# FACIAL REPRESENTATION OF MULTIVARIATE DATA[1]

Robert J. K. Jacob

Naval Research Laboratory
Washington, D. C.

Computer-generated cartoon faces, first proposed by Herman Chernoff in 1971, appear to combine a number of desirable properties for representing multivariate data graphically, including the integrality of the display dimensions and the general familiarity of faces. A series of experiments revealed that, for some useful tasks, the faces do indeed constitute a superior representation for multidimensional Euclidean data. A further series examined how the face displays could be applied to a particular multivariate application. There, it was found that the stereotype meaning already present in faces could be measured and exploited to construct an inherently meaningful display.

## I. INTRODUCTION

People are well-known to be proficient at processing visual information (Entwisle and Huggins, 1973). They can do sophisticated processing tasks, almost below the level of consciousness, when the data are presented graphically (Arnheim, 1969). Until the advent of computer graphics, however, people were not nearly as good at generating graphical--or iconic--information as they were at assimilating it. Hence most data

[1]This research was supported by a contract between the Johns Hopkins University and the Engineering Psychology Programs, Office of Naval Research; and by the U. S. Public Health Service Hospital in Baltimore, Maryland.




ISBN 0-12-734750-X

From: P. Wang (ed), Graphical Representation of Multivariate Data, New York: Academic Press, 1978, pp. 143-168.

were actually communicated using symbols--in the symbolic mode, rather than the iconic mode. Now, the problem has become how best to use the iconic mode for communicating information. While there are some traditional iconic techniques, such as maps and Cartesian graphs, given the new capabilities, it becomes worthwhile to look for new, better, and richer ways to use the iconic mode for communicating information.

A particularly clever iconic device for communicating multidimensional numerical information was proposed by Herman Chernoff (1971; 1973). This was the cartoon human face. Humans look at and process faces constantly. They have become well adapted to this task and are extremely good at performing it. Hence humans would be expected to perform visual processing on faces better than on otherwise comparable visual stimuli. In fact, some evidence suggests that the perception of faces is a special visual process (Yin, 1970).

In order to represent multidimensional numerical data facially, variation in each of the coordinates of the data is represented by variation in one characteristic of some feature of the cartoon face. For example, the first component of the data might be represented by the length of the nose. Other components would be represented by others of the eighteen possible parameters, such as the curvature of the mouth, separation of the eyes, width of the nose, and so on. Then, the overall value of one multidimensional datum would be represented by a single face. Its overall expression--the observer's own synthesis of the various individual features--would constitute a single image depicting the overall position of the point in its multidimensional space. The variety of possible facial expressions would represent the variation possible in a set of numerical data. By looking at the faces and applying one's innate visual processing abilities to them, an observer could perform the visual equivalents of such tasks as multivariate clustering or identifying outliers as easily as he notices family resemblances between people, and by precisely the same, almost unconscious mental mechanism.

Figure 1 shows how the faces are used to represent data. Here, each face represents the value of an eight-dimensional datum chosen from an uncorrelated multivariate normal distribution. One datum differs significantly from the remaining nineteen on several dimensions. It is rather clearly and rapidly identifiable (by a facial expression which differs from the remaining nineteen), despite the presence of considerable noise from the normal distribution. (It is Face 4.)

Several changes were made to Chernoff's original faces for these studies. Most obvious is that the nose was changed from a line to a triangle, and its width is now an additional variable. Chernoff's face height and width parameters were replaced by size and aspect ratio, which better match perceived dimensions. There were some discontinuities and anomalies in the way small changes in the face outline parameters affected the appearance of the face. These were remedied by devising a system of ratio parameters for the face outline. Finally, it was found that reducing the range of variation on most parameters gave a more realistic set of faces; these were preferred because people are especially attuned to very small variations in realistic faces. (The computer program used to generate the faces can be found in Jacob, 1976a or 1976b.)

