

# Transformation and Relational-Structure Schemes for Visual Pattern Recognition

## Two Models Tested Experimentally with Rotated Random-Dot Patterns

David H. Foster and Robert J. Mason

Department of Communication and Neuroscience, University of Keele, Keele, Staffordshire, U.K.

**Abstract.** Two models for visual pattern recognition are described; the one based on application of internal compensatory transformations to pattern representations, the other based on encoding of patterns in terms of local features and spatial relations between these local features. These *transformation* and *relational-structure* models are each endowed with the same experimentally observed invariance properties, which include independence to pattern translation and pattern jitter, and, depending on the particular versions of the models, independence to pattern reflection and inversion ( $180^\circ$  rotation). Each model is tested by comparing the predicted recognition performance with experimentally determined recognition performance using as stimuli random-dot patterns that were variously rotated in the plane. The level of visual recognition of such patterns is known to depend strongly on rotation angle. It is shown that the relational-structure model equipped with an invariance to pattern inversion gives responses which are in close agreement with the experimental data over *all* pattern rotation angles. In contrast, the transformation model equipped with the same invariances gives poor agreement to the experimental data. Some implications of these results are considered.

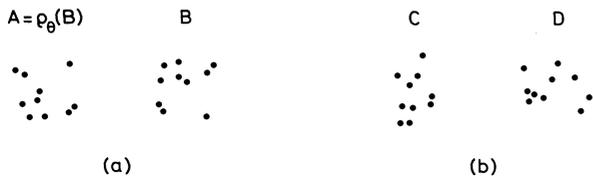
### 1. Introduction

Two basic mechanisms which have been proposed in the functional modelling of human visual pattern recognition are *compensatory pattern transformation* and *relational-structure encoding*. In schemes based on compensatory pattern transformation (see, for example, Pitts and McCulloch, 1947; Hoffman, 1970; Marko, 1973), it is supposed that visual recognition is achieved by the application in some appropriate internal perceptual space of certain restoring transfor-

mations, for example, translations and dilatations. Thus, two patterns under visual inspection are judged to be the same, independent in this case of pattern position and size, if some combination of these transformations can be used to bring the internal representation of one pattern into coincidence with that of the other, the coincidence usually being evaluated by some form of correlation operator. When one of the patterns is a standard, the transformation process is sometimes referred to as *normalization*. The transformations may be applied before the correlation operation according to a prescribed procedure, or contemporaneously with the correlation operation to maximize the coincidence measure (Marko, 1973; Ullmann, 1974).

In schemes based on relational-structure encoding (see, for example, Sutherland, 1968; Barlow et al., 1972) it is supposed that stimulus patterns are internally encoded in terms of local features and spatial relations between these local features. Suggested local features include spots, edges and bars, and suggested relations include "left of", "above", and "joined to". The sameness of two patterns is then determined by some operation which specifies the extent of the concurrence of the structural descriptions. For the local features and relations just given, recognition independent of pattern position and size is automatically obtained. Some discussion of the merits of transformation and relational-structure schemes is given by Marko (1973), Sutherland (1973), Reed (1973), Leeuwenberg and Buffart (1978) and Foster (1977, 1978b). For general reviews of machine-oriented pattern-recognition techniques, see Fu and Rosenfeld (1976) and Ullmann and Rosenfeld (1977).

In general, it is straightforward to endow either type of model with an invariance to a given objective transformation of a pattern. For transformation models, the procedure is in fact trivial: for each objective pattern transformation  $\psi$ , the model is equipped with



**Fig. 1a and b.** Illustrations of stimulus pattern pairs. In **a** the patterns  $A, B$  have the same shape and differ only in orientation,  $\theta = 270^\circ$ ; in **b** the patterns  $C, D$  are paired at random and have different shape

• the capacity to effect the inverse transformation  $\psi^{-1}$ . For relational-structure models, the procedure is not trivial, but provided the selected objective pattern transformations form a group, appropriate invariant quantities can usually be computed. In the following, the objective transformations  $\psi$  for which a pattern  $A$  and its transform  $\psi(A)$  have a high probability of being recognized as “the same” are called *invariance transformations*, and the models are referred to as having the corresponding *invariance properties*.

Given a transformation model and a relational-structure model, each of which is endowed with precisely the same invariance properties, a fundamental problem becomes evident when a comparison of the two models is attempted. Suppose that  $\psi$  is any invariance transformation and that  $A$  is any pattern; if  $A$  and  $\psi(A)$  are presented to each of the models, then the response of one model will be indistinguishable from that of the other, the outputs in both cases necessarily signalling (with high probability) that the two patterns are “the same”. One situation in which a functional difference between the models should, however, be observed is when the two patterns for comparison appear neither completely different, nor “the same”, so that there exists no invariance transformation relating the two. For such pairs of patterns, the measure of their residual similarity determined by the transformation model might be expected to be different from that determined by the relational-structure model. The former involves the correlation of pattern representations; the latter, the comparison of encodings in terms of structural descriptions.

In this paper we describe a transformation model and a relational-structure model each of which has certain invariance properties, based on observed visual recognition performance. These invariances include independence to pattern translation and local pattern distortion (“jitter”). We test the two models by the method described above. The pairs of patterns we consider are such that the one is obtainable from the other by a rotation in the plane. As is already known (Dearborn, 1899; Aulhorn, 1948; Rock, 1973) visual recognition varies strongly with pattern rotation: performance falls off steadily with rotation angle and then increases for angles near  $180^\circ$ . We determine whether

each model equipped with the same selected invariances can correctly predict the reduced levels of recognition performance found for rotation angles between  $0^\circ$  and  $180^\circ$ . As will be seen, the two models differ significantly in their pattern responses in this region.

