Facial recognition and the von Restorff effect

MICHELLE E. COHEN and W. J. CARR
Beaver College, Glenside, Pennsylvania 19038

The accuracy of recognition for faces of varying distinctiveness was investigated in this study. In Phase I, the subjects ranked 12 male faces from most to least distinctive. Interjudge agreement proved to be reliable. An analysis of the mean ranks assigned to the 12 faces showed them to be significantly different. In Phase II, other subjects responded to the 12 faces in a test of recognition. One week after inspection, distinctive faces produced significantly fewer errors of omission and commission than did nondistinctive faces. The findings suggest a von Restorff effect for faces.

Appropriate interactions among members of a group are facilitated if members respond to each other as individuals, as well as members of subgroups; e.g., male vs. female, immature vs. mature (Howells, 1938). Among humans, individual recognition is mediated largely by visual perception of the facial configuration and some believe that we are especially adapted to remember many different faces for long periods of time so that individuals may be recognized as having been seen or not seen before, despite changes in configuration, orientation, and expression (Galper, 1970; Gombrich, 1961; Scapinello & Yarmey, 1970; Shepard, 1967; Warr, 1968; Yin, 1969).

Faces vary in distinctiveness, and Ford (1958) hypothesized that high-distinctive faces are more recognizable than low-distinctive faces, in the sense that the former produce both fewer errors of omission (i.e., failing to recognize faces previously seen) and fewer errors of commission (i.e., falsely recognizing faces not previously seen). Recently, Going and Read (1974) reported findings supporting Ford's hypothesis. Subjects estimated the distinctiveness of each face in a set of faces, the reference point being the many faces with which the subjects were familiar from everyday experience. Later, in a test of facial recognition administered to other subjects, high-distinctive faces produced significantly fewer errors of omission and commission than did low-distinctive faces. The present experiment provides an additional test of Ford's hypothesis, using a different reference point from which to estimate facial distinctiveness, namely, the distinctiveness of each face relative to the other faces in the set.

METHOD

The experiment was conducted in two phases. In Phase I, subjects rank-ordered 12 faces from most to least distinctive. On the basis of the ranks, three categories were established, consisting

M. E. Cohen is now at the Department of General Educational Psychology, Temple University, Philadelphia, Pennsylvania. Requests for reprints should be addressed to W. J. Carr, Department of Psychology, Beaver College, Glenside, Pennsylvania 19038.

of 3 high-, 6 medium-, and 3 low-distinctive faces. Phase II consisted of recognition tests, administered 1 week after inspection, to determine the accuracy of recognition for high- vs. low-distinctive faces.

Subjects

Eighteen female undergraduates served as subjects in Phase I of the experiment and 94 female undergraduates served as subjects in Phase II. None had previously seen any of the faces that were presented.

Apparatus

Twelve photographs of adult males (ages 22-40) were selected randomly from a large file of faces, the faces selected having no unusual features (scars, beards, eyeglasses, etc.). These photographs were converted to standardized line drawings via the Idnti-kit (Smith-Wesson Law Enforcement Group) procedures widely used by law enforcement agencies. The kit consists of transparent overlays, with a major facial feature on each.

Procedure

In Phase I, the relative distinctiveness for each of the 12 faces was established. At the time of testing, the faces were placed in random order, and each subject was asked invidiaully to rank-order the faces from most to least distinctive, by selecting first the most and least distinctive and then working toward the middle of the distribution. Distinctiveness was to be judged in terms of the context of the 12 faces presented, not in terms of the general population of faces with which the subjects were familiar. No time limit was imposed on the task. The subjects were allowed to change the ranking of the faces until they were satisfied with the order of distinctiveness.

Phase II consisted of recognition tests for the 12 faces varying in distinctiveness. In accord with a technique similar to that described by Munn (1961), subjects individually examined a single inspection sheet containing six faces for a 1-min period. The subjects were informed that they would later be asked to identify these faces. One week later, the subjects were presented individually with a test sheet containing all 12 faces used in Phase I, 6 of which having been present on the inspection sheet and six not. Within a 5-min period, the subjects indicated (forced choice) whether each face was or was not on the inspection sheet.

For Group A (N = 48), three of the faces on the inspection sheet had been previously judged (in Phase I) to be high-distinctive and three low-distinctive. The additional six faces that were on the test sheet were those previously judged to be medium-distinctive. For Group B (N = 46), the six faces on the inspection sheet were medium-distinctive and the six additional faces on the test sheet were the three high- and three low-distinctive faces. Thus, with respect to the high- and low-distinctive faces, subjects in Group A could make only errors of omission (i.e., failing to recognize faces previously seen), and subjects in Group B could only make errors of commission (ie., falsely recognizing faces not seen previously). The

possible effect of a given face's position on either the inspection or test sheet upon the recognition of that face was eliminated by using five different spatial orders of faces with each of the two groups.

