Aspects of learning in detection and identification processes in visual perception

U. Mortensen, G. Meinhardt

Version: August 2003[1]

FB Psychologie und Sportwissenschaften
Institut III
Westfälische Wilhelms-Universität Münster
Fliednerstr. 21
48149 Münster

[1] artikel/identif/learndetid.tex

Abstract: The sensory representation of a stimulus may change as the result of learning processes during a detection experiment. We present the data (i) of a detection experiment, followed by (ii) an identification experiment, followed again by (iii) a detection experiment identical to the experiment (i). The stimulus in the detection experiment was one of the four patterns used in the detection experiment. There is no difference in the detection thresholds from the two detection experiments, suggesting that the identification experiment has no influence on the detection performance. The detection data further suggest that the pattern is detected by a matched filter for the stimulus pattern, and that the formation of such a filter is fast compared to the learning processes taking place during the identification experiment. The identification data are analysed - apart from testing some standard models of identification processes - employing *Dual Scaling* (Correspondence Analysis (CA)). It is shown that under certain conditions the scale values of the stimuli represent the conditional expected values of the random variable representing the aspect of the sensory representation of the stimulus patterns with respect to which the stimuli are discriminated; CA caters, under certain conditions, for the multivariate case.

Contents

1 Introduction

There is evidence that in detection experiments perceptual learning takes place even for very simple stimuli (e.g. Beard, Levi & Reich (1995), Fahle, Edelman and Poggio (1995)), where perceptual learning could mean the adjustment of some threshold level on the decision axis; or even a change in the sensory representation of the stimulus pattern itself. One question of interest with respect to learning refers to the time scale with respect to which learning takes place.

Consider, on the other hand, identification experiments. Here the subject has to decide which of a set of m stimulus patterns was presented. Suppose that the stimulus pattern employed in the detection experiment is one of the m stimulus patterns of the identification experiment. In the identification experiment, learning will take place as well, and may mean (i) finding the optimal decision bounds allowing to identify a presented pattern with smallest possible probability of error, and possibly (ii) complementary to (i) a fine tuning of the sensory representation of the patterns. If there is such a fine tuning it may have an effect on the detection performance. For instance, the neural processes implying the fine tuning of the sensory representation may imply the formation of matched filters for the patterns, and such filters would also optimize the detection performance; however, as Townsend and Landon (1983) point out, such filters are optimal only if the discriminant functions are linear, which need not be the case in a given experiment.

So our investigation consisted of a series of three experiments: (1) a detection experiment, (2) an identification experiment, and (3) a detection experiment. The stimuli and the experimental conditions where identical in experiments (1) and (3), and the purpose of the identification experiment (2) was to find out whether the learning processes that can be assumed to have taken place during the identification experiment have an effect on the detection performance.

The effect of the identification experiment on the detection performance can easily be discovered by comparing the results of the experiments (1) and (3). In order to find out whether learning has taken place in the identification experiment one may simply test whether the estimates of the probability of correct responses increase with the number of an experimental session. But the structure of confusions of stimulus patterns with each other may change es well with increasing number of experimental session, and in order to investigate these changes the data have to be summarized in a way that allows to see what is going on. We therefore decided to apply Dual Scaling (DS) (Nishisato (1980)), or, equivalently, Correspondence Analysis (CA) (Greenacre (1984)) to the data of each session. DS provides, for a given confusion matrix, scale values (i) for the stimuli and (ii) for the responses. It will be shown that under certain conditions the scale values for the stimuli are proportional to the conditional expected values $E(x|s_i)$ of the stimulus dimension, given that the pattern s_i was presented. There may exist more than one stimulus dimension with respect to which the stimuli are evaluated, and CA as a method equivalent to DS will provide us with the set of latent dimensions necessary to "explain" the χ^2 of a given confusion matrix; whether each of these dimensions indeed represents a stimulus dimension is another question. In any case, stimuli s_i and responses r_j can simultaneously be represented by points in the coordinate system defined by the latent dimensions (this is the biplot), and to the extent that learning implies an increasingly systematic choice of correct responses the positions of points representing s_i and the corresponding response r_i should become more and more similar; the configuration of the points should tell us something about the stimulus dimensions underlying the identification responses.

The DS/CA approach to our data is mainly descriptive, and so it is of interest whether a particular model of the identification process can be fitted to the data. In the light of the results of the detection experiment (the data are compatible with the hypothesis that the pattern is detected by a matched filter for this pattern) the all-or-none model of Townsend (1971a, 1971b) is of particular interest since it it can be taken as a special form of a template matching model: if there exists a matched filter, then identification could be the result of a template matching. On the other hand, subjects may try to identify the patterns by evaluating the representations of stimulus components that vary independently of each other from trial to trial, and one may try to fit the independent decisions model (Ashby and Gott (1988)) to our data. This model is also of particular interest here: if each pattern is represented by the activity of a pattern-specific channel one may say that the representation of the pattern is holistic, and a fit of the independent decisions model would mean that

- that there is the possibility of a decomposition of the sensory representation into independent components, or
- the subjects can adjust the decision bounds such as to allow for an independent evaluation of the components; this point will be taken up again in the Discussion section.