## 2. Visual Recognition Experiments

### 2.1. Main Experiment

The experimental visual recognition data against which the responses of each model are compared are taken from a previous investigation (Foster, 1978b) into the effect of pattern rotation on the recognition of random-dot patterns. Experimental details relevant to the present study are given below. For a more complete account, see Foster (1978b). The experimental data are presented later.

Subjects (24 in all) were briefly presented with pairs of side-by-side random-dot patterns  $A, B$  for which  $A = \rho_\theta(B)$ , where  $\rho_\theta$  is a clockwise rotation in the plane through angle  $\theta$  and  $\theta$  has values  $0^\circ, 15^\circ, \dots, 345^\circ$ . Each pattern contained ten dots constrained to lie within a fixed limiting circle which subtended  $0.75^\circ$  at the eye. Each dot subtended  $0.05^\circ$  and the centre-to-centre separation of the patterns was  $1.25^\circ$ . Figure 1a shows a typical stimulus pair, except that the dots appeared bright against a uniform background field. (Strictly, the relationship between  $A$  and  $B$  should be written  $A = \tau_0 \rho_\theta(B)$  where  $\tau_0$  denotes the translation through  $1.25^\circ$ , but for clarity  $\tau_0$  is omitted in the following.) The mean rotation autocorrelogram for these patterns is given in the Appendix. After each presentation, subjects indicated (forced choice) whether the two patterns had the same shape, in that one pattern could be obtained from the other by some combination of translation and rotation in the plane and reflection about a vertical axis. Note that the inclusion of reflections in this definition of “same shape” renders it equivalent to a metric definition, where distances between pairs of points in one pattern are required to be the same as those between corresponding pairs of points in the other pattern. For a relational-structure scheme, the metric definition (and therefore the above transformational equivalent) is appropriate since there is then no requirement that relations define the “parity” of a pattern.

As controls, random-dot patterns  $C, D$  paired at random were also presented (see, for example, Fig. 1b). “Same” responses were thus obtained for patterns that had the same shape and for patterns that had different shape. Recognition performance at a given rotation angle  $\theta$  is specified in terms of Tanner and Swets’ (1954) discrimination index  $d'$ , which derives from the

simple equal-variance normal-distributions model of signal detection theory (Green and Swets, 1966). This index provides a measure of the visual distinguishability of same-shape and different-shape patterns that is relatively insensitive to changes in level of subject response criterion. When  $d' > 0$ , the two types of pattern pair are inferred to be visually distinguishable; when  $d' = 0$ , they are inferred to be visually indistinguishable.

## • 2.2. Auxiliary Experiment

It was mentioned in the Introduction that there is an elevation in recognition performance at rotation angles close to  $180^\circ$ . We wished to consider the possibility that this elevation at  $180^\circ$  is due in part or in all to the *independent* operation of processes which include invariance to pattern reflection  $\mu_y$  about a vertical axis, invariance to pattern reflection  $\mu_x$  about a horizontal axis, and invariance to pattern inversion  $\iota$ , that is, rotation in the plane through  $180^\circ$ . Obviously, if the first two invariances can be effected jointly, then the last invariance may be omitted. Accordingly, we performed a short additional experiment with conditions similar to those of the main experiment, in which the side-by-side patterns  $A, B$  for visual comparison were such that (i)  $A$  was identical with  $B$ , i.e.  $A = \text{Id}(B)$ , (ii)  $A = \mu_y(B)$ , (iii)  $A = \mu_x(B)$ , (iv)  $A = \iota(B)$ , and (v)  $A$  and  $B$  were paired at random. The use of these data in the construction of the models is described in the next section.

## 3. Models

The models to be constructed here have certain fixed invariance properties, namely insensitivity to pattern position and pattern jitter (see, for example, Sutherland, 1968). In the transformation model, these invariances are achieved by operating with the corresponding inverse transformations; in the relational-structure model, these invariances arise naturally by the construction of the internal encoding. In addition to these fixed invariances, we also consider the effect of the introduction of certain discrete invariances to account for the observed elevation in recognition performance at  $180^\circ$  rotation. From the results of the auxiliary recognition experiment (Sect. 2.2), it was concluded that there is (i) invariance to reflection  $\mu_y$  about a vertical axis (visual recognition performance equal to that for identical patterns), (ii) invariance to pattern inversion  $\iota$ , i.e. rotation  $\varrho_{180^\circ}$  through  $180^\circ$  (recognition performance approaches that for identical patterns), and (iii) *non*-invariance to pattern reflection  $\mu_x$  about a horizontal axis (recognition performance greatly reduced with respect to that for identical patterns).

These results are consistent with the findings of Sekuler and Rosenblith (1964) and others using a similar stimulus arrangement, but different stimulus patterns. It follows that the elevation in recognition data at  $180^\circ$  cannot be attributed to the joint operation of two processes, the one invariant to  $\mu_x$  and the other to  $\mu_y$ . The additional discrete invariances to be tested are therefore restricted to  $\mu_y$  and the inversion  $\iota$ .

For a pair of patterns  $A, B$ , where  $A$  is a rotated version of  $B$ ,  $A = \varrho_\theta(B)$ , each model initially provides a measure of the similarity of  $A$  and  $B$ . This measure is a deterministic quantity and may be normalized to range between zero and unity. In order that a comparison may be made with the experimental data expressed in terms of discrimination index  $d'$ , this deterministic output is also converted into a discrimination index  $d'$ . Recall that  $d' > 0$  implies that the pair  $(\varrho_\theta(B), B)$  is in general distinguishable from a random pair  $(C, D)$ , and that  $d' = 0$  implies that it is not. This transformation of the initial output merely constitutes a linear scaling. Thus suppose that  $d'_{\text{id}}^E$  denotes the experimentally determined value of the discrimination index for visual responses to identical ( $\theta = 0^\circ$ ) patterns and randomly paired patterns. If  $c(\varrho_\theta(B), B)$  denotes the (normalized) measure of similarity of the same-shape pair  $(\varrho_\theta(B), B)$  computed by the model, then the predicted discrimination performance  $d'^T(\varrho_\theta(B), B)$  is given by:

$$d'^T(\varrho_\theta(B), B) = d'_{\text{id}}^E \cdot c(\varrho_\theta(B), B). \quad (1)$$

