this was precisely as expected. There were 17, 14, 14, and 11 components with eigenvalues greater than 1.0 in the White male, White female, Black male, and Black female picture sets. These accounted for 74.1%, 71.2%, 74.1%, and 71.7%, respectively, of the total variance in judgments.

Confirmatory factor analyses, variance accounted for. Table 5 indicates that the White + Black model accounts for more than the random model. These differences, which range from 1% with White male pictures to 4% with Black male pictures, are smaller than those observed in Experiment 1, but, in conjunction with the factor correlation data to be presented below, are sufficiently great to establish the separability of White and Black judgmental criteria.

The biggest disparity between Black and White factors was 2.1% in the case of Black male faces. The other disparities were 1.2% or less, and in the "wrong" direction for White female faces (the Black factor accounted for more variance than the White

factor). Considering how overall judgmental consistency (percent variance accounted for) varied across the four picture sets, the difference is clearly trivial.

The variation among picture sets is of the same form as that observed in Experiment 1 in that there was higher agreement found when people, regardless of sex or ethnicity, judged "minority" groups. Another way to describe this is that stereotypy as to attractiveness exists to a greater extent among these groups.

As we noted earlier, the centroid and one principal component factor yield similar results, and the random-factor model's two factors account for more variance than the single-factor models.

Confirmatory analyses, factor correlations. Table 6 contains the factor correlations. Except for the pictures of Black males, these do, in fact, run higher than the corresponding entries of Table 3, which is in accord with expectation. In other words, although Black and White ratings are separable, they are more similar to one another than are Chinese and White ratings, reflecting the hypothesized greater cultural dissimilarity of the latter two groups.

Sex differences. Confirmatory analyses of possible sex differences were also conducted, paralleling the analyses performed in Experiment 1. As was true in that study, none were found; that is, differences between male and female judges were of the same magnitude as randomly generated differences.

Rating means and standard deviations. Table 7 contains the cell means and standard deviations of individual ratings with the standard deviations of these statistics, paralleling Table 4.

The ANOVA of the means revealed three significant effects: (1) race of picture, $F = 30.28$, $p < .001$, $MSe = .2167$, (2) sex of picture, $F = 46.57$, $p < .001$, $MSe = .2118$, and (3) race of picture × race of judge, $F = 9.18$, $p < .01$, $MSe = .2167$ (all $df = 1$ and 56). White (mean = 4.24) and female (mean = 4.11) pictures were rated lower (more favorably) than Black (mean = 4.51) and male (mean = 4.64) pictures. White judges rate White pictures much more favorably than Black pictures (means = 4.02 and 4.46), but the

Table 5
Proportions of Variance Accounted for by Alternative Models in Experiment 2

Model	White Males	White Females	Black Males	Black Females
One Principal Component	.275	.351	.381	.454
One Centroid	.274	.350	.380	.454
Two Principal Components	.328	.402	.441	.502
White + Black	.308	.396	.437	.494
White Group Factor	.258	.322	.330	.426
Black Group Factor	.246	.324	.351	.432
Random	.297	.371	.397	.472
Random Factor 1	.264	.340	.370	.444
Random Factor 2	.260	.339	.371	.446

Table 6
Factor Correlations for Alternative Models: Raw (Component) and Corrected for Attenuation (Common Factor) in Experiment 2

Model	Type of Picture			
	White Males	White Females	Black Males	Black Females
	White vs. Black			
Raw	.82	.82	.76	.88
Corrected	.93	.84	.77	.90
	Random			
Raw	.91	.93	.95	.96
Corrected	.99	.97	.97	.98

differences were smaller, although, perhaps surprisingly, in the same direction, for Black judges (means = 4.47 and 4.62).