A third alternative is the famous similarity choice model of Luce (1963) where it is assumed that judgments depend on the stimulus similarities. For instance, Nosofsky (1980) and Townsend and Ashby (1982) sucessfully fitted the model to their data. This model may be considered together with Shephards (1957) model, where similiarity is defined as being inversely related to psychological distance between stimuli, and distance may be related to the values stimuli have on certain *stimulus dimensions* (a term introduced by Garner and Morton (1965) for continuously varying features).

There is quite a variety of possible models, see Townsend and Landon (1983) for an overview. A discussion of all these models would be beyond the scope of this paper. We will concentrate on an analysis of the data employing correspondence analysis.

2 Models of detection and identification

2.1 Models of detection

The main purpose of the detection experiments reported in this paper is to test for possible learning effects with respect to the detection task. To this end, the data from two experiments with identical conditions will be compared.

The detection experiments were conceived as superposition experiments in the sense of Kulikowski and King-Smith (1973): the stimulus pattern was presented superimposed on a background pattern. The background pattern had a contrast being a constant fraction q of the contrast m of the stimulus pattern. The stimulus pattern was radially symmetric, and the background pattern was a (radially symmetric) Bessel function $J_0(\omega r$ of order zero, $r = \sqrt{x^2 + y^2}$, $\omega = 2\pi f$, $f = \sqrt{u^2 + v^2}$, and u and v are spatial frequencies; of course, f may also be taken as a spatial frequency. J_0-functions are known to be eigenfunctions of radially symmetric linear systems (Papoulis (1980)), as sinusoidal functions are eigenfunctions of 1-dimensional linear systems. The threshold contrast m of the stimulus pattern is determined for various values of the spatial frequency parameter f of the J_0-pattern; the resulting threshold curves allow to test in particular the hypothesis of detection by a matched filter for the stimulus pattern.

We will not go into the details of testing different models of detection: apart from the matched filter model, models of detection by probability summation (e.g. Graham, Nachmias & Robson (1972), Graham (1977), Wilson & Bergen (1979)), local energy (Morrone and Owens (1987)) could be tested. However, such an undertaking would be beyond the scope of this paper, since our main interest here is whether an identification task changes the detection behaviour. However, the threshold curves (for details see the Methods section) may be compared with regard to their shape; if learning takes place that changes the detection mechanism, these curves should differ.

2.2 Models of identification

Further, suppose there exists some numerical representation $x = x(\nu)$ for ν; usually, x will be a vector $x = (x_1, \ldots, x_n)'$, where x_i represents the value of the i-th dimension, $1 \leq i \leq n$. Let $p_{ij} = p(r_j|s_i)$ denote the probability of confusing stimulus s_j with stimulus s_i (the *confusion probability*). p_{ij} has to be specified in terms of (i) $p(r_j|x)$, i.e. the probability that the response r_j is chosen given the internal representation x, and $p(x|s_i)$, the probability that x was generated by the pattern s_i. Possibly, bias parameters b_j, corresponding to the a priori probability that the pattern s_j was presentd, have to taken into account.

2.2.1 The Similarity Choice Model

This model was introduced by Luce (1963). Let η_{ij} be a measure of similarity between the stimuli s_i and s_j. The the probability of confusing the pattern s_i with s_j is assumed to be given by

$$p_{ij} = \frac{\eta_{ij} b_j}{\sum_k \eta_{ik} b_k}, \quad \eta_{ij} = \eta_{ji} \tag{1}$$

The similarities η_{ij} have to be estimated, alongside with the bias parameters b_j, as free parameters, where it has to be assumed that $\eta_{ij} = \eta_{ji}$, - there are not sufficiently many degrees of freedom to cater for the asymmetrical case $\eta_{ij} \neq \eta_{ji}$.

The similarity η_{ij} may be related to psychological distance d_{ij} between the patterns s_i and s_j via the relation $\eta_{ij} = \exp(-d_{ij})$ (Shepard (1957)). If the model can be fitted to the data it is then of interest to interprete the estimates for η_{ij} with respect to the values of d_{ij} as computed from estimates of the stimulus dimensions; possibly a particular metric for the d_{ij} can be found, which again may tell us something about the way stimulus dimensions are integrated in order to arrive at a decision about the presented stimulus pattern.

2.2.2 The all-or-none model

Townsend (1971a, 1971b) proposed the all-or-none model of identification. According to this model, a stimulus pattern is either identified correctly, and then the response will be correct, or it is not identified at all, and then the subjects simply guesses which stimulus has been presented. The probability of confusing stimulus S_i with S_j, i.e. of giving the response R_j, is then given by

$$p_{ij} = \begin{cases} c_i + (1 - c_i) b_j, & j = i \\ \\ (1 - c_i) b_j, & j \neq i \end{cases} \tag{2}$$

The way the individual components determine the probability c_i is not specified, but the assumption that with probability $1 - c_i$ the subject has to guess suggests that the components interact in some way. According to another interpretation the all-or-none model describes a version of template matching. For instance, there may exist a matched filter for the stimulus pattern, and in an identification experiment, the subject may decide according to the maximally activated matched filter. The distribution of the activity for such filters could be such that only the filter corresponding to the presented pattern has nonzero probability of reaching a threshold value; in that case the all-or-non model may fit the data.

2.2.3 The independent decisions model

Suppose that any stimulus pattern is defined by a combination of values of components, and the set of stimulus patterns is given by the set of all possible combinations of values.

