

Effects of training on Japanese face recognition: Reduction of the other-race effect

ALVIN G. GOLDSTEIN and JUNE E. CHANCE
University of Missouri, Columbia, Missouri

Subjects with demonstrated poor memory for Japanese faces were either intensively trained to learn Japanese face-digit pairs or assigned to a no-training control condition. Immediately following training, and at 1 and at 5 months later, trained subjects showed improved recognition when tested with new sets of Japanese faces. The no-training controls, who were tested on the same schedule, showed improvement only on the 5-month test. Data are discussed in relation to schema development and the "other-race" effect.

The phrase "they all look alike to me" is so well known that it is often used as a joke. The phrase expresses the feeling that faces of "other-race" people appear to be much more homogeneous than the faces of the people from one's own group. Recent laboratory research (e.g., Chance, Goldstein, & McBride, 1975; Malpass & Kravitz, 1969; see Shepherd, 1981, for a review) has elevated this folk belief to the status of confirmed hypothesis (with a dissenter here and there, e.g., Lindsay & Wells, 1983). Briefly, research has demonstrated that white subjects' recognition memory for briefly seen unfamiliar white faces is superior to their memory for briefly seen black or oriental faces. Several studies have shown that black subjects remember unfamiliar black faces better than they remember unfamiliar white faces (e.g., Shepherd, Deregowski, & Ellis, 1974). Of the several explanations offered for the other-race effect, the most likely interpretation—differential experience with other-race faces—is also the most intuitively satisfying. The argument is straightforward. Remembering faces is a complex skill continuously developed and improved upon from infancy on. "Training" of this skill occurs without awareness over many years in normal day-to-day social interactions with people. Before reaching adulthood, a person living in an urban environment is exposed to hundreds of thousands of social interactions involving facial stimuli. For the large majority of the population in our society, these "learning trials" take place largely with own-race faces. Our relative inexperience with other-race faces appears in our relative inability to encode efficiently and process "foreign" faces. This deficit in experience is reflected in poor recognition memory performance for other-race faces, especially when encoding circumstances are atypical or unusually demanding (e.g., brief exposure, high similarity between target and foils).

Some evidence in support of the differential experience

This research was (reluctantly) supported in part by funds from the Research Council of the University of Missouri, Columbia, MO. We would like to express our gratitude to Brent Darnell for his help in collecting the data. Our mailing address is: Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211.

hypothesis is found in both everyday experience and laboratory research. Travelers to foreign countries often notice that many natives appear to be members of the same family, but after more exposure to the natives, they begin to individualize people and discrimination among their faces improves. The results of several experiments (e.g., Elliott, Wills, & Goldstein, 1973; Lavrakas, Buri, & Mayzner, 1976; Malpass, Lavigueur, & Weldon, 1973; see Malpass, 1981, for review) suggest that training improves other-race face recognition and has little effect on own-race recognition performance. However, since these studies assessed only immediate effects of training on performance, there remain the questions of whether the salutary effects of training last over appreciably longer periods of time and whether the increments in performance persist even though the trainee returns to a world mostly devoid of other-race faces? The present experiment assessed long-term effects of intensive paired-associate training on recognition of other-race faces. Essentially, we were interested in effects of "training aimed at erasing a deficit in recognition (improving individuals' recognition for faces of another 'racial' group)" (Malpass, 1981, p. 278). Furthermore, since we wished to determine whether the improved performance was transitory or whether it represented an enduring change in memory functioning—one that could be maintained without continued practice—our subjects were trained on Japanese faces, which are rarely encountered in their everyday environment. Finally, note that the purpose of the training sessions was to improve ability to learn general rules about Japanese faces, not to improve memory for particular Japanese.

METHOD

Subjects

Eight undergraduate women served as paid subjects. The subjects were each paid approximately \$5.00 per session for either three or nine sessions. Four of the women (the "trainees") were assigned randomly to the experimental group, and the remaining four served as the untrained control group.

Pretesting

The subjects were selected on the basis of their performance on two

screening tasks, one measuring white face recognition and the second measuring Japanese face recognition. On the basis of their screening test performance, eight subjects were recruited to serve as either trainees or control subjects. The subjects were selected because their recognition memory performance (d') on white faces was average or above while their d' scores on Japanese faces were well below their white d' scores (see Tables 1 and 2 for pretest d' scores). These selection criteria identified subjects who were attentive to the task and demonstrated normal memory for white faces, but who appeared deficient in recognition skills when faces were Japanese.