RESULTS AND DISCUSSION

Phase I. The mean ranks assigned by the 18 subjects to the faces in the high-, medium-, and low-distinctive categories were 9.2, 6.1, and 4.6, respectively; and the differences between these means were statistically reliable (Friedman's analysis of variance for ranked data, $\chi^2 = 48.3$, df $= 11$, p $< .01$). Moreover, intersubject agreement proved to be reliable (Kendall's coefficient of concordance, W $= 0.26$, $\chi^2 = 52.6$, df $= 11$, p $< .01$). We conclude that some faces are reliably more distinctive than others and that subjects agree fairly well in their estimates of distinctiveness.

Phase II. Considering the brief inspection period (1 min) and the long delay between inspection and recognition tests (1 week), the subjects proved to be quite capable of recognizing the 12 faces as having been seen or not seen previously. Twenty-one of the 94 subjects made no errors of omission or commission, and the overall error-rate for Groups A and B combined was only 17.4%; i.e., only 196 errors out of 1,128 responses.

But the important issue at hand is the effect of high vs. low distinctiveness of face upon the two kinds of error (omission and commission) inherent in the overall error rate. First, consider the performance of Group A. On inspection, this group saw only high- and low-distinctive faces. Therefore, on the recognition test, errors of omission (i.e., failing to recognize faces seen previously) could be made only on high- or low-distinctive faces. Of necessity, all errors of commission (falsely recognizing a face not seen previously) involved medium-distinctive faces. Of the 53 errors of omission made by Group A, 19 were made on high-distinctive faces and 34 on low-distinctive faces. Errors of omission were significantly less frequent on high- than on low-distinctive faces (single sample test of proportion, $z = 2.12$, df $= 52$, p $< .05$.

Second, consider the performance of Group B. On inspection, this group saw only medium-distinctive faces. Therefore, on the recognition test, errors of omission could be made only on medium-distinctive faces. Of necessity, all errors of commission involved high- or low-distinctive faces. Of the 45 errors of commission made by Group B, 11 were made on high-distinctive faces and 34 on low-distinctive faces. Errors of commission were significantly less frequent on high- than on low-distinctive faces ($z = 3.45$, df $= 44$, p $< .01$).

The results of Phase II support the hypothesis suggested by Ford (1958) that high-distinctive faces are more recognizable than low-distinctive faces, in the sense that the former produce fewer errors of omission and commission. However, it is not entirely

clear how facial distinctiveness is to be measured. The individual faces within a set of faces can be said to vary in distinctiveness about either of two reference points, specified by the way the judges rank-order the faces for distinctiveness.

The first reference point (here called "familiarity") is used when the subjects rank-order faces for distinctiveness, relative to the many faces with which they are familiar from everyday experience. Such a reference point is defined by the subjects' concept of "Mr. Average" (Goldstein, Harmon, & Lesk, 1971), which, of necessity, is relatively inaccessible and labile across susjbects. The second reference point (here called "stimulus context") is used when the subjects rank-order faces for distinctiveness, relative to other faces in the set. Such a reference point is a property of the set itself and is, therefore, more accessible and less labile across subjects. Unfortunately, this reference point is also highly dependent upon the particular set of faces selected, especially if the number of faces in the set is small.

The hypothesis suggested by Ford (1958) that high-distinctive faces are more recognizable than low-distinctive faces has been tested using both ways of measuring distinctiveness described above. Going and Read (1974) found support for the hypothesis using "familiarity" and the present experimenters found support for it using "stimulus context." Our findings suggest a phenomenon akin to the von Restorff effect (Green, 1958) for facial recognition. But further research is necessary to determine the relative contributions of the two indices of distinctiveness on the recognizability of human faces and other stimulus materials.

REFERENCES

Ford, C. *Has anybody seen me lately?* Garden City, N.Y: Doubleday, 1958.

Galper, R. E. Recognition of faces in photographic negative. *Psychonomic Science*, 1970, **19**, 207-208.

Going, M., & Read, J. D. Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills*, 1974, **39**, 109-110.

Goldstein, A. J., Harmon, L. D., & Lesk, A. B. Identification of human faces. *Proceedings of the Institute of Electrical and Electronic Engineers*, 1971, **59**, 748-760.

Gombrich, E. H. *Art and illusion.* New York: Pantheon, 1961.

Green, R. T. The attention-getting value of structural change. *British Journal of Psychology*, 1958, **49**, 311-314.

Howells, T. H. A study of the ability to recognize faces. *Journal of Abnormal and Social Psychology*, 1938, **33**, 124-127.

Munn, N. L. *Psychology* (4th ed.). Boston: Houghton Mifflin, 1961.

Scapinello, K. H., & Yarmey, D. The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, 1970, **21**, 329-330.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156-163.

Warr, R. B., & Knapper, C. *The perception of people and events.* New York: Wiley, 1968.

Yin, R. K. Looking at upside-down faces. *Journal of Experimental Psychology*, 1969, **81**, 141-145.