To be specific, in our case there are two components (the "outer" and the "inner" disc), each of which may assume two possible values (radii or diameters). Stimuli are presented in random order so that the value of a component varies independently from one trial to the next. The subject may realize this and try to evaluate each component independent of the other. We will say that the patterns are identified by *independent decisions* (Asby and Gott (1988)).

To avoid unnecessary generality let us assume that there are just two components having two possible values each. Let c_{uv} denote the u-th component with value v; $u, v = 1, 2$. To each c_{uv} there exists a set C_{uv} such that if $x_u \in C_{uv}$, then the subject will decide that the u-th component has the value v. Let $p(x_1 \in C_{11}|c_{11}) = p_1$, $p(x_1 \in C_{21}|c_{21}) = p_2$, $p(x_2 \in C_{12}|c_{12}) = q_1$ and $p(x_2 \in C_{22}|c_{22}) = q_2$ (see the summary in table 1). The sets A_j

Table 1: component values, sets and probabilities

	Values	
Dim.	big	small
D_1	c_{11}, C_{11}, p_1	c_{12}, C_{12}, q_1
D_2	c_{21}, C_{21}, p_2	c_{22}, C_{22}, q_2

are then defined as unions of the sets C_{rs}, e.g. $A_1 = C_{11} \cup C_{21}$, $A_2 = C_{11} \cup C_{22}$, etc. The components x_k of x are stochastically independent random variables, and corresponding to (6) one has

$$
\begin{aligned}
p_{11} &= p(x_1 \in C_{11}|c_{11})p(x_2 \in C_{21}|c_{21}) = p_1 q_1 \\
p_{12} &= p(x_1 \in C_{11}|c_{11})p(x_2 \in C_{12}|c_{21}) = p_1(1 - q_1), \quad \text{etc.}
\end{aligned}
$$

The confusion probabilities are summarized in table 2. Let $f_i(x_1, x_2)$ be the 2-dimensional

Table 2: Identification probabilities according to the independent decisions model

	r-bb	r-bs	r-sb	r-ss
s-bb	$p_1 q_1$	$p_1(1-q_1)$	$(1-p_1)q_1$	$(1-p_1)(1-q_1)$
s-bs	$p_1(1-q_2)$	$p_1 q_2$	$(1-p_1)(1-q_2)$	$(1-p_1)q_2$
s-sb	$(1-p_2)q_1$	$(1-p_2)(1-q_1)$	$p_2 q_1$	$p_2(1-q_1)$
s-ss	$(1-p_2)(1-q_2)$	$(1-p_2)q_2$	$p_2(1-q_2)$	$p_2 q_2$

distribution of $x = (x_1, x_2)'$ when the pattern s_i is presented. The representation of s_i in a given trial is representable as a point in $\mathbb{R}^2$; let X_1 and X_2 be the coordinate axes. The sets C_{11} and C_{12} are then defined as $C_{11} = \{x|x \leq x_{01}\}$, $C_{12} = \{x|x > x_{01}\}$, and analogously $C_{21} = \{x|x \leq x_{02}\}$, $C_{22} = \{x|x > x_{02}\}$. The decision bounds are lines parallel to the coordinate axes X_1 and X_2. Following Ashby and Townsend (1986) we will speak of *decisional independence* if the decision bounds satisfy this condition.

It is intuitively clear that decisional independence does not yet imply *perceptual independence*. The notions of decisional and perceptual independence have been thoroughly discussed by Ashby and Townsend (1986) and related to the notions of perceptual separability and integrality as introduced by Garner and Morton (1969); we return to these concepts in the Discussion.

2.2.4 Dual Scaling (Correspondence Analysis)

As pointed out in the Introduction we prefer to approach our data in a more descriptive manner instead of searching for a model that could explain the data. Dual Scaling (DS) provides scale values u_i for the stimulus patterns s_i and, at the same time, scale values v_j for the responses r_j, $i, j = 1, \ldots, m$, such that

$$u_i = \hat{p}_{i1} v_1 + \cdots + \hat{p}_{im} v_m, \quad \hat{p}_{ij} = n_{ij}/n_{i+}, \tag{3}$$

with $n_{i+} = \sum_j x_{ij}$. The u_i and v_j refer to the same (latent) dimension and are chosen such that the variance of the u_i is maximal relative to that of the v_j (Nishisato (1980), Greenacre (1984)). DS is equivalent to Correspondence Analysis (CA) (Greenacre (1984)), catering, however, very elegantly with the multidimensional case, i.e. with the case that the relation between stimuli and responses cannot be described with respect to a single latent dimension. Quite generally - i.e. without assuming that the $\hat{u}_i$ are estimates of the expected values $E(x|s_i)$ - one should expect similar values for $\hat{u}_i$ and $\hat{v}_j$ if the subject has learned to assign the correct response to any of the presented stimuli. To show that this is the case (3) is written in matrix form: let $\hat{u} = (\hat{u}_1, \ldots, \hat{u}_m)'$, $\hat{v} = (\hat{v}_1, \ldots, \hat{v}_m)'$ and $\hat{P} = (\hat{p}_{ij})$ the matrix of $\hat{p}_{ij}$-values. Then (3) is equivalent to

$$\hat{u} = \hat{P}\hat{v}.$$

If the subject does not make any errors at all, $\hat{P}$ will be a diagonal matrix and $\hat{u} = \hat{v}$ follows. This is only a sufficient condition; in an identification experiment the stimuli are usually shown such that confusions will occur, meaning that $\hat{P}$ is not diagonal. Under conditions specified below the case $\hat{u} \approx \hat{v}$ or even $\hat{u} = \hat{v}$ is still possible.