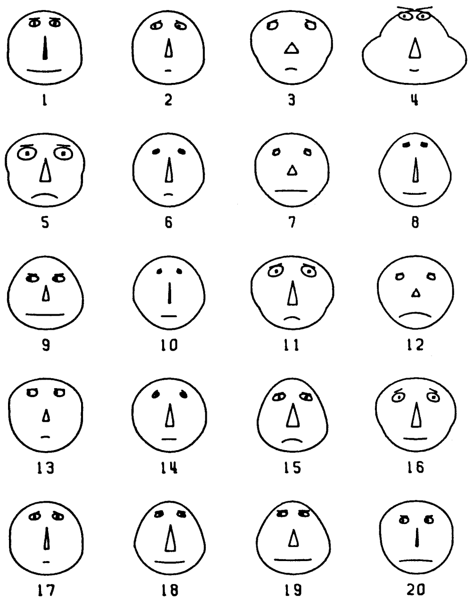


FIGURE 1. Facial display of multivariate data

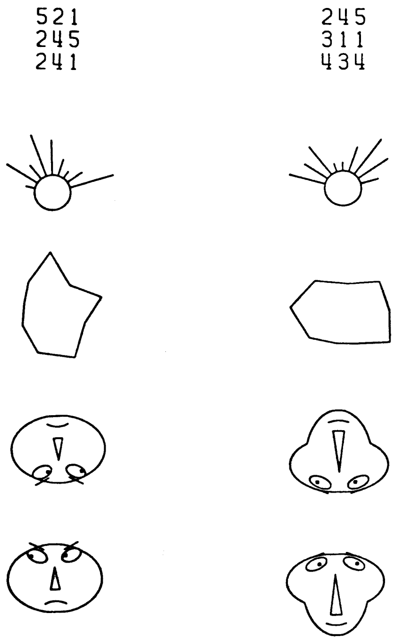


FIGURE 2. Examples from experiment 1

## II. PSYCHOLOGICAL EXPERIMENTS

The first set of experiments was intended to ascertain whether subjects could perform common or useful tasks better when the data were displayed as faces or as traditional iconic or symbolic displays. (Jacob, Egeth, and Bevan, 1976 provides more details on these experiments.) In each of the two experiments, subjects were given a simple task to perform involving a set of random data. Performance was compared between subjects who were given the facial representation for the random data and those who were given other representations.

### A. Experiment 1

The task in the first experiment was paired-associate learning, a simple, standard task in psychological studies. It consists of asking subjects to learn to associate a name with each data point. Twelve such points were represented by digits, "glyphs" (see Anderson, 1960), polygons (see Siegel, Goldwyn, and Friedman, 1971), upside-down faces, and faces. Each of these displays is illustrated in Figure 2. The entire procedure was repeated for three different dimensionalities. A total of 120 subjects were used.

Results revealed a variety of effects, some mutually confounding. There was a clear dimensionality effect as expected; subjects performed better on points in higher-dimensional spaces, since they contained a greater amount of memorizable information (Egeth, 1966). Because the digit displays lent themselves to rote rehearsal, they induced rather good performance. The overall result, however, was that faces were at least as good as any of the other displays, and often better.

The most interesting observation is that the conventional faces were substantially better than the upside-down faces. Upside-down faces have all the geometrical characteristics of conventional faces, but they lack the familiarity of faces. They were included to determine whether the face is just a geometrically well-designed display (in which case the upside-down face should be just as good) or whether the face is a unique display; results indicated the latter.

## B. Experiment 2

While the paired-associate learning task was a standard research task, it was not the sort of task to which the faces were intended to be applied in practice. The second experiment investigated a realistic and practically useful task. This was clustering, or sorting into categories, or pattern recognition.

The task consisted of a set of 50 points in a nine-dimensional space, which were to be organized into 5 groups. They were generated in 5 clusters, each normally distributed around a center point, named the prototype. The subject's task was to look at the 5 prototypes and then assign each of the 50 deviants to a cluster surrounding one of the 5 prototypes. The correct answers were those which put deviants with the prototypes from which they were generated, and to which they were closest in Euclidean distance. While this was a contrived task in that the questions were derived from the answers, it was outwardly similar to many realistic tasks. In a real task, the subject would have the 5 prototypes in his mind, abstracted from his experience or training. He would look at a new data point and assign it to one of the groups he knew. For example, a doctor would examine the data on a patient and then assign him to a cluster which represents a particular disease.

As in the first experiment, the 55 data points were represented in several different ways, and subjects performed the same task with the different displays: faces, a second set of faces with the range of possible variation reduced to three-fourths that of the first set, polygons as in the first experiment, and digits. Figures 3 through 6 present the prototypes (top row of each figure) and examples of their deviants (succeeding rows) for the four different display types respectively. Polygons were used here because they had been found to be the better of the two alternate graphic displays used in the first experiment, probably because their elements are better integrated (Garner, 1974).

Results consisted of the number of errors subjects made in classifying the 50 points. Table I shows the mean number or errors (chance performance would give 40 errors) they made and the mean time (in minutes) they took in sorting the 50 cards. The two types of faces were found clearly to be superior to both the polygons and the digits at $\underline{p} < 0.001$.

TABLE I. Results of Experiment 2 -- 24 Subjects

| | Faces | Faces (3/4 range) | Polygons | Digits |
|---|---|---|---|---|
| Mean no. wrong | 15.33 | 17.08 | 27.96 | 31.88 |
| Standard dev. | 5.16 | 5.76 | 4.98 | 7.30 |
| Mean time (mins.) | 4.14 | 4.07 | 3.69 | 8.24 |
| Standard dev. | 1.63 | 1.46 | 1.43 | 3.22 |