Clearly, if  $\theta = 0^\circ$ , then  $c(\varrho_\theta(B), B) = 1$  and  $d'^T(B, B) = d'_{\text{id}}^E$ , as required; if  $c(\varrho_\theta(B), B)$  is zero, then  $d'^T(\varrho_\theta(B), B)$  is zero, which implies that  $\varrho_\theta(B)$  and  $B$  are no more recognizable as each other than are, in general, an arbitrary pair  $(C, D)$ . For the additional invariance operations relating to the reflection  $\mu_y$  and the inversion  $\iota$ , scaling equations similar to (1) are used. From the main experiment, the measured values of the discrimination indices  $d'_{\text{id}}^E$  and  $d'_i{}^E$ , corresponding to identical pattern pairs  $(B, B)$  and inverted pattern pairs  $(\varrho_{180^\circ}(B), B)$  respectively, are given in Table 1 below, along with the value of  $d'_{\mu_y}^E$  which from the results of the auxiliary experiment is set equal to  $d'_{\text{id}}^E$ .

Table 1

Identity	Inversion	Reflection
$d'_{\text{id}}^E = 1.513$	$d'_i{}^E = 1.396$	$d'_{\mu_y}^E = 1.513$

### 3.1. Transformation Model

Suppose that  $A$  and  $B$  are any two patterns in the plane, not necessarily composed of randomly positioned dots. Assume that each pattern is assigned an internal representation which, to within the limits implied by visual acuity, is in one-to-one correspon-

dence with the original, that is, the representation does not correspond to two or more visually distinguishable patterns. Although the compensatory transformations and correlation operation are applied to these representations internally, we shall, for simplicity, consider these operations as being applied to the patterns  $A$  and  $B$  directly. Provided both patterns and operations are defined only to within visual indistinguishability, the two processes are equivalent. The compensatory transformations are the planar translations  $\tau_{(x,y)}$  through  $(x, y)$ , where the plane is equipped with the usual coordinate system, non-linear transformations  $\sigma_{\{(x_i, y_i)\}}$  which are defined below, reflection  $\mu_y$  about the  $y$ -axis, and inversion  $\iota$ . When some combination of these transformations is applied to the pattern  $A$  we obtain the transformed pattern  $A'$ . The correlation between  $A'$  and the other pattern  $B$  of the pair is measured by a modification of the usual overlap integral:

$$\iint A'(\xi, \eta)B(\xi, \eta)d\xi d\eta,$$

where  $A'(\xi, \eta)$  and  $B(\xi, \eta)$  are the planar luminance distributions of  $A$  and  $B$  respectively. The modification to the overlap integral is specific to the particular patterns used here and consists of the replacement of the rectangular functions describing the luminance distributions of the individual dots in each pattern by delta functions at the centres of the dots. A dot in one pattern is thus effectively either overlapped completely with a dot in the other pattern, or not overlapped at all. This corresponds to the subjective correlation between the two patterns being established on a “dot-to-dot” basis. The non-linear transformations  $\sigma_{\{(x_i, y_i)\}}$  act upon the dot pattern by shifting the centre of each dot  $p_i$  through  $(x_i, y_i)$ , where  $x_i^2 + y_i^2 < r^2$ ,  $i = 1, 2, \dots, n$  ( $n$  dots in each pattern). The parameter  $r$  is referred to as the *jitter parameter*.

Suppose then that  $A$  and  $B$  are random-dot patterns and that  $A^*(x, y)$  and  $B^*(x, y)$  are their luminance distributions after the delta-function substitutions described above. The compensatory transformations are adjusted to maximize the correlation coefficient under three conditions, corresponding to the different discrete invariance operations considered, namely, identity alone, reflection  $\mu_y$ , and inversion  $\iota$ . Thus:

$$c_{\text{id}} = \max_{(x,y)} \max_{\{(x_i, y_i)\}} \iint \tau_{(x,y)} \circ \sigma_{\{(x_i, y_i)\}}^r (A^*)(\xi, \eta) \cdot B^*(\xi, \eta) d\xi d\eta, \quad (2)$$

$$c_{\mu_y} = \max_{(x,y)} \max_{\{(x_i, y_i)\}} \iint \tau_{(x,y)} \circ \sigma_{\{(x_i, y_i)\}}^r \circ \mu_y (A^*)(\xi, \eta) \cdot B^*(\xi, \eta) d\xi d\eta, \quad (3)$$

$$c_{\iota} = \max_{(x,y)} \max_{\{(x_i, y_i)\}} \iint \tau_{(x,y)} \circ \sigma_{\{(x_i, y_i)\}}^r \circ \iota (A^*)(\xi, \eta) \cdot B^*(\xi, \eta) d\xi d\eta, \quad (4)$$

where the integrations are over the plane. Each of these correlation coefficients is normalized to unity ( $\max_{A,B} \{c_{\alpha}\} = 1, \alpha = \text{Id}, \mu_y, \iota$ ). Let  $c_{\alpha}(\theta)$ ,  $\alpha = \text{Id}, \mu_y, \iota$ , denote the value of the correlation coefficient when  $A = \varrho_{\theta}(B)$ , and let  $\tilde{c}_{\alpha}$ ,  $\alpha = \text{Id}, \mu_y, \iota$ , denote the value of the correlation coefficient for randomly paired patterns  $C, D$  averaged over all such pairings. ( $c_{\alpha}(\theta)$  depends on  $B$ , but we shall eventually average over all  $B$  and it is convenient to shorten the notation here.) To convert the same-shape coefficient  $c_{\alpha}(\theta)$  to a form suitable for linear scaling by the corresponding experimental discrimination index  $d_{\alpha}^E$  (Table 1), we subtract the corresponding averaged random-pair coefficient  $\tilde{c}_{\alpha}$  from  $c_{\alpha}(\theta)$ , and then normalize. Thus, for each discrete invariance property, the predicted discrimination index  $d_{\alpha}^T(\theta)$  for the same-shape pair  $(\varrho_{\theta}(B), B)$  is given by

$$d_{\alpha}^T(\theta) = d_{\alpha}^E \frac{c_{\alpha}(\theta) - \tilde{c}_{\alpha}}{1 - \tilde{c}_{\alpha}}, \quad \alpha = \text{Id}, \mu_y, \iota. \quad (5)$$