DS or, equivalently, CA are descriptive methods in so far as no assumptions concerning the the processes constituting the dependencies among stimuli and responses have to be made in order to find the scale values. On the other hand, we may ask what the scale values tell us about the processes we are interested in. To see this let us first recall the Bayes model of pattern recognition; this model is rather general and may serve as a general meta-model for our purposes.

The general Bayes model: According to Bayes' Theorem

$$p(r_j|x) = \frac{p(x|s_j)b_j}{p(x)}, \quad p(x) = \sum_k p(x|s_k)b_j, \tag{4}$$

with b_j the a priori probability that pattern s_j was shown. Suppose the subject sets $p(r_j|x) = 1$ if $p(s_j|x) = \max_k p(s_k|x)$. Let A_j be the set of vectors x such that

$$p(s_j|x) = \max_k p(s_k|x) = \max_k p(x|s_k)b_k. \tag{5}$$

The subject is then behaving optimally, conditional upon the choice of the sets A_j (equivalently: the decision bounds) and the probabilities b_j. In any case, if follows that

$$p_{ij} = P(x \in A_j|s_i)b_j = b_j \int_{A_j} p(x|s_i)dx. \tag{6}$$

Discriminant functions: The subject may also decide with respect to the value of discriminant functions $y_j(x)$, $j = 1, \ldots, m$; the subject decides for r_j (indicating the belief that s_j was shown), if $y_j(x) > y_k(x)$ for all $k \neq j$. A particular choice for the y_j is $p(s_j|x)$. Suppose x is some vector $x = (x_1, \ldots, x_d)'$. Another choice for the $y_j(x)$ is $y_j = a_{j1}x_1 + \cdots + a_{jd}x_d$. i.e. y_j is some weighted sum of the components of x, and the $a_{j1}, \ldots, a_{jd}$ represent the weights the subject assigns to each component of x. However, y_j may also be defined with respect to (6), e.g. $y_j(x) = \log P(x \in A_j|x) + \log b_j$ (the probability $p(x)$ can be neglected here since it is the same for all j); formulating a model according to which the subject decides with respect to the values of discriminant functions may thus be equivalent to adopting the general Bayes model.

For reasons already pointed out in the Introduction we will not test any specific model. On the other hand, the way the scale values are chosen in DS is equivalent to adopting the criteria of discriminant analysis (Nishisato (1980)). In so far one may say that applying DS or CA may well be related to the Bayes or discriminant functions model. We will not pursue this line of argument any further here and point out instead how the u_i in particular may be related to the sensory representation x. We will do this with respect to DS, assuming a single latent dimension represented by x; CA is then the generalisation to the multivariate case.

Let $E(x|s_i)$ be the conditional expectation of x, give that the pattern s_i was presented, and let $A = \cup_j A_j$. Then

$$E(x|s_i) = \int_A xp(x|s_i)dx = \sum_j \int_{Aj} xp(x|s_i)dx \tag{7}$$

If we introduce the factor

$$w_{ij} = \frac{\int_{Aj} xp(x|s_i)dx}{\int_{Aj} p(x|s_i)dx},$$

we can write

$$\int_{Aj} xp(x|s_i)dx = w_{ij} \int_{Aj} p(x|s_i)dx = w_{ij}\, p(x \in A_j|s_i), \tag{8}$$

so that

$$E(x|s_i) = w_{i1}\, p(x \in A_1|s_i) + \cdots + w_{im}\, p(x \in A_m|s_i). \tag{9}$$

We make now the
Assumption:

$$w_{ij} \approx v_j \text{ a constant for all } i \tag{10}$$

Then

$$E(x|s_i) \approx u_i = v_1 p(x \in A_1|s_i) + \cdots + v_m P(x \in A_m|s_m), \tag{11}$$

i.e. the u_i are approximations for the $E(x|s_i)$. A particular choice for the v_j is

$$v_j = \frac{1}{m} \sum_i w_{ij} \tag{12}$$

and the v_j approximates the w_{ij} in the least-squares-sense.

> **Conditions:** For (11) to hold, the condition (10) is sufficient, but - as simulations surprisingly show - not necessary. Simulations have been run assuming x to be Gaussian distributed or, alternatively, to be Beta-distributed; for the latter the shape of the distribution can be changed considerably with parameters. If the differences between expected values and between the variances of the distributions are not too different the plot of the scale values (coordinates on the first dimension provided by a CA) of the stimuli versus the expected values turned out to be linear with $r^2 > .96$; so it is fairly safe to interprete the coordinate values on the first dimension as some linear transformation of the conditional expected values of some underlying ("latent") variable with respect to which the stimulus patterns are evaluated. A full discussion of the simulations will be given elsewhere.
>
> The prediction $u_i = E(x|s_i)$ can be directly tested if stimuli differing with respect to a single physical attribute only are employed, e.g. which vary with respect to the carrier frequency f of a Gabor-patch. Such an experiment was carried out (Meinhardt, unpublished data), and the relation between the f_i and the corresponding u_i-values was perfectly linear ($r^2 = .98$), suggesting that the above conditions were actually met.