Training Trials

What kind of training would most effectively improve face recognition? Examination of the relevant literature produced little direct information, but one idea seemed especially salient, viz., that the quantity of exposure was perhaps less effective in modifying behavior than the quality of the interaction with other-race people. Following this lead, we provided our trainees with a challenging learning task embodying the elements of everyday social encounters while at the same time demanding active participation of the trainees. Therefore, the trainees were given intensive paired-associate training, with a Japanese face as one member of each pair and a digit as the other member. The stimulus pairs were Japanese portraits (1-in.-square achromatic photographs of college-age men and women) individually mounted on 5 × 7 in. cards, with an identifying number inscribed to the right of the photo. Depending on the particular trainee (see below), between 67 and 94 different Japanese faces were learned during the training sessions. Five [or more (see below)] face-digit pairs constituted each learning set. The cards were presented to each trainee manually by the experimenter. Before training began, each subject was familiarized with the paired-associate task format by learning a set of five white face-digit pairs in a procedure identical to the method to be used in the subsequent Japanese face training process. Trainees began each session by studying the paired associates of the first set one at a time for about 5 sec each. After the cards were reshuffled, the identifying digits were occluded and the trainee attempted to remember the correct digit for each portrait before the experimenter uncovered it. If during the first test trial the trainee's performance was error free, the experimenter reshuffled the cards and continued with a second test trial, instead of progressing to a study trial. If the trainee's performance was perfect, a third and final test trial would follow. However, the occurrence of a single error on any one of the test trials was reason to initiate another study trial and at the same time again require the learning criterion to be reset to three errorless test trials. Trial sequences continued until completion of three consecutive errorless trials had been achieved.

During each training session, which lasted between 30 and 45 min, the trainees learned two different sets of faces. The minimum number of faces in each set was five, but on several occasions, if the trainee were doing very well, as many as eight faces were included in each set. Before each training session was terminated, and after the trainee had reached criterion on both sets of faces, she was retested one more time on each set. Approximately 10 min separated the final test trial from the first retest trial. If no errors were committed on this retest, the training session was terminated. If one or more errors were committed, study-test sequences were again initiated until a trainee had completed three errorless trials. Training procedures were deliberately flexible to encourage self-pacing. Changes were permitted in the number of pairs constituting a learning set, in the speed of presentation of the individual photos, and in the rate at which testing trials proceeded, all depending on the trainee's wishes and the ease with which she acquired the face-number pairs. For example, at the trainee's request, she could be permitted to look simultaneously at two faces or to increase from five to eight the number of pairs in a study-test set. Each trainee was also permitted to use her own learning strategy; the experimenter never mentioned mnemonics nor suggested that a memorizing technique was needed.

In summary, although details of training necessarily differed slightly for each trainee, all four trainees learned between 67 and 94 face-number pairs in a total of six practice sessions distributed over a 2-3-week period. All faces were Japanese, and all were portraits of different individuals. Control subjects received no training.

Recognition Memory Testing

The purpose of the posttraining recognition memory testing was to assess whether paired-associate discrimination practice with Japanese had improved the trainees' performance with new sets of Japanese faces and whether, if improvement occurred, it was merely temporary or more long lasting. Since it seemed possible that just viewing a large number of Japanese faces in the recognition test procedure might improve performance, apart from any effect of the discrimination practice, control subjects who never had practiced were tested on the same schedule as the experimental subjects.

A standard face recognition memory testing procedure was used with all subjects. Target faces were presented for a few seconds in the study trial, and in the test trial that followed, "seen before" and "not seen before" decisions were made in response to each face as it appeared on the screen. The face recognition memory tests consisted of presentation of sets of white and Japanese faces; the tests were administered to the trainees on four occasions: one several days before the first practice trial (i.e., the screening test), and subsequent tests 2-4 days after training, 1 month after, and 5 months after the final training session. The control subjects were tested 1, 2, and slightly more than 5 months after their screening test. All portraits—that is, targets and foils—used in the recognition tests were new and had never been used before in any paired-associate training session. In the recognition memory screening tests (i.e., the "pretest") and in the first, second, and third post-training recognition tests, 10 faces served as targets in the study trial, and these 10, mixed with 22 foils, were presented in the following test trial, which occurred immediately after the study trial. In the fourth post-training test, 13 targets were used as study stimuli, and 30 distractors were added in the test trial. This fourth other-race recognition test (5 months after the final training session) differed from its predecessors in another way: Portraits of Chinese men and women were substituted for those of the Japanese. Prior research had suggested that Chinese and Japanese faces were indistinguishable from each other (Farnsworth, 1943, 1965).