[An alternative to converting  $c_{\alpha}(\theta)$  in the above way is to compute a discrimination index directly from  $c_{\alpha}(\theta)$  and  $\tilde{c}_{\alpha}$ , the coefficients considered as theoretical “hit” and “false-alarm” rates respectively, and then to scale the result if necessary. This approach does not materially affect the fit of the model to the experimental data, and we use (5) since it is then of the same form as the scaling used in the relational-structure model.]

We evaluate the importance of the discrete invariances to  $\mu_y$  and  $\iota$  by testing four versions of the transformation model with discrimination indices (5) combined as follows:

- I.  $d_1^T(\theta) = d_{\text{id}}^T(\theta)$ ,
- II.  $d_2^T(\theta) = \max \{d_{\text{id}}^T(\theta), d_{\mu_y}^T(\theta)\}$ ,
- III.  $d_3^T(\theta) = \max \{d_{\text{id}}^T(\theta), d_{\iota}^T(\theta)\}$ ,
- IV.  $d_4^T(\theta) = \max \{d_{\text{id}}^T(\theta), d_{\mu_y}^T(\theta), d_{\iota}^T(\theta)\}$ .

Version I relies solely on the fixed invariances to translation and jitter; version II includes the contribution for reflection about a vertical axis; version III includes the contribution for inversion; and version IV includes the contribution for reflection and inversion. In II, III, and IV, it is assumed that the discrimination performance effected is the best of those possible.

The expressions I, II, III, and IV were each evaluated by computer for the 21 same-shape pattern pairs and 24 different-shape pattern pairs made use of in the main experiment (Sect. 2.1), at angles  $\theta = 0^\circ, 15^\circ, \dots, 345^\circ$ . There is one adjustable parameter in the model, namely the jitter parameter  $r$ , but as will be seen the predicted discrimination indices are relatively insensitive to changes in the value of  $r$ .

In Fig. 2, the predicted recognition performance in terms of discrimination index is shown as a function of rotation angle  $\theta$  for each version of the transformation model with various values of the jitter parameter  $r$  (expressed in terms of dot diameters). Experimental data, pooled over intervals of  $45^\circ$ , are indicated by the solid points. Both theoretical and experimental data are averaged over all pattern pairs. For all versions of the model the fit of the theoretical data to the experimental data is poor. The best fit is for version IV which has both  $\mu_y$  and  $\iota$  as discrete invariance transformations, but even here the predicted results are too low at intermediate  $\theta$  values. A statistical analysis of the deviations of the experimental data from the  $d_4^T(\theta)$  curve with  $r=2.5$  (based on formulae derived by Gourevitch and Galanter, 1967) shows the lack of fit to be highly significant ( $\chi^2 = 59$ ,  $df = 6$ ,  $P < 0.00001$ ).

### 3.2. Relational-Structure Model

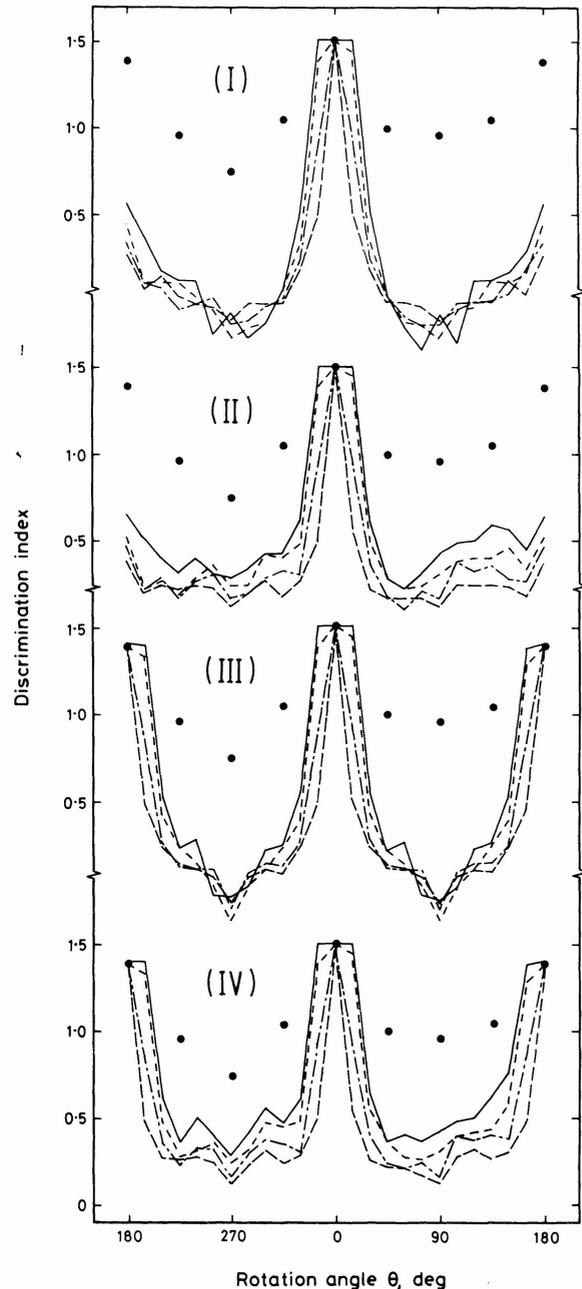
As before, let  $A$  and  $B$  be any two patterns in the plane, not necessarily composed of randomly positioned dots. We suppose that each pattern is visually encoded in terms of local features and spatial relations between local features, and specifically that the relations are of the form "left of" and "above". (Other relations such as "joined to" and "near to" may be included in the pattern representation, but the following analysis does not depend upon their being specified.) The pattern  $B$  is thus assigned a structural description consisting of a set  $F(B)$  of local features