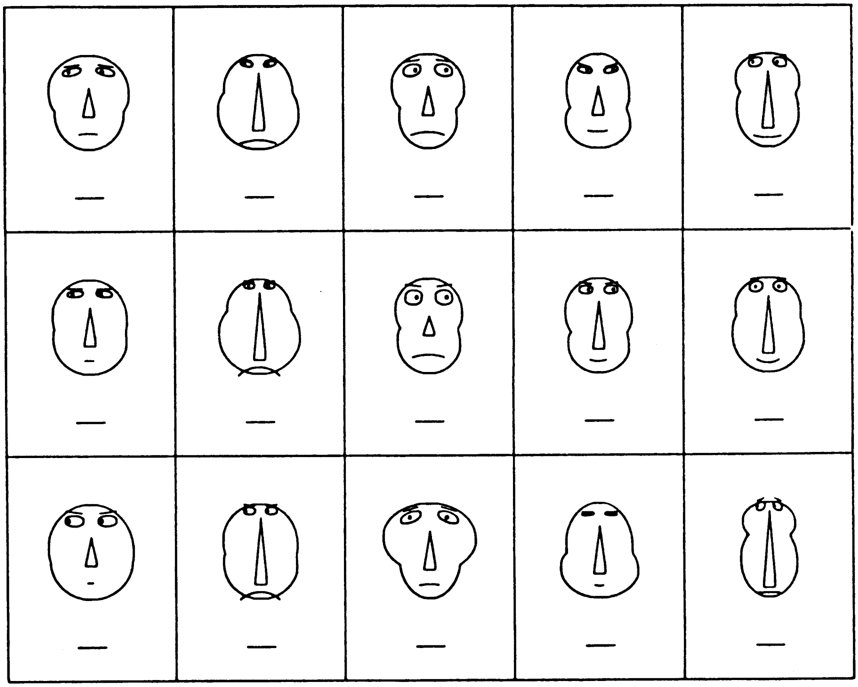

FIGURE 3. Examples from experiment 2: faces

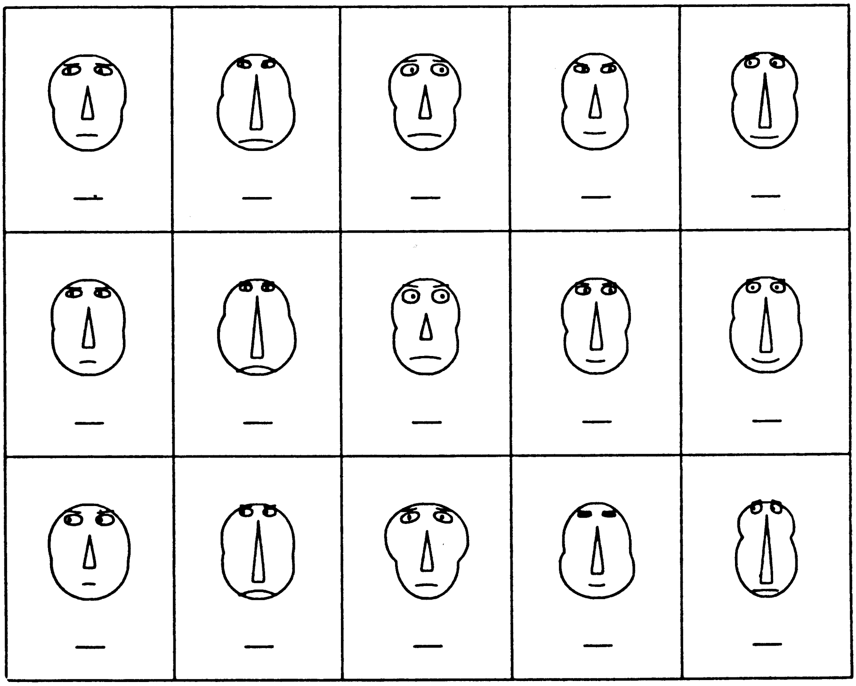

FIGURE 4. Examples from experiment 2: 3/4 range faces

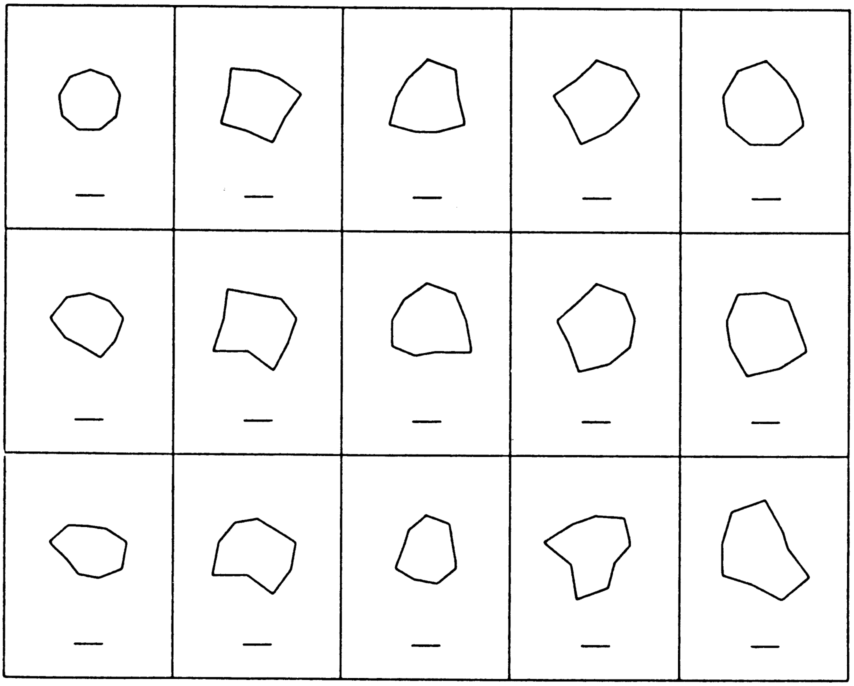

FIGURE 5. Examples from experiment 2: polygons

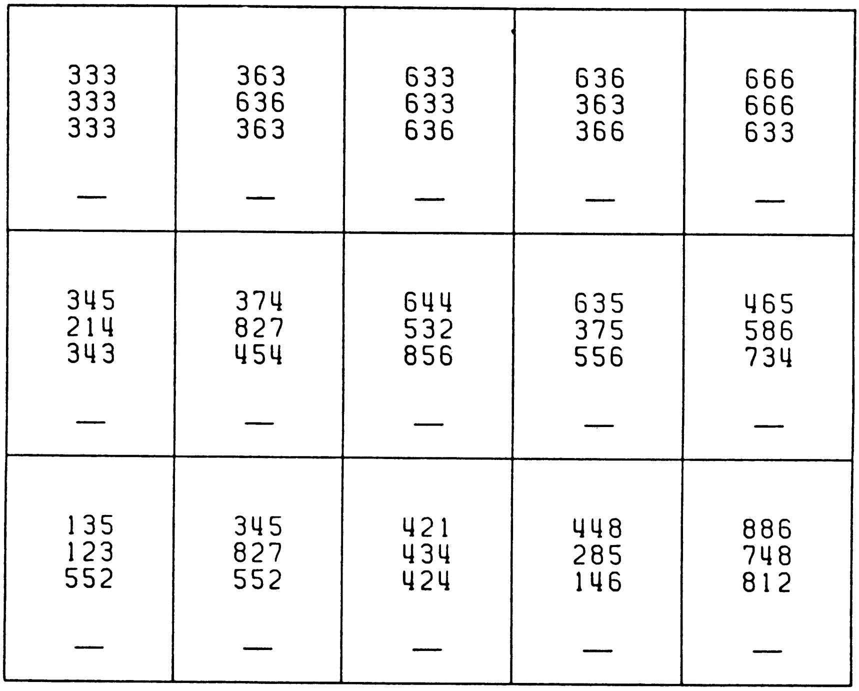


FIGURE 6. Examples from experiment 2: digits

Significant differences were not found between the two face types or between the polygons and digits. While the polygons could be sorted as quickly as the faces, they were not sorted correctly.

The conclusion drawn was that subjects performed a realistic and useful task significantly better when the data were represented by faces than when they were represented by a conventional display (digits) or by a well-integrated graphic display (polygons). As the experimental task is a fairly general one, one underlying many specific data analysis tasks such as diagnosis, pattern recognition, and cluster analysis, it is claimed that faces provide a superior display for many multivariate applications. Subjects' comments on the experiment help explain this result. They reported that they put all the "happy" faces in one pile, "angry" ones in another, and so on; they found this easy to do. In fact, because of the representation, they were performing a fairly sophisticated multivariate clustering task accurately using only their visual processing abilities. For the other displays, they reported inventing more complicated strategies, which turned out to be self-defeating.