Remarks:

1. The u_i and the v_j are values on an interval scale; so if u_i is an estimate of $E(x|s_i)$ this value is an estimate that is unique up to a scale parameter and an additive constant.

2. Note that $\int_{A_j} xp(x|s_i)dx$ may be called the conditional expected value of x on A_j, given s_i was presented; w_{ij} then equals this conditional expectation, weighted by $1/p(x \in A_j|s_i)$. One may regard w_{ij} as a characteristic value of x for A_j, given s_i was presented. If (10) holds this characteristic value is independent of s_i and may thus be interpreted as the value of an indicator variable v, i.e. $v = v_j$ if $x \in A_j$, and $u_i = E(v|s_i)$, i.e. u_i is the conditional expectation of v if s_i was presented.

3. x may represent the combination of two stimulus dimensions, e.g. $x = a_1x_1 + a_2x_2$, or $x = x_1/x_2$.

Determination of scale values We consider the general case $x = (x_1, \ldots, x_d)'$, i.e. there may exist d stimulus dimensions. In this case DS has to be carried out such that more than a single set of scale values $(\hat{u}, \hat{v})$ have to be found. It may now be shown that DS is equivalent to Corrrespondence Analysis (CA) (e.g. Greenacre, (1984)), which automatically provides as many sets $(\hat{u}, \hat{v})$ as necessary. We have therefore employed CA in order to analyse the confusion matrices.

Let $x_{ij} = (n_{ij} - e_{ij})/\sqrt{e_{ij}}$, where the e_{ij} are the expected values under the null hypothesis of no relationship between stimuli and responses, i.e. $e_{ij} = n_{i+}n_{+j}/N$, N the total number of responses (stimulus presentations) in a table. CA consists basically (i) of a *Single Value Decomposition* (SVD) of the matrix $X = (x_{ij})$, and (ii) a re-scaling of the resulting scores for the row categories (here: the stimuli) and the column categories (here: the responses) such that the coordinates of stimuli and responses can be related to the χ^2 of the confusion matrix.

The SVD of X yields a the decomposition $X = Q\Lambda^{1/2}T'$ of X given by where Q is the $m \times m$-Matrix of eigenvectors of XX', T is the $m \times m$-matrix of eigenvectors of $X'X$ and $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m})'$, $\sqrt{\lambda_k}$, $1 \leq k \leq m$ the k-th eigenvalue of $X'X$ and $X'X$. The matrices Q contains scores of the s_i on some latent dimension, and T contains scores of the r_j on the same dimensions. These scores are re-scaled according to

$$F = D_r^{-1/2}Q\Lambda^{1/2}, \quad V = D_c^{-1/2}T\Lambda^{1/2}, \tag{13}$$

where D_r is the diagonal matrix of row sums n_{i+} of the confusion matrix (since all stimuli are presented equally often the elements of D_r in the diagonal are all identical, $n_{i+} = n$ for all i), D_c is the diagonal matrix of column sums n_{+j}. Let us, for simplicity's sake, rename in (3) the $\hat{u}_i$ into u_i again and the $\hat{v}_j$ into v_j. Thus

$$u_{ik} = q_{ik}\frac{\sqrt{\lambda_k}}{\sqrt{n}}, \quad v_{jk} = t_{jk}\frac{\sqrt{\lambda_k}}{n_{+j}}. \tag{14}$$

The χ^2-value of the confusion matrix is given by $\chi^2 = \sum_{ij} x_{ij}^2$. The *inertia* of a table is defined as χ^2/N. The u_{ik} and v_{jk}, representing the u_i and v_j in (3) for the k-th latent dimension; they have been defined such that each latent dimension "explains" a certain proportion of the total inertia χ^2/N, in decreasing order, i.e. the first latent dimension caters for the largest proportion, the second for the second largest proportion, etc. The inertia components explained by the latent dimensions correspond to the percent of variance explained in a PCA of measurements.

The stimuli and the responses may be simultaneously represented by points with coordinates given by (13), or (14); this is called a *biplot* (usually, two latent dimensions are sufficient to explain the data in a confusion matrix). The Eucledian distance between two stimulus points represents the proportion of the total inertia (or χ^2) generated by the difference between the two stimuli; the Eucledian distance of a stimulus point to the origin of the coordinate system represents the proportion of the inertia generated by that stimulus. An analogous interpretation holds for the distances between response points. The distance between a stimulus and a response point is not defined, however; stimuli and response points have to be interpreted with respect to the scalar products of the vectors pointing from the origin to the points.

As before, one may ask when the stimuli and the corresponding responses are presented by points at similar positions. (13) implies that in the extreme (perfect correspondence), $F = G$ if $D_c D_r^{-1} Q = T$. The matrix $D_c D_r^{-1} Q$ is diagonal with elements n_{+j}/n in its main diagonal, n the number of stimulus presentations. So the n_{+j} need not be identical and consequently the confusion matrix needs not be symmetrical. If the n_{+j} are identical, then $n_{+j} = n_{i+}$ for all i and j and $D_r = D_c$, i.e. $D_c D_r^{-1} = I$ the identity matrix so that $Q = T$. This means that stimuli and responses have identical scores. So $D_c D_r^{-1} \neq I$ reflects the bias towards certain responses.