$$\{f_i | 1 \leq i \leq n\}$$

and a set  $R(B)$  of relations

$$\{r_x(f_i, f_j), r_y(f_i, f_j) | 1 \leq i < j \leq n\},$$

where  $r_x(f_i, f_j)$  indicates whether  $f_i$  is to the left or right of  $f_j$ , and  $r_y(f_i, f_j)$  indicates whether  $f_i$  is above or below  $f_j$ . The pattern  $A$  is assigned a similar description. These descriptions are independent of pattern position and small local distortions in the pattern. For the random-dot patterns used in the present study, the local features  $f_i$  formed by the various clusters of dots



**Fig. 2.** Recognition performance of the transformation model. Predicted discrimination performance of same-shape patterns from different-shape patterns is plotted as a function of pattern rotation angle for four versions I, II, III, and IV of the model (Sect. 3.1). The continuous and various interrupted lines are for different values of the jitter parameter  $r$ , which in terms of dot-diameters are thus: --- 1.0, - · - · 1.5, - - - - 2.0, — 2.5. The experimental data, pooled over  $45^\circ$  intervals, are indicated by the solid points. The differences between the experimental and theoretical data are highly significant in all cases (each  $P < 0.00001$ )

are unknown. We consider instead the set  $R'(B)$  of relations between the individual dots  $p_i$  of the pattern  $B$ :

$$\{r_x(p_i, p_j), r_y(p_i, p_j) | 1 \leq i < j \leq n\},$$

and compare this with the set  $R'(A) = R'(\varrho_\theta(B))$  of relations between the individual dots of the rotated pattern  $A = \varrho_\theta(B)$ :

$$\{r_x(q_i, q_j), r_y(q_i, q_j) | 1 \leq i < j \leq n\},$$

where  $q_i = \varrho_\theta(p_i)$ ,  $1 \leq i \leq n$ . Let

$$m_{ij}^x(\theta) = 1 \quad \text{if } r_x(p_i, p_j) = r_x(q_i, q_j) \\ = 0 \quad \text{otherwise;}$$

similarly, let

$$m_{ij}^y(\theta) = 1 \quad \text{if } r_y(p_i, p_j) = r_y(q_i, q_j) \\ = 0 \quad \text{otherwise.}$$

A measure of the concurrence of the relational descriptions  $R'(B)$  and  $R'(\varrho_\theta(B))$  is given by the sums:

$$m^x(\theta) = \sum_{i < j} m_{ij}^x(\theta), \quad m^y(\theta) = \sum_{i < j} m_{ij}^y(\theta). \quad (6)$$

The variables  $m^x(\theta)$  and  $m^y(\theta)$  specify the number of horizontal and vertical relations unchanged by rotating pattern  $B$  through  $\theta$ . Thus, if  $\theta = 0^\circ$ ,  $m^x(\theta) = m^y(\theta) = n(n-1)/2$ , and if  $\theta = 180^\circ$ ,  $m^x(\theta) = m^y(\theta) = 0$ .

As descriptions of the changes in relations between local features, the variables  $m^x(\theta)$  and  $m^y(\theta)$  incorporate unnecessary "microstructure", that is, information concerning changes in relations between dots *within* local features. The effect of these contributions for a given pattern is to smooth the dependence of  $m^x(\theta)$  and  $m^y(\theta)$  on  $\theta$ . This smoothing may be neglected, however, in view of the subsequent general smoothing that occurs when responses are averaged over all patterns. Note that the derivations of  $m^x(\theta)$  and  $m^y(\theta)$  entail the assumption that individual dots in a pattern may be "labelled", that is, one dot distinguished from another by virtue of its position relative to that of other dots in the pattern. If such specification were not possible, then one could find at least two dots  $p_1$  and  $p_2$  such that by some combination of non-zero rotation and translation, the pattern could be brought into coincidence with itself and the dots  $p_1$  and  $p_2$  superimposed. The fact that on average same-shape patterns only coincide at  $\theta = 0^\circ$  (see Appendix) supports the assumption that dots may indeed be "labelled".

To incorporate invariance to reflection  $\mu_y$ , about a vertical axis and inversion  $\iota$ , we suppose that in the first case all the relations "left of" are replaced by "right of" and vice-versa, and in the second case all relations "left of" are replaced by "right of" and vice-

versa, and all relations "above" replaced by "below" and vice-versa. Both  $m^x(\theta)$  and  $m^y(\theta)$  are normalized to 0.5 ( $m^x(0) = m^y(0) = 0.5$ ). Corresponding to each discrete-invariance property, the predicted discrimination index  $d'^T(\theta)$  for the same-shape pair  $(\varrho_\theta(B), B)$  is then given by

$$d'_{\text{id}}{}^T(\theta) = d'_{\text{id}}{}^E \cdot (m^x(\theta) + m^y(\theta)), \\ d'_{\mu_y}{}^T(\theta) = d'_{\mu_y}{}^E \cdot (1 - m^x(\theta) + m^y(\theta)), \quad (7) \\ d'_\iota{}^T(\theta) = d'_\iota{}^E \cdot (1 - m^x(\theta) + 1 - m^y(\theta)),$$

where the  $d'_z{}^E$  are given in Table 1. In contrast to the transformation model, the relational-structure model used here is designed to respond only to patterns that have the same local features and there is no general capacity to respond to different-shape patterns. The above approach assumes that when all the relations of a pattern  $A$  are changed (not necessarily by reversal of all horizontal and all vertical relations) to produce some pattern  $B$ , the two patterns are then no more recognizable as each other than are, in general, a pair chosen at random.