This synthesis by the observer himself of the various graphical elements of the facial display into a single gestalt is one of the principal advantages of this type of iconic display. Many other common types of displays contain several variable elements and could thus be used for graphing multivariate data; but often such displays predispose toward a piecemeal, sequential mode of processing, which obscures the recognition of relationships among elements. By contrast, faces induce their observer to integrate the display elements into a meaningful whole. Previous research with simple cartoon faces and with photographs of real faces has indicated that observers do indeed process these stimuli in such a wholistic fashion (Yin, 1969; Smith and Nielsen, 1970; Reed, 1972).

## III. APPLICATION TO MEANINGFUL DATA

Having provided support for the initial supposition that the faces provide a demonstrably good display for Euclidean data of several dimensions, the problem of displaying a specific type of actual (rather than synthetic) data using faces was addressed. The data selected for this purpose were the results of a psychological test intended to determine a patient's psychological personality profile. It was thought that such a profile might possess a more natural facial representation than most other sorts of data. (More details on these experiments, as well as discussions of some related issues, are contained in Jacob, 1976a.)

The form the data took was the results of five particular scales of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway and McKinley, 1942). The U. S. Public Health Service Hospital in Baltimore administers this test to patients as part of a comprehensive health testing and evaluation and was interested in alternate ways to display the test results. The hospital uses five of the clinical scales of the MMPI: Hypochondriasis, Depression, Paranoia, Schizophrenia, and Hypomania.