As a guideline to find interpretations for the latent dimensions one may adopt (11), assuming that the Conditions for (10) are met. The latent dimensions then reflect directly the stimulus dimensions, that is the components x_k of x, i.e. the u_{i1} are the (more precisely: are linear transformations of) the conditional expectations of the representation of the stimuli on the first dimension x_1, and u_{i2} represent the patterns with respect to x_2, etc. With respect to the above Remark we may interprete the v_{jk} as values of the indicator variable $v_{\cdot k}$, i.e. $v_{\cdot k} = v_{jk}$ if $x_{jk} \in A_{jk}$.

3 Experiments

3.1 Subjects

Two subjects (male, 27 and 32 years of age) served as observes. Both are myope with corrected eyesight, and both are well acquainted with psychophysical experiments, in particular with detection experiments. Subject KF was naive with respect to the purpose of the experiment.

3.2 Apparatus

Patterns were generated using a VSG2/3 stimulus generator and displayed on a VM3640 14" grayscale monitor. A pattern was scaled in contrast using a table loaded with 512 entries equidistant in contrast where entry No. 511 pointed to the pattern in maximum contrast and entry No. 0 pointed to the pattern in zero contrast. Each of the 512 entries referred to a lookup table being a linear gray staircase consisting of 256 steps chosen from a palette of 4096 possible gray values, the medium step (128) always referring to gray value No. 2048. The relation between the gray level entries 0 to 4095 and the luminance on the screen was linearised by means of RGB-translation tables. This linearity was checked before each experimental session using a calibration program which determined the relation between the digital gray values of the VSG2/3 and luminance in cd/m^2 measured by an LMT 1003 photometer. The coefficient of determination of the regression line was in all cases greater than .98. The refresh rate of the monitor was 72 Hz at a horizontal frequency of 39 kHz, the pixel resolution was set to 640 $\times$ 480 pixels. The room was darkened so that the ambient illumination matched the illumination on the screen to a fair degree of appoximation. The mean luminance of the screen was set to 50 cd/m^2. Patterns were viewed monocularly at a distance of 180 cm. The subjects used a chin rest and an ocular. the ocular limited the visible area of the screen to a circular field of 4.6^0 in diameter. The subjects responded by pressing a button on an external response box.

3.3 Stimuli

Stimulus patterns were defined either according to

$$l(r) = l_0(1 + m\, s(r)), \quad r = \sqrt{x^2 + y^2}, \tag{15}$$

(identification experiment), or to

$$l(r) = l_0(1 + m(s(r) + q\, J_0(a\, r))) \tag{16}$$

(detection experiment; r as defined in (15)), where l_0 defines the space average luminance of the screen, and $m = (l_{max} - l_{min})/2l_0$ (maxwell contrast). $s(r)$ is defined as the superposition of two radially symmetric "discs" d_1 and d_2, specified by

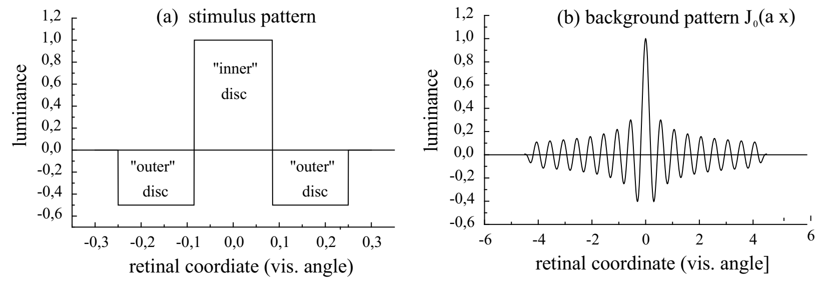
$$d_k(r) = \begin{cases} 1, & r \leq c_k \\ 0, & r > c_k \end{cases}, \quad k = 1, 2 \tag{17}$$

and $s(r)$ is defined as

$$s(r) = (1 + \alpha)d_1(r) - \alpha d_2(r), \quad \alpha = 1/2. \tag{18}$$

i.e. $s(r)$ defines two two concentrically superimposed circular discs of different radius

Figure 1: Example of stimulus (a) and background pattern (b): luminance profiles



and luminance. For the identification experiment four patterns $s(r)$ were defined: r_1 could assume either the value r_{11} or r_{12}, and r_2 could assume the values r_{21} or r_{22}; a particular pattern s_i is defined by a combination (r_{1i}, r_{2i}), so there are four possible patterns. The actual values of the r_{ij} are summarized in table 3. These patterns were

Table 3: Stimulus parameter for the identification experiment; c_1, c_2: radii of components.

Stimulus	c_1	c_2	c_1/c_2	c_2/c_1
bb	.25	.075	3.333	.300
bs	.25	.07	3.571	.280
sb	.23	.075	3.067	.320
ss	.23	.07	3.285	.304

not presented superimposed on a Bessel-pattern. Thus the total width (diameter) of the stimulus pattern is defined by the "outer" disc and was either $2c_1 = .5^0$ or $2c_1 = .46^0$. The "inner" disc had either the diameter $2c_2 = .15^0$ or $2c_2 = .14^0$.