The ANOVA of the standard deviations revealed four significant effects: (1) race of picture, $F = 20.20$, $p < .001$, MSe = .0300, (2) sex of picture, $F = 9.76$, $p < .01$, MSe = .0479, (3) sex of picture × sex of judge, $F = 14.02$, $p < .001$, MSe = .0479, and (4) race of picture × race of judge × sex of judge, $F = 4.11$, $p < .05$, MSe = .300 (all df = 1 and 56). Blacks ($\bar{s} = 1.37$) and female ($\bar{s} = 1.36$) pictures evoked more varied ratings than White ($\bar{s} = 1.27$) and male pictures ($\bar{s} = 1.28$). The difference between standard deviations of male and female pictures was relatively large for male judges ($\bar{s} = 1.20$ and 1.39) but very small for female judges ($\bar{s} = 1.35$ and 1.34). The triple interaction arose because there was a very small Black vs. White difference when females judged White pictures ($\bar{s} = 1.30$ and 1.30) or when males judged Black pictures ($\bar{s} = 1.35$ and 1.34), but there were larger differences when males judged White pictures ($\bar{s} = 1.20$ and 1.28) or when females judged Black pictures ($\bar{s} = 1.35$ and 1.44).

Common to the two experiments (viewing both Chinese and Blacks as "minorities" within the context of a primarily White American university) were four findings: (1) White pictures were rated more favorably, even when minorities did the judging, (2) female pictures were rated more favorably than male pictures, (3) the difference between ratings of White and minority pictures were greater when Whites were judges than when minorities were judges, and (4) pictures of females evoked more varied ratings than did pictures of males. However, Chinese pictures evoked less varied ratings than White pictures (possibly due to the selection of the Chinese stimuli), but Black pictures evoked more varied ratings than White pictures. The mean differences do not necessarily reflect how minorities "really" perceive Whites, and, of course, neither these nor the standard deviation data bear upon the covariances among raters, which are of main concern to this paper.

In summary, the overall results of Experiment 2 are quite similar to those of Experiment 1. The various ethnic groups do use slightly different standards of attractiveness, but they are all capable of applying these rules to other groups as well as to their own.

DISCUSSION

The results clearly fit the second of the four models described in Table 1. That is, judges from differ-

Table 7
Cell Means (C) of Individual Subject Means (S) and Standard Deviations (SD) in Experiment 2

Judge's Sex	Judge's Race		Race and Sex of Picture							
			White				Black			
			Male		Female		Male		Female	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Male	White	C	4.25	.51	3.85	.61	4.54	.45	4.49	.69
		S	1.19	.37	1.36	.30	1.24	.36	1.44	.37
Female	White	C	4.16	.60	3.81	.62	4.43	.50	4.17	.53
		S	1.35	.30	1.25	.35	1.44	.33	1.45	.34
Male	Black	C	4.82	.89	4.07	.59	4.82	.83	4.17	.69
		S	1.11	.37	1.30	.28	1.25	.36	1.46	.29
Female	Black	C	4.85	.68	4.11	.61	5.25	.61	4.23	.81
		S	1.30	.38	1.30	.32	1.34	.30	1.36	.32

Note— Each C mean is the mean of 15 subjects' individual means; each S mean is the mean of 15 subjects' individual standard deviations.

ent ethnic groups perceive the *same degree* of variation in rating faces for attractiveness although they use *different criteria*. This is true whether the pictures are of Whites, so that they are relatively unfamiliar to Chinese, or vice versa, and holds for Whites and Blacks, who are relatively more familiar with each other.

Initially, there was reason to believe that Model C or, especially, Model D would have fit the data, that is, that there would have also been differences in perceived variation due to differential familiarity. This expectation was based upon the literature on cross-racial recognition. There are several reasons one might suspect our failure to find a difference. One is simply that the processes involved in judging attractiveness are different from those involved in making specific identifications. However, it would seem premature to abandon the hope for a single, parsimonious explanation, since there are ways to unify the two judgmental processes, as we will note below.

A second possibility is that even the familiarity of the present White and Chinese with each other was sufficient to eliminate any difference in perceived variation. This would not be the case for more clearly isolated groups, such as native Taiwanese with no prior direct exposure to Whites. However, the degree of familiarity that these two racial groups had with one another was probably less than the pairs of groups used in most accuracy studies, especially those comparing Whites and Blacks, who share so many experiences, including American television. The Whites and Chinese were clearly more different from one another than the Whites and Blacks of Experiment 2. Also, differences between both pairs of cultures were sufficient to produce a difference in criteria for attractiveness (although it could be argued that such are maintained after differences in perceived variation dissipate with familiarity). Moreover, many of the Chinese students reported difficulties in identifying White faces, so that they are at least as appropriate a sample as others previously used.