As with the transformation model, we test four versions of the relational-structure model with discrimination indices (7) combined as follows. These versions are exactly equivalent to those used for the transformation model.

- I.  $d'_1{}^T(\theta) = d'_{\text{id}}{}^T(\theta)$ ,
- II.  $d'_2{}^T(\theta) = \max\{d'_{\text{id}}{}^T(\theta), d'_{\mu_y}{}^T(\theta)\}$ ,
- III.  $d'_3{}^T(\theta) = \max\{d'_{\text{id}}{}^T(\theta), d'_\iota{}^T(\theta)\}$ ,
- IV.  $d'_4{}^T(\theta) = \max\{d'_{\text{id}}{}^T(\theta), d'_{\mu_y}{}^T(\theta), d'_\iota{}^T(\theta)\}$ .

Version I considers only the total number of relations unchanged; version II includes the reversal of all horizontal relations; version III includes the joint reversal of all horizontal and vertical relations; and version IV includes the reversal of all horizontal relations and the joint reversal of all horizontal and vertical relations.

These expressions were evaluated by computer for the 21 same-shape pattern pairs, at angles  $\theta = 0^\circ, 15^\circ, \dots, 345^\circ$ . Note that there is no jitter parameter  $r$  to be adjusted here. In Fig. 3, the predicted recognition performance in terms of discrimination index is shown as a function of rotation angle  $\theta$  for each version of the relational-structure model. Experimental data, pooled over intervals of  $45^\circ$ , are indicated by the solid points, and both theoretical and experimental data are averaged over all pattern pairs. Both versions III and IV of the model, involving  $\iota$  alone or  $\iota$  and  $\mu_y$  as invariance transformations, give good fits of the theoretical data to the experimental data. Deviations of the experimental data from the  $d'_3{}^T(\theta)$  and  $d'_4{}^T(\theta)$  curves are not significant ( $\chi^2 = 2.6$  and

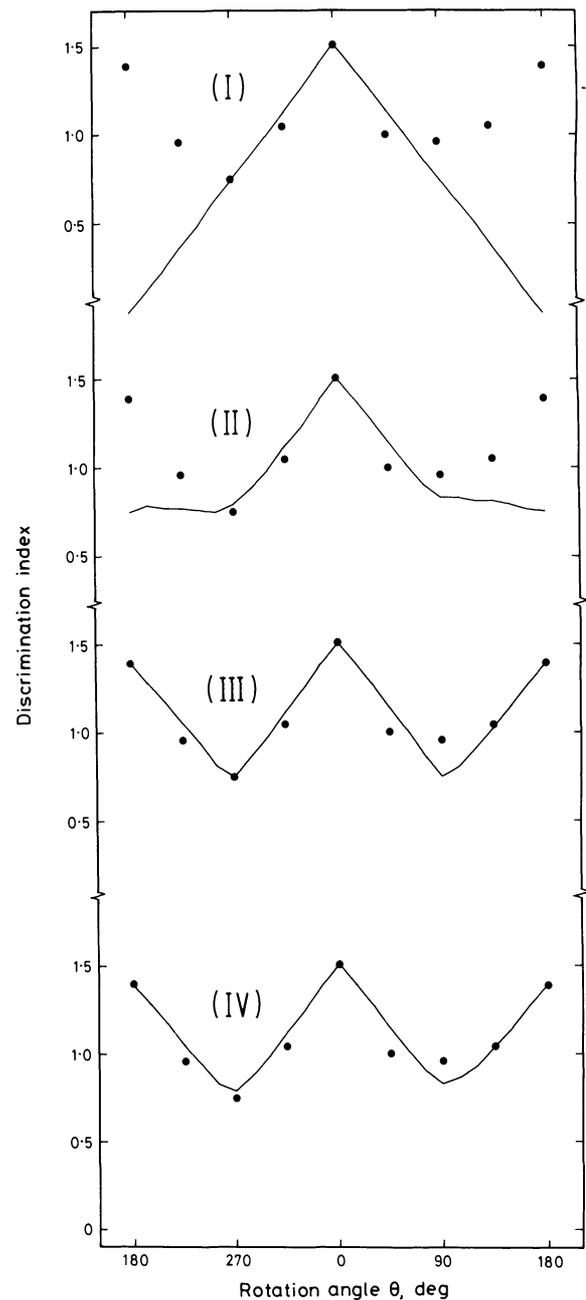
1.7 respectively,  $df=6$ ,  $P>0.5$  for both). Versions I and II are clearly inadequate ( $\chi^2=84.6$  and  $16.2$ ,  $df=7$  and  $7$ ,  $P<0.00001$  and  $P<0.05$  respectively).

#### 4. Discussion

In transformation schemes for visual recognition, one pattern is recognized as being the same as another by the application of certain internal compensatory transformations which transform the internal representation of one pattern into that of the other. In relational-structure schemes, this recognition is achieved by the assignment to each pattern of the same internal encoding based on local features and relations.

The transformation and relational-structure models considered here each display the experimentally observed invariances to pattern translation and to pattern jitter, and, depending upon the particular version of the model, to reflection of the pattern about a vertical axis and to inversion ( $180^\circ$  rotation) of the pattern. Only the transformation model has an adjustable parameter, the jitter parameter  $r$ , but predicted recognition performance, in terms of discrimination of same-shape from different-shape patterns, is shown to be relatively insensitive to changes in the value of  $r$ . To test these models, we compared the predicted recognition performance of each with experimentally determined recognition performance using as stimuli random-dot patterns that had been variously rotated in the plane. The outcome of this comparison may be summarized as follows. First, for either type of model to fit the level of experimental response at  $180^\circ$  pattern rotation, it is necessary for it to be endowed with a specific discrete invariance to pattern inversion, since none of the other invariances produce sufficient elevation at  $180^\circ$ . Second, equipped with this invariance property, the relational-structure model then gives responses which are in good agreement with the experimental data for *all* angles of pattern rotation. Third, independent of whether invariance to inversion or reflection is included in the transformation model, it fails significantly to fit the experimental data over a large range of rotation angles. A corollary of the last result is that the fall-off in visual recognition performance with rotation angle is not attributable to a simple progressive difference in pattern overlaps.