RESULTS AND DISCUSSION

Evidence that the trainees improved in their ability to process Japanese faces can be seen first in the number of trials each trainee needed to attain criterion (3 errorless trials) during each of the six training sessions. In the first training session, the four trainees each viewed two sets of five face-digit pairs, and they took an average of 8 trials to reach the prescribed learning criterion. In their sixth and final training session, the trainees reached criterion on an average of nine face-digit pairs in only 7.1 trials. Thus, in their last training session, the trainees learned more (nine) face-digit associates in fewer trials than it had taken them to learn a smaller number (five) of associates in their first session. Even the poorest of the four trainees, who in her first session took 10 trials to learn five pairs, demonstrated improvement by learning seven pairs in 9 trials in her last training session. Since the trainees learned different Japanese faces at each training session, the most likely explanation for their improved performance over sessions is that they were developing a Japanese face schema (Goldstein & Chance, 1980); that is, they were learning some general strategies in relation to Japanese faces.

The hypothesis of this study also implies that, in addition to improving their paired-associate performance, the trainees' performance in recognition memory for briefly seen unfamiliar Japanese faces should improve over that in pretraining. Thus, comparison of the trainees' pretest

d' scores with their d' scores obtained on the first, second, and third posttraining tests should reveal whether what they learned during the paired-associate training transferred to the recognition memory tasks.

A 2 × 4 repeated-measures analysis of variance, with training as one factor and repeated recognition test scores as the second factor, was computed separately on white d' scores and on Oriental d' scores.¹ For each subject in the experiment, four d' scores were derived from the white face recognition pretest and from the three posttraining tests (see Table 1). Parallel d' measures were derived from the four Oriental face recognition tests (see Table 2).

The analysis of the white face recognition data clearly indicates that the overall average recognition memory performance of the trainees and the control subjects was indistinguishable [trainees mean = 2.25; control mean = 2.38; F(1,6) < 1.00]. There was no indication that the subjects improved their white face recognition performance over the four testing periods [pretest mean = 2.29; posttest mean 1 = 2.08; posttest mean 2 = 2.56; posttest mean 3 = 2.32; F(3,18) < 1.00]. Finally, the interaction between the training variable and testing periods was not significant [F(2,12) = 1.098, p > .05], suggesting that trainees, compared with controls, did not improve their white face recognition memory performance as a function of their training on Japanese faces.

Table 2 gives the d' scores for the trainees and for controls on each of the four recognition memory tests involving Oriental faces. First, these data show that both groups of subjects started the experiment with almost identical recognition memory handicaps; they were unable to recognize previously seen Japanese faces at better than chance levels of performance (trainees mean = -0.24 and control mean = -0.26). Statistical analysis (ANOVA) of the data of Table 2 revealed that, by the third posttest, the overall performance of the trainees became significantly better than the performance of the con-

Table 2
Recognition Memory Performance (d') of Trainees and Control Subjects on Oriental Faces

Subjects	Pretest	Posttest 1	Posttest 2	Posttest 3
Trainees				
1	-.16	1.34	1.76	2.63
2	-.16	3.03	3.16	3.72
3	-.16	1.86	2.18	2.82
4	-.48	3.60	2.18	3.28
Mean	-.24	2.46	2.32	3.11
Controls				
5	-.40	.36	-.12	3.16
6	+.22	.48	1.08	2.87
7	-.71	.92	-.23	2.87
8	-.16	.61	-.61	(2.97)*
Mean	-.26	+.59	+.03	2.97

*Subject 8 did not return for the final test; this value is the mean of the three control subjects d' scores on Posttest 3.

rol subjects [F(1,5) = 17.44, p < .01]. Although all subjects improved in performance across the four tests [F(3,17) = 63.06, p < .01], the significant interaction [F(3,17) = 11.72, p < .01] between the training variable and repeated measures of recognition memory clearly indicates that the rate of improvement differed in the two groups. The trainees profited almost immediately from their training sessions; they improved their mean test performance from -0.24 in the pretraining test to +2.46 in their first posttraining test. In contrast, the control groups' mean improvement from the pretest to the first posttest was quite small (from -0.26 to +0.59), and, more importantly, the magnitude of their d' scores suggest that they still cannot reliably discriminate "old" from "new" Japanese faces.

On the second posttraining test, 1 month after their final training session, the trainees' performance was almost unchanged from their first posttest performance. In contrast, over the same time period, the performance of the control subjects deteriorated from +0.59 to +0.03. In the third and final recognition test session, 5 months after the end of the training period, the trainees demonstrated, with an impressive average d' score of 3.11, that they had not lost their newly acquired skill in remembering Oriental faces. Unexpectedly, however, all three of the remaining members of the control groups also improved their performance [F(3,9) = 45.40, p < .01]. Further statistical analysis indicated that all their improvement occurred on the final recognition test (see notes to Tables 1 and 2). This puzzling and interesting result can be interpreted in two ways.

Although the control subjects received no training, they did view more than 140 Japanese faces in three recognition tests prior to the final test. It is possible that the subjects acquired the schema and strategies necessary to recognize other-race faces from repeated, spaced attempts to learn target faces, even though they received no explicit information about the accuracy of their responses.