A third possibility is that we chose an inappropriate attribute to have the judges rate, in the sense that attractiveness is insensitive to differences in familiarity. However, there is both informal as well as formal data to support the basic nature of attractiveness ratings in interpersonal evaluations (Walster, Aronson, & Abrams, 1966). Of course, it is quite possible to determine this empirically by looking at ratings for other attributes such as friendliness or intelligence, but this is beyond the scope of the present study.

Reconciling our findings with the identification literature requires a further consideration of what can give rise to difference in criteria for cross- vs. within-racial attractiveness judgments. There are at least two ways that differences in the cues (attributes, dimensions) of judgment can have implications for identifications.

One explanation is that people of race B use fewer cues in judging members of race A than do people of race A in judging members of race A. Interjudge reliabilities could be similar for the two groups of judges if each used similar cues among themselves. For example, if judges of race B all rated tall faces as attractive and disregarded everything else, they could be just consistent as judges of race A, who would use a richer set of cues. However, members of race B would be more inclined to identification errors, since they would classify two people with tall faces whose eye separations were different as similar, whereas members of race A would pay attention to both attributes. If one were to use linear regression as a device to emulate their judgments, the equations for cross-racial attractiveness would have a larger set of predictors, each with relatively small beta weights. However, similarities between equations derived from pairs of judges in race A and equations derived from pairs of judges in race B could still be parallel.

Alternatively, judges rating other races could still respond to as wide a range of cues as judges rating members of their own race, yet still show an identification decrement if the former tended to rely on easily modified or contextual cues. In this case, cross- and within-racial regression equations would look the same as far as number of beta weights and their magnitudes went, but the specific predictors would be different. We are all aware of presumed racial differences that are really cultural, especially hair styles, which are prominent cues and yet are quite easily altered. Because we tend to habituate to characteristics of our own culture, it is possible that we pay more attention to the more biologically invariant features of people belonging to our culture. This explanation involves type rather than number of cue.

Deciding whether either factor is relevant could in principal be done with real life stimuli like those used in this study. However, their complexity is a serious disadvantage. Consequently, we are using schematic faces in our present research to control the cues that are varied.

REFERENCES

BERNSTEIN, I. H., & EVELAND, D. C. State vs. trait anxiety: A case study in confirmatory factor analysis. *Personality and Individual Differences*, 1982, 3, 361-372.

BOWER, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press, 1967.

HARMON, H. H. *Modern factor analysis* (3rd. ed., rev.). Chicago: University of Chicago Press, 1976.

Joreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, **34**, 183-202.

Loftus, E. F. *Eyewitness testimony*. Cambridge, Mass: Harvard University Press, 1979.

Malpass, R. S., & Kravitz, J. Recognition of faces of own and other races. *Journal of Personality and Social Psychology*, 1969, **13**, 330-334.

Nunnally, J. C. *Psychometric theory*. New York: McGraw-Hill, 1978.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559-573.

Walster, E., Aronson, V., & Abrams, D. Importance of physical attractiveness in dating behavior. *Journal of Personality and Psychology*, 1966, **4**, 508-516.

NOTES

1. The two largest groups of ethnically Chinese, foreign-born students at the University of Texas at Arlington and other American universities are from Taiwan and Hong Kong. Culturally, these groups are fairly distinct, because the latter are raised in a society with a heavy British influence. We limited our subjects to native Taiwanese because of their lesser experience in seeing White faces.

2. The present study involves a transposed or Q-type factor analysis, since correlations are computed between people over stimuli rather than the converse. Nunnally (1978, pp. 426-429) points out how problems may arise in a Q-type analysis if differences in stimulus means and standard deviations reflect arbitrary scale differences. This is not the case here, since the stimuli within sets were selected in an essentially random manner, paralleling subject selection in conventional factor analytic designs. Thus, differences in means and standard deviations for pictures reflect natural variations in attractiveness and do not contain arbitrary scale differences.