The fundamental nature of the inadequacy of the transformation model (in any of its versions) is evident when we examine the response of the model to rotated linear patterns. Suppose  $A$  is a line of ten dots and  $B$  is  $A$  rotated through  $90^\circ$ . If the dots are sufficiently well spaced, then the value of any of the (normalized) correlation coefficients (2), (3), and (4) for the pair  $(A, B)$  (Sect. 3.1) is then 0.1. Since the mean correlation coefficient for randomly paired patterns is found here



**Fig. 3.** Recognition performance of the relational-structure model. The continuous lines show predicted discrimination performance of same-shape patterns from different-shape patterns as a function of pattern rotation angle for four versions I, II, III, and IV of the model (Sect. 3.2). The experimental data, pooled over  $45^\circ$  intervals, are indicated by the solid points. The differences between the experimental and theoretical data are significant in I and II ( $P<0.00001$  and  $P<0.05$  respectively) and not significant in III and IV (each  $P>0.5$ )

to be not less than 0.3, discrimination scores predicted from versions I, II, III and IV (Sect. 3.1) will be negative. To resolve this difficulty, additional rotations can be introduced into the repertoire of the model's compensatory transformations, providing these rotations are given a suitably diminished weighting to

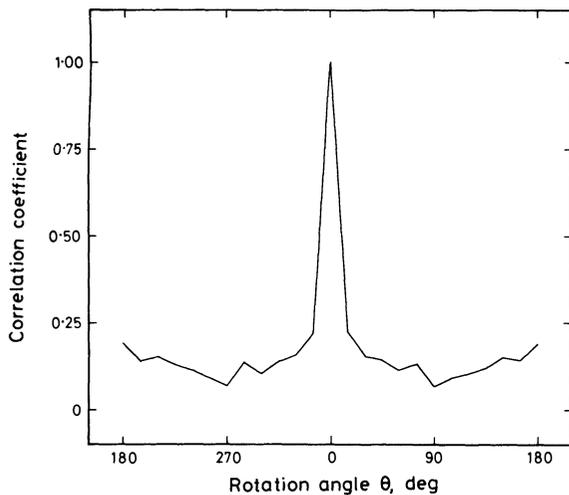


Fig. 4. Mean rotation autocorrelogram of the 21 dot patterns

match the observed recognition response levels between  $0^\circ$  and  $180^\circ$ . These rotations might be effected by some dynamical process so that a particular rotation  $q_{\theta_0}$  would thus be achieved by the execution of the 1-parameter family:

$$q_{t\theta_0}, \quad 0 \leq t \leq 1,$$

which smoothly transforms the pattern  $A$  into the rotated pattern  $q_{\theta_0}(A)$ . [Shepard and Metzler (1971) and Cooper (1975) have suggested that such a process occurs in some reaction-time sense-discrimination tasks.] A dynamical scheme of this kind would be compatible with previous analyses of pattern recognition and visual apparent motion effects (Foster, 1972a, 1973a, b, 1978a; also see Kolers, 1972). It would, however, be necessary to assume that the probability of a rotation being effected by this process

decreases monotonically with angle. For if  $0 < \theta_1 < \theta_2 < 180^\circ$ , and rotation  $q_{\theta_1}$  has probability  $p_1$  and  $q_{\theta_2}$  probability  $p_2$ , then  $p_2$  is the product of  $p_1$  and the probability of the remainder of the process from  $\theta_1$  to  $\theta_2$  being effected, given that it has reached  $\theta_1$ . Hence  $p_1 \geq p_2$ . Experimental support for such a result for visual apparent motion has already been reported (Foster, 1972b; Frisby, 1972). The fact that the probability of visual recognition increases with angle near  $180^\circ$  might then be accounted for by supposing that these 1-parameter families are also executed after application of the inversion operator.

Given the ad hoc inclusion of additional compensatory transformations for a particular class of patterns, it seems likely that the performance of the transformation model may then be made to match that of the relational-structure model as closely as is wished.

If, in addition, the possibility of *pattern-dependent* transformations is introduced into the transformation scheme, the transformation and relational-structure models should then be operationally indistinguishable over all classes of patterns (see comments by Nickerson, 1972, Sect. 3.1). In view of this potential indistinguishability, the most appropriate general recognition scheme might be a hybrid one, where there is both an encoding of a stimulus pattern in terms of local features and relations, and an application at possibly more than one level of compensatory transformations (see Amari, 1968, 1978).

Within the context of the present study, however, the relational-structure model is clearly the more economical of the two types. Given the relations "left of" and "above", and the possibility of the essentially trivial en bloc reversal of the sense of these relations, the model exhibits invariance to translations, jitter, reflection and inversion of a pattern, and correctly predicts the diminished level of recognition performance for patterns rotated through angles between  $0^\circ$  and  $180^\circ$ . The principal disadvantage of the model, not manifested here, is the difficulty in computing its response to patterns which have some, but not all, local features and relations the same. To overcome this problem, it would be necessary to know what constitutes an appropriate assignment of local features to a pattern. For same-shape patterns, this question does not arise.