Following the approach both of Chernoff and of the previous two experiments, the five components of an MMPI data point could simply have been assigned arbitrarily to five of the facial features (while the unused features were kept at constant values). The resulting facial expressions and the personality traits or disorders which each represents would then be learned by doctors, just as they have learned the meanings of the numerical data and the graphs presently in the patient reports. However, it has been widely observed (e.g., Secord, Dukes, and Bevan, 1954; Harrison, 1964) that particular facial expressions tend to signify particular personality traits to observers with great consistency. Therefore, if the face displays could be devised in such a way that the expression on the cartoon face suggested the same personality traits as those in a particular MMPI report, the resulting face displays would tend to communicate the meaning of the data they represent intuitively. To this extent, a self-explanatory display would have been constructed, somewhat like an hypothetical graph in which it is not necessary to label the axes, because the meaning of the curve is inherently obvious.

Consider, for example, a particularly unfortunate arbitrary assignment of MMPI scales to facial features, in which a smile on the face signified a patient suffering from severe depression. While this could certainly be learned, just as the letters depression or the shape of the personality profile graph are learned, such training would clearly be a poor utilization of the observer's skills.

Therefore experiments were undertaken to attempt to obtain a positive relationship between the 5 components of the MMPI score vector and the 18 variable parameters of the face construction. It was hoped that the resulting face displays would be highly intuitive and suggestive; unlike most computer output formats which require the human observer to learn to understand the computer's language, the power of the computer would here be used to tailor the display format to suit the person's intuition and preconceptions.

## A. Experiment 3

The first experiment in this study attempted to measure a relationship between MMPI scores and face parameters based on one observer's preconceptions or stereotypes. This corresponded to a transformation between the 5-space of MMPI scores and the 18-space of face parameters. Because of the imprecision in the process of perception of personality from faces, it was hoped that a linear model would be sufficiently accurate for useful results. Subsequent analysis of the experimental results for higher-order interactions showed this to be a reasonable choice. Moreover, the dimensionality of the problem made any other model very much more difficult to study. Thus a matrix ( T ) was proposed to define a linear transformation from the space of MMPI score vectors ( d for diagnosis) into that of face parameters ( p ).

A set of 200 faces was generated using parameter ( p ) vectors chosen from an 18-variate uniform random distribution. Figure 7 shows a sample of these faces. Dr. Faith Gilroy, a research psychologist at the Public Health Service Hospital, then rated each of the faces on the five scales. She was, in effect, indicating what MMPI results each of the faces signified to her, or, more specifically, what MMPI score she thought a person who looked like each of the 200 faces would receive.

A multiple linear regression of the p vectors on the d's was computed from 200 pairs of such vectors, producing a T matrix of regression coefficients (Jacob, 1976a). That matrix could then be used to estimate a p vector (or face drawing) for any given d vector (or personality score). Such estimated p vectors were computed and compared to the original (stimulus) p vectors; the mean squared error over all components of all the vectors was 0.07497 (components of the p vectors ranged between 0 and 1).

The T matrix was displayed graphically by computing the p vectors which correspond to equally-spaced points along the axes of the d space (that is, points which represent patients who have only one psychological disorder). Figure 8 shows the resulting display. In it, each row depicts a series of patients with increasing amounts of a single disorder. Because of the rating scale used, 0 (the first column) represents an inverse amount of the disorder, 1 represents no disorder (the origin of the d space), and 4 represents a large, extrapolated amount of the disorder. It was thought that these faces (particularly those in the column labelled 3) actually corresponded to common stereotypes of the personality traits they were claimed to represent. The subject had never seen these faces nor any resembling them; rather, they had been deduced from the linear regression using faces reported to have more than one disorder.

Some comparisons were made between this T matrix and results obtained by previous investigators. While no studies had used stimuli of this complexity or the same rating scales, some of the observed relations between basic facial feature variations and basic emotions were confirmed. Comparison to the work of McKelvie (1973) and of Harrison (1964) corroborated both the major axes of facial variation found in the T