In the detection experiment only the pattern defined by (c_{1j}, c_{2j}) was employed. The parameter a in the background pattern J_0 is defined as $a = 2\pi f$, and f could assume the values 2.0, 2.5, 3.0 or 3.5 c/deg; these values are dictated by the spectrum of the "test"pattern, since the background patterns will have no effect on the detectability of the testpattern if the frequency parameter f has zero amplitude in the spectrum of $s(r)$.

3.4 Procedure

Pre- and posttest detection thresholds. The pre- and posttest measurements aimed at data collection for a measure of sensitivity for superimposed patterns composed of the small aperiodic test pattern $(r_{12,}r_{22})$ and Besselfunctions of zero order and various spatial frequencies. The measure of sensitivity introduced here follows the standard of Kulikowski's and King-Smith's summation to threshold paradigm. Briefly, let the response to $m(s(r) + q\,J_0(a\,r))$ be given by $g(r) = m(h_t(r) + q\,h_b(r)$, where h_t is the unit response to the "test"-pattern $s(r)$ (i.e. the response to $s(r)$ with unit contrast) and h_b the unit response to the background pattern J_0. Suppose the pattern is detected at the position r_0, where g assumes its maximum, and if $g(r_o) = c$ a certain constant. Let $m_1 = m$, $m_2 = q\,m$. Dropping r_0 for simplicity it follows, with m_{01} the threshold contrast for $s(r)$ when presented without the background pattern, $m_1 = m_{01} - m_{01}m_2h_b/c$,so for different values of m_2 the value of m_1 should be a linear function of m_2, provided the values of m_2 are not too large; this is the *contrast interrelation function* (CIF)). The slope of this CIF is proportional to h_b, and h_b is the response to the background pattern J_0. If the detecting channel is radially symmetric, then the response $h_b(r)$ is proportional to the input $J_0(ar)$, and the proportionality factor equals the system function of the detecting channel. If the detecting channel is a matched filter for the stimulus pattern $s(r)$, then the system function is proportional to the Hankel transform (i.e. the Fourier transform in case of radial symmetry) of $s(r)$. So the slopes of the CIFs for different values of $s = 2\pi f$, i.e. of f should be proportional to the Fourier spectrum of $s(r)$.

Contrast threshold measuring procedures. For the pretest and the posttest sensitivity measurements two experimental sessions were carried out with each subject. Within each experimental session the contrast thresholds for compound patterns (superpositions of the test pattern on Besselfunctions of zero order and each of the six possible spatial frequencies) and the testpatten alone were determined. For each spatial frequency f four values of q were employed in order to estimate linear approximations to contrast interrelation functions. These values were $q_1(f), q_2(f), -q_1(f), -q_2(f)$, where $q_2 = q_1/2$. Negative factors of q denote a 180^0 phase shift of the background pattern (contrast reversal technique, c.f. Kulikowski and King-Smith 1973). The factor $q_1\,(f)$ was always chosen such that the relative component strength of the Besselfunction was not greater than 40% of its threshold contrast. In order to determine the appropriate values of q for each spatial frequency of the Besselfunction, measurements of the thresholds for simple J_0 targets were carried out prior to measuring compound pattern thresholds. For each experimental session the computer generated a random list of all superimposed patterns, i.e. the subject was not able to generate valid hypotheses about the sequence of spatial frequencies of the superimposed patterns. Contrast thresholds were measured with the method of limits. This was done as follows: The initial contrast was set to the starting value. For the first measurement, this was a value well beyond threshold, for the following measurements for the same pattern, this value was the mean of all foregoing threshold contrasts for that pattern, plus 25% of contrast. Then the first down-run started: The contrast was decremented using a temporal staircase until the subject signaled that the signal was no longer visible by pressing a button on a small response keyboard. The temporal staircase comprised 512 contrast steps (equidistant in contrast, s. above), each of them with a contrast amplitude of 3.9×10^{-5} and 24 msec duration. The image was always present on the screen, it was not temporally gated from one contrast step to the next (rectangular temporal staircase). Then the contrast was diminished by 25% and the contrast was incremented using the temporal staircase until the subject signaled that the pattern was barely distinguishable from the screen background. The average of both threshold contrast values m_{up} and m_{down} was taken and used as the contrast threshold m of this measurement. At least twelve threshold measurements were carried out for each pattern, the mean of the measurements served as estimate of the threshold contrast. The subjects were advised to press the button if they had the impression that the screen deviated in any way from the mean pale gray, or turned to pale gray again, respectively, and to fix the center of the screen, which was marked by a small permanent dot. Each subject made one experimental session per day, i.e. pretest and posttest, respectively, were done at two consecutive days.

3.5 The identification experiment

Firstly, the subjects were trained to establish the correct assignment of response button and stimulus pattern. 20 training trials proved to be sufficient.

The stimuli were presented in pseudo random order; within a single session each stimulus was pesented $n = 25$ times, so there were altogether $4n$ stimulus presentations and the same number of responses. The contrast of the patterns was held constant at $m = .18^0$; this is about six times the detection threshold of a pattern.

A trial was initiated by a tone; after 500 ms a stimulus pattern appeared for the duration of 560 ms. The subject responded by pushing one of the buttons 1 to 4 according to his decision about the presented stimulus pattern. After his response he was provided with a feedback about his decision: one of four possible tones indicated which pattern had actually been shown.