## Appendix

In terms of the notation of Sect. 3.1, let  $B$  be an arbitrary dot pattern and

$$c_{1d}(\theta) = \max_{(x,y)} \max_{\{(x_i,y_i)\}} \int \int \tau_{(x,y)} \circ \sigma_{\{(x_i,y_i)\}}^r \circ q_\theta(B^*)(\xi, \eta) \cdot B^*(\xi, \eta) d\xi d\eta,$$

with jitter parameter  $r$  set equal to one dot radius. This expression coincides with (2) of Sect. 3.1 after the substitution  $A = q_\theta(B)$ . Figure 4 shows the mean rotation autocorrelogram obtained by averaging the  $c_{1d}(\theta)$  over the 21 dot patterns used to form the same-shape pairs  $(q_\theta(B), B)$ .

*Acknowledgements.* We thank J. I. Kahn and R. J. Watt for stimulating discussions and for critical readings of the manuscript. RJM was supported by an award from the Science Research Council.

## References

- Amari, S.: Invariant structures of signal and feature spaces in pattern recognition problems. *RAAG Memoirs* 4, 553-566 (1968)
- Amari, S.: Feature spaces which admit and detect invariant signal transformations. 4th Int. Joint Conf. on Pattern Recognition, Kyoto, Japan (1978)

- Aulhorn, O.: Die Lesegeschwindigkeit als Funktion von Buchstaben und Zeilenlage. *Pflügers Arch.* **250**, 12–25 (1948)
- Barlow, H. B., Narasimhan, R., Rosenfeld, A.: Visual pattern analysis in machines and animals. *Science* **177**, 567–575 (1972)
- Cooper, L. A.: Mental rotation of random two-dimensional shapes. *Cognit. Psychol.* **7**, 20–43 (1975)
- Dearborn, G. V. N.: Recognition under objective reversal. *Psychol. Rev.* **6**, 395–406 (1899)
- Foster, D. H.: A method for the investigation of those transformations under which the visual recognition of a given object is invariant. I. The theory. *Kybernetik* **11**, 217–222 (1972a)
- Foster, D. H.: A method for the investigation of those transformations under which the visual recognition of a given object is invariant. II. An example experiment: The group of rotations  $SO(2)$  acting on a Landolt ring. *Kybernetik* **11**, 223–229 (1972b)
- Foster, D. H.: A hypothesis connecting visual pattern recognition and apparent motion. *Kybernetik* **13**, 151–154 (1973a)
- Foster, D. H.: An experimental examination of a hypothesis connecting visual pattern recognition and apparent motion. *Kybernetik* **14**, 63–70 (1973b)
- Foster, D. H.: Visual pattern recognition by assignment of invariant features and feature-relations. *Opt. Acta* **24**, 147–157 (1977)
- Foster, D. H.: Visual apparent motion and the calculus of variations. In: *Formal theories of visual perception*, pp. 67–82. Leeuwenberg, E. L. J., Buffart, H. F. J. M., eds. Chichester: Wiley 1978a
- Foster, D. H.: Visual comparison of random-dot patterns: Evidence concerning a fixed visual association between features and feature-relations. *Quart. J. exp. Psychol.* In press (1978b)
- Frisby, J. P.: The effect of stimulus orientation on the phi phenomenon. *Vision Res.* **12**, 1145–1166 (1972)
- Fu, K. S., Rosenfeld, A.: Pattern recognition and image processing. *IEEE Trans. Comput.* **C-25**, 1336–1346 (1976)
- Gourevitch, V., Galanter, E.: A significance test for one parameter isosensitivity functions. *Psychometrika* **32**, 25–33 (1967)
- Green, D. M., Swets, J. A.: *Signal detection theory and psychophysics*. New York: Wiley 1966
- Hoffman, W. C.: Higher visual perception as prolongation of the basic Lie transformation group. *Math. Biosci.* **6**, 437–471 (1970)
- Kolers, P. A.: *Aspects of motion perception*. Oxford: Pergamon Press 1972
- Leeuwenberg, E. L. J., Buffart, H. F. J. M., eds.: *Formal theories of visual perception*. Chichester: Wiley 1978
- Marko, H.: Space distortion and decomposition theory: A new approach to pattern recognition by vision. *Kybernetik* **13**, 132–143 (1973)
- Nickerson, R. S.: Binary-classification reaction time: A review of some studies of human information-processing capabilities. *Psychon. Monogr.* **65**, Suppl. **4**, 275–318 (1972)
- Pitts, W., McCulloch, W. S.: How we know universals. The perception of auditory and visual forms. *Bull. Math. Biophys.* **9**, 127–147 (1947)
- Reed, S. K.: *Psychological processes in pattern recognition*. New York: Academic Press 1973
- Rock, I.: *Orientation and form*. New York: Academic Press 1973
- Sekuler, R. W., Rosenblith, J. F.: Discrimination of direction of line and the effect of stimulus alignment. *Psychon. Sci.* **1**, 143–144 (1964)
- Shepard, R. N., Metzler, J.: Mental rotation of three-dimensional objects. *Science* **171**, 701–703 (1971)
- Sutherland, N. S.: Outlines of a theory of visual pattern recognition in animals and man. *Proc. Roy. Soc.* **B171**, 297–317 (1968)
- Sutherland, N. S.: Object recognition. In: *Handbook of perception*. Vol. III, pp. 157–185. Carterette, E. C., Friedman, M. P. eds. New York: Academic Press 1973
- Tanner, W. P. Jr., Swets, J. A.: A decision-making theory of visual detection. *Psychol. Rev.* **61**, 401–409 (1954)
- Ullmann, J. R.: A review of optical pattern recognition techniques. *Optoelectron.* **6**, 319–332 (1974)
- Ullmann, J. R., Rosenfeld, A.: Picture recognition and analysis. *Radio Electron. Eng.* **47**, 33–48 (1977)

Received: October 20, 1978

Dr. D. H. Foster  
 Department of Communication and Neuroscience  
 University of Keele  
 Keele  
 Staffordshire ST5 5BG  
 England