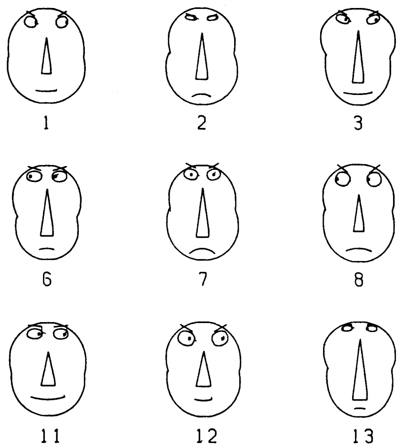


FIGURE 7. Examples from experiment 3

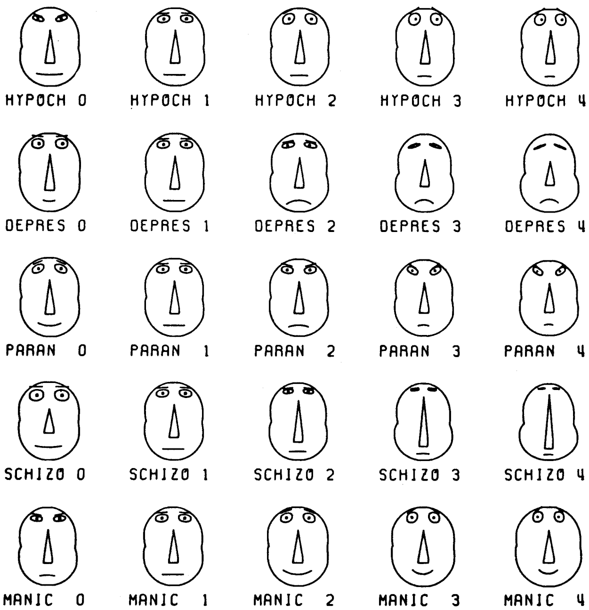


FIGURE 8. Facial representation of the T matrix

matrix and their relationships to variation in emotional states. (As one might expect, these all suggest that the joint variation in the mouth and eyebrows are the major determinants of emotional content of the facial expression; that variation induces variations along axes comparable to Paranoia, Depression, and Hypochondriasis.)

An additional computation showed that the angles in the p space between the facial representations of the orthogonal axes of the d space ranged from 70 to 112 degrees, suggesting that the orthogonal d axes were indeed perceived as being related to orthogonal variations in their facial representations.

Thus, Experiment 3 provided a linear transformation from MMPI scores to faces which was both intuitively appealing and internally consistent. Further study was undertaken in order to validate and then apply this relationship.

## B. Experiment 4

An attempt was made to replicate the previous experiment with the same and with another subject. A new set of random faces, generated similarly to the first set, was presented to two subjects who rated them as in the previous experiment. Actual responses were compared to those predicted using the T matrix of Experiment 3.

The comparison was confounded by the appearance of significant response bias. That is, subjects gave consistently higher or consistently lower ratings to the faces on certain scales. It could be determined that, in those cases where the response <u>magnitudes</u> matched the predictions (approximately half of the data), the present results supported the previous ones in <u>direction</u> as well. In the remaining cases, neither support nor contradiction could be asserted. This experiment could have been improved by embedding the stimulus faces in a larger group which would have induced subjects to attain the same mental set (and thus the same response bias) as that of the subject during Experiment 3. Instead, the insights gained from this experiment were used to devise a new experiment which would provide a more powerful test of the transferability of the T matrix relationship.

## C. Experiment 5

For the relationship T to be valid and transferable to other observers, it must appeal to intuitive stereotypes which are already present in the minds of most observers. Such stereotypes need not possess any absolute validity; they need only be widely and uniformly held in order to be exploitable in devising a facial display for MMPI scores. Thus, this experiment was designed to test the applicability of the stereotypes already discovered. Untrained subjects in the experiment were asked to match facial representations of random hypothetical MMPI scores to alternate representations of the same data. Since the numerical MMPI scores were not meaningful to the subjects (or to the intended final users of the display), an independently-developed textual representation for MMPI scores (Rome *et al.*, 1962) was used in this study.

The 30 subjects were each given 50 stimuli, an example of which appears in Figure 9. In each, the subjects were asked to indicate which of the five faces given best corresponded to the given text description. In fact, that description was the textual representation of a particular point in the d space.

NO. 22

-RESENTFUL AND SUSPICIOUS OF OTHERS

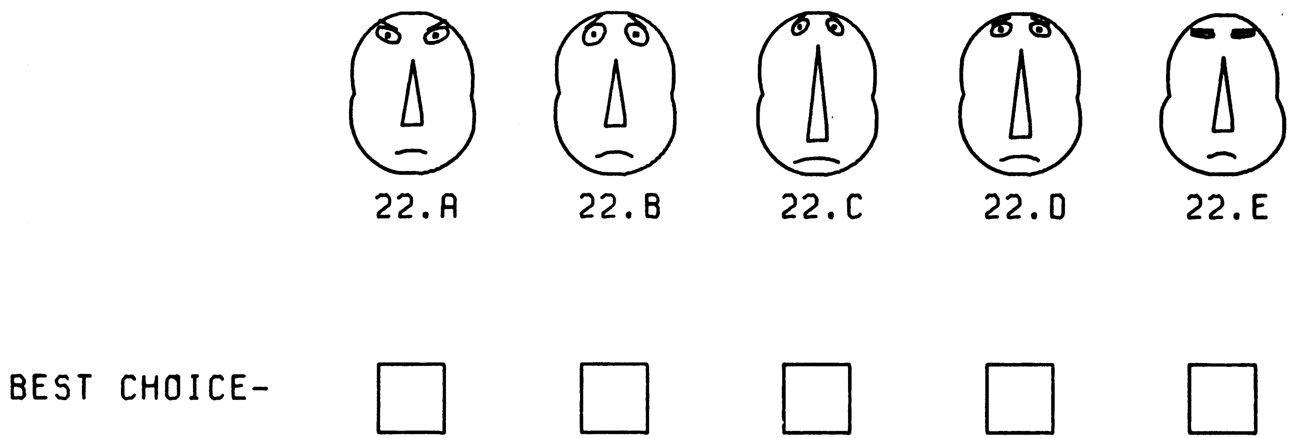


FIGURE 9. Example from experiment 5

One of the five faces was the facial (using the T matrix) representation of the same point, and the remaining faces were representations of other, randomly-selected points.