Table 1
Recognition Memory Performance (d') of Trainees and Control Subjects on White Faces

Subjects	Pretest	Posttest 1	Posttest 2	Posttest 3
Trainees				
1	1.76	1.34	2.28	1.87
2	2.62	3.60	2.62	2.02
3	1.60	1.44	2.84	2.32
4	2.18	1.92	3.16	2.38
Mean	2.04	2.08	2.73	2.15
Controls				
5	2.20	1.60	1.92	3.27
6	1.92	2.28	2.28	1.78
7	3.66	1.86	2.36	2.21
8	2.36	2.62	3.03	(2.67)*
Mean	2.54	2.09	2.40	2.48

*Subject 8 did not return for the final test; this value is the mean of her three earlier d' scores.

It is also possible that, simply by chance, the last recognition test was comparatively "easy." Although the latter explanation is unlikely—the trainees' recognition performance on the identical set of faces was not much different from their performance on the preceding test—both explanations will be tested in future experiments.

In a recent critical review of the literature on face recognition training, Malpass (1981) concluded that although evidence concerning this topic was sparse, short-term improvement in other-race face recognition had been demonstrated in several studies using a variety of training procedures. The results of the present study add further information to the literature on face recognition training. First, we have shown that relatively intensive training with pairs of digits and Japanese faces over a period of several weeks substantially improves subjects' ability to learn additional pairs. Second, whatever is learned during the training can be transferred to a later recognition memory task and also enhances performance there significantly. Third, whatever is learned is retained without evidence of loss for a much longer period (5 months) than has up to now been found in other investigations, suggesting that subjects improved their Oriental face schema. And, finally, although much less secure than the first three conclusions, there is some evidence suggesting that face recognition performance also may be enhanced simply as a consequence of repeated attempts to remember and identify briefly seen faces. This finding is intriguing because it implies that face recognition improvement can take place in the absence of explicit information about right and wrong responses.

Although we have shown that other-race face recognition performance remains at a high level several weeks and months after the last training session, we have no way of determining what factor—or combination of factors—of our training regime was effective. Is it total number of faces, total number of trials, spacing of trials, the discrimination task? What? These questions remain to be answered in future research.

Because of the possibility of misunderstanding our results, it is important to emphasize the exact nature of the long-term effect. All performance test data in this investigation were collected on tasks that tested memory for faces over very brief retention intervals. At no time were subjects tested for retention of particular faces for periods longer than about 1 h. Thus, we do not at this time know whether the training would have improved the subjects' ability to remember a stranger's face over an extended period of time.

In conclusion, we have presented evidence that gains in performance attained at the end of the Japanese face recognition training can be maintained without measurable loss and without additional exposure to Japanese faces. This finding supports the hypothesis that the other-race effect is a performance deficit that reflects relative lack of experience with other-race faces.

REFERENCES

- CHANCE, J. E., GOLDSTEIN, A. G., & McBRIDE, L. (1975). Differential experience and recognition memory for faces. *Journal of Social Psychology*, *97*, 243-253.
- ELLIOTT, E. S., WILLS, E. J., & GOLDSTEIN, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society*, *2*, 71-73.
- FARNSWORTH, P. R. (1943). Attempts to distinguish Chinese from Japanese college students through observation of face-photographs. *Journal of Psychology*, *16*, 99-106.
- FARNSWORTH, P. R. (1965). A social effect on the perception of facial resemblance. *Journal of Social Psychology*, *65*, 221-223.
- GOLDSTEIN, A. G., & CHANCE, J. E. (1980). Memory for faces and schema theory. *Journal of Psychology*, *105*, 47-59.
- LAVRAKAS, P. J., BURI, J. R., & MAYZNER, M. S. (1976). A perspective on the recognition of other-race faces. *Perception & Psychophysics*, *20*, 475-481.
- LINDSAY, R. C. L., & WELLS, G. L. (1983). What do we really know about cross-race eyewitness identification? In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence*. New York: Wiley.
- MALPASS, R. S. (1981). Training in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 271-285). London: Academic Press.
- MALPASS, R. S., & KRAVITZ, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*, 330-334.
- MALPASS, R. S., LAVIGUEUR, H., & WELDON, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics*, *14*, 285-292.
- SHEPHERD, J. (1981). Social factors in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 55-79). London: Academic Press.
- SHEPHERD, J. W., DEREGOWSKI, J. B., & ELLIS, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology*, *9*, 205-212.

NOTE

1. In addition to the d' analyses, separate statistical analyses of hits and false alarms were also performed both on the raw data and on the white-Japanese difference scores. The results of all of these analyses essentially corroborated each other.

(Manuscript received for publication December 19, 1984.)