The principal result of interest was whether entirely naive subjects could select the face which was claimed (by the results of Experiment 3) to represent the same MMPI data as the text at better than chance performance. If the T matrix had no wider validity than for one subject at one point in time, the subjects would not perform the present task; if, however, the matrix relationship corresponded to widely-held stereotypes, the subjects would use such to perform this task better than a random guessing hypothesis would predict. Results were obtained by measuring the Euclidean distance in the 5-dimensional d space between the expected answer and the answer a subject chose. Such a distance could range from 0 (correct choice) to 4.5 (the maximum diagonal dimension of the hypercube). Table II presents these data. A matched $\underline{t}$ test on the data revealed that subjects were able, with highly significant ($\underline{p} < 0.0005$) accuracy, to choose those faces which were designed to communicate the same information as the text items.

Such a result suggests that the faces plus the T transformation obtained provide a data display which requires no

TABLE II. Results of Experiment 5 -- 30 Subjects

| | |
|---|---|
| Chance performance | 1.571 |
| Mean observation | 1.226 |
| Standard deviation | 0.146 |
| $t_{29}$ | 12.975 |

training of the observer. Without any prior information other than their innate facial stereotypes, subjects were able correctly to perceive the data being displayed.

## D. Experiment 6

Experiment 5, then, demonstrated that the faces could be used to communicate psychological data to naive subjects. Experiment 2 showed that a particular useful task could be performed better and more quickly with facially-represented data than with several other representations. Together, the experiments suggest that the face might be a superior mode of displaying the MMPI data under consideration. The present experiment was intended to test this composite hypothesis by having subjects perform a meaningful and realistic task which requires apprehension of MMPI data. Various subjects would perform the same task using the facial and the textual representations of the same MMPI data, and their performance would be compared.

A truly realistic task would be the diagnosis and treatment of a real patient; the results would be measured by evaluating the patient's well-being at the conclusion of the treatment. Unfortunately, there would be far too many confounding variables in such an experiment (as well as ethical considerations). Instead, a crude task, analogous to psychological triage, was devised. Subjects were asked to rate the overall emotional well-being of an hypothetical patient, given his MMPI test scores presented in one of two ways. Their success would be measured by comparing their responses to the responses of a clinical psychologist who studied the unprocessed numerical MMPI scores. Thus, to the extent that a naive subject's responses, using the facial or textual representation, corresponded to this baseline, it could be claimed that, through the use of that representation for the data, he was able to perform the same task as the trained psychologist.

Thirty-two subjects were each given fifty stimuli, each of which resembled either Figure 10 or Figure 11. In each case, the subject was being asked to rate a random point in the d space (represented facially or textually) for emotional well-being.

Results were obtained by measuring the correlation coefficient between a subject's ratings and those of the psychologist. A chance hypothesis would have predicted zero correlation. The mean correlation scores over subjects are presented in Table III. First, one can observe that subjects' performance exceeded chance expectation significantly ($p < 0.005$) for both faces and text. Next, conventional and also paired-observation $t$ tests were made to find the difference

TABLE III. Results of Experiment 6 -- 32 Subjects

| | Text | Faces |
|---|---|---|
| Mean correlation score | 0.644 | 0.399 |
| Standard deviation | 0.095 | 0.111 |
| Difference from chance--$t_{31}$ | 38.485 | 20.416 |
| Difference between means--$t_{62}$ | 9.533 | |
| Paired observations difference--$t_{31}$ | 8.823 | |
| Score using text algorithm | 0.667 | |

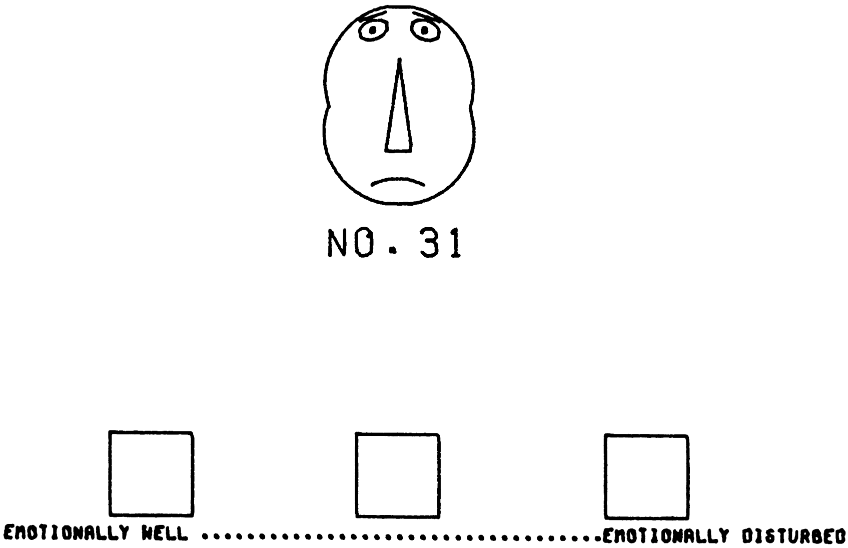


FIGURE 10. Example from experiment 6

```
-ABOVE AVERAGE NUMBER OF PHYSICAL COMPLAINTS, UNDUE CONCERN
 WITH BODILY HEALTH
-TOUCHY, UNDULY SENSITIVE, SUSPICIOUS, INCLINED TO BLAME OTHERS
 FOR OWN DIFFICULTIES

                         NO. 31




              I-----I        I-----I        I-----I
              I     I        I     I        I     I
              I     I        I     I        I     I
              I-----I        I-----I        I-----I

   EMOTIONALLY WELL......................EMOTIONALLY DISTURBED
```

FIGURE 11. Example from experiment 6

between the two display types. Both tests showed that subjects performed the task significantly ($p < 0.005$) better when given the text than when given the faces.

Some insight into this unexpected situation may be gained by studying the text displays in more detail. It appeared that the more disturbed a patient was, the longer his text description was. Hence subjects' responses to the text could have been based on this unexpected iconic content of the text display; they could have been responding to the quantity of text rather than to its meaning. To test this, an algorithm which rated the emotional well-being of a patient based only on the quantity of text in the textual representation of his MMPI score was applied to the experimental stimuli. As shown in the table, the algorithm achieved slightly better performance than the subjects who used the text display. Thus the superior performance of the text displays could be explained by their unintentional iconic content; or, illiterate subjects could have produced the same responses from the text displays as did college students.

The conclusions of this experiment are, then, unclear. While the text displays induced better performance, this turned out to be explainable by an irrelevant property they were found to possess. Nevertheless, the usefulness of faces for inducing good performance in processing Euclidean data was established by Experiment 2; and the ability of the transformation discovered in Experiment 3 to transmit data facially without training was established by Experiment 5. These continue to suggest that an improved version of Experiment 6 would indicate superiority for the facial representation.

Experience in constructing Experiment 6, however, suggests that it would not be a trivial task to devise an unassailable experimental test of this hypothesis.

## IV. CONCLUSIONS

Two principal conclusions are drawn from this study. First, computer-produced faces are a particularly good representation for inducing superior performance of useful tasks on multivariate metrical data. Experiments with other iconic and symbolic displays indicate that it is the face display itself, not merely the iconic mode, that accounts for this superiority. Second, the stereotype meaning already present in faces can be utilized in constructing a display. It was possible to measure and then exploit such meaning in order to create a demonstrably self-explanatory display for a particular set of data.

## ACKNOWLEDGMENTS

Prof. William Huggins was the author's advisor while the author was a graduate student in Electrical Engineering at the Johns Hopkins University; he provided guidance and insight throughout this work. Profs. Howard Egeth and William Bevan at Johns Hopkins guided the work on the first set of experiments. Drs. Richard Hsieh and Faith Gilroy of the U. S. Public Health Service Hospital in Baltimore provided important assistance for the second set of experiments.

## REFERENCES

Anderson, E. (1960). Technometrics 2:387.

Arnheim, R. (1969). "Visual Thinking." Univ. of California Press, Berkeley.

Chernoff, H. (1971). Stanford University Department of Statistics Technical Report No. 71.

Chernoff, H. (1973). American Statistical Assoc. J. 68:361.

Egeth, H. (1966). Perception and Psychophysics. 1:245.

Entwisle, D. and Huggins, W. (1973). Child Development. 44:392.

Garner, W. (1974). "The Processing of Information and Structure." Lawrence Erlbaum, Potomac, Md.

Harrison, R. (1964). "Pictic Analysis." Doctoral dissertation, Michigan State University.

Hathaway, S. and McKinley, H. (1942). "The Minnesota Multiphasic Personality Schedule." Univ. of Minnesota Press, Minneapolis.

Jacob, R. (1976a). "Computer-produced Faces as an Iconic Display for Complex Data." Doctoral dissertation, Johns Hopkins University.

Jacob, R. (1976b). "PLFACE Program." Available from the author upon request.

Jacob, R., Egeth, H., and Bevan, W. (1976). Human Factors. 18:189.

McKelvie, S. (1973). Perception and Psychophysics. 14:343.

Reed, S. (1972). Cognitive Psychol. 3:382.

Rome, H. _et al_. (1962). Proc. Mayo Clinic. 37:61.

Secord, P., Dukes, W., and Bevan, W. (1954). Genetic Psychol. Monographs. 49:231.

Siegel, J., Goldwyn, R., and Friedman, H. (1971). Surgery. 70:232.

Smith, E. and Nielsen, G. (1970). J. Experimental Psychol. 85:397.

Yin, R. (1969). J. Experimental Psychol. 81:141.

Yin, R. (1970). "Face Recognition: A Special Process?" Doctoral dissertation, M.I.T.