In an experimental session each pattern was presented 25 times. For each session, the responses were summarised in *confusion matrices* having the form given in table 4, where each row contains the number of a certain stimulus pattern. The notation bb, bs

Table 4: Confusion Matrix

	BB	BS	SB	SS	$\sum$
bb	n_{11}	n_{12}	n_{13}	n_{14}	n
bs	n_{21}	n_{22}	n_{23}	n_{24}	n
bs	n_{31}	n_{32}	n_{33}	n_{34}	n
ss	n_{41}	n_{42}	n_{43}	n_{44}	n
$\sum$	n_{+1}	n_{+2}	n_{+3}	n_{+4}	N

etc represents the components defining a pattern. Since the radius of a disc can assume one of two possible values they may be coded with either b (*big*)or s (*small*). So bb represents the pattern with $.5^0$ diameter "outer" disc and $.15^0$ diameter of the "inner" disc, bs represents the pattern with $.5^0$ diameter of the outer and $.14^0$ diameter of the inner disc, etc.. Capital letters indicate responses: BB thus represents the response that the pattern bb was presented, etc.. n_{+j}, $j = 1, \ldots, 4$ represents the sum of the n_{ij} for a given value of j, and $n = 25$, $N = 100$.

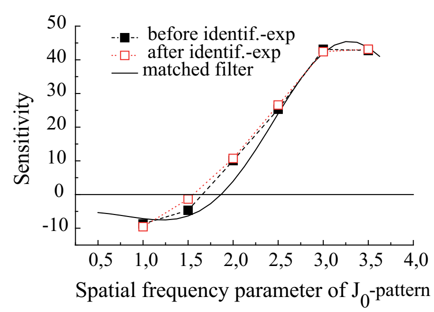
Subject KF provided confusion matrices of the form given in table 4 at five consecutive days, with three sessions each, i.e. altogether 15 matrices. Subject GM provided a single confusion matrix per day at five consecutive days. Usually, identification experiments comprise many more sessions than our experiment, and data from practise trials or sessions are not included in the analysis[2]. In our experiment we collected far less data because we were mainly interested in the effects of this experiment on the detection performance. Still, the data show stable aspects pointing to interesting research perspectives.

4 Results

4.1 Detection data

The results of the detection experiments - one before and one after the identification experiment - are summarised in Fig. 2 (subject KF, the data of subject GM are similar).

Figure 2: Results of the detection experiments



Obviously, the sensitivities of the pre- and post test do not differ. The prediction of the matched filter model is presented as well, the data appear to be compatible with this model.

It follows that the identification experiment does not seem to have any effect on the detection data. Further, if one assumes detection by a matched filter, then this filters

[2]For instance, Falmagne (1972) discusses a confusion matrix from an identification experiment with 6 alternatives based on altogether 54 000 trials, of which 539452 trials were actually analysed

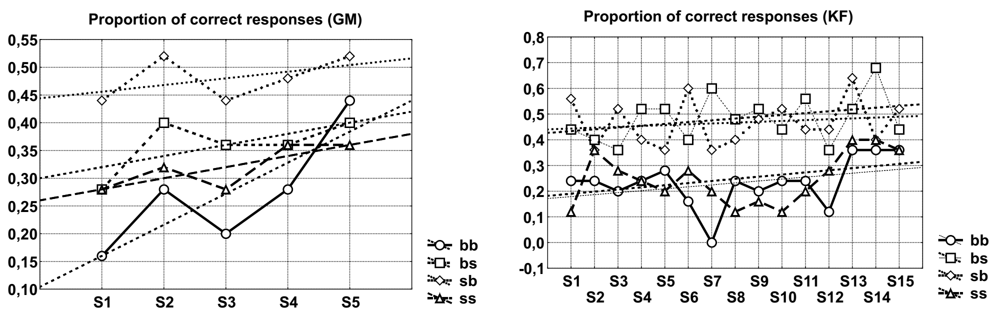
appears to exists from the first measurements on. If the matched filter is formed during the course of the experiment one has to conclude that it forms fastly.

4.2 The identification data

4.2.1 Correct responses

To begin with, let us consider the change of the proportion of correct responses with increasing number of experimental sessions. Fig. 3 shows the data for subjects GM and

Figure 3: Correct responses



KF. The regression lines are not meant to represent a model of learning; they are just meant to indicate the *trend* in the data (which may even be nonlinear). According to these trends the subject do learn. Note that the regression lines are about parallel, indicating a similar learning rate for the different patterns, - with the exception of the line for the pattern *bb*, subject GM. On the other hand, the parallelity of the regression lines should not be overinterpreted; for *bb* and *sb* positively accelerated curves fit the data better than linear functions.

For both subjects, the pattern *bb* and *ss* appear to be the more difficult to identify than the patterns *bs* and *sb*. While the possibility that after continued learning the probabilities of correct identification become identical it remains to state that the improvement of the identification performance for the patterns *ss* and *bb* is paralleled to an improvement for the patterns *bs* and *sb*. This indicates that the pairs (bs, sb) and (ss, bb) differ with respect to some aspect of their respective sensory representations that makes it easier to discriminate *bs* or *sb* from the alternatives than to discriminate *ss* or *bb* from the alternatives.

4.2.2 The all-or-none, the biased similarity and the independent decisions model

Figures 4 and 5 show the results for the all-or-none, the biased similarity and the

Figure 4: χ^2-values for confusion matrices, test of the independence, the biased similarity and the all-or-none model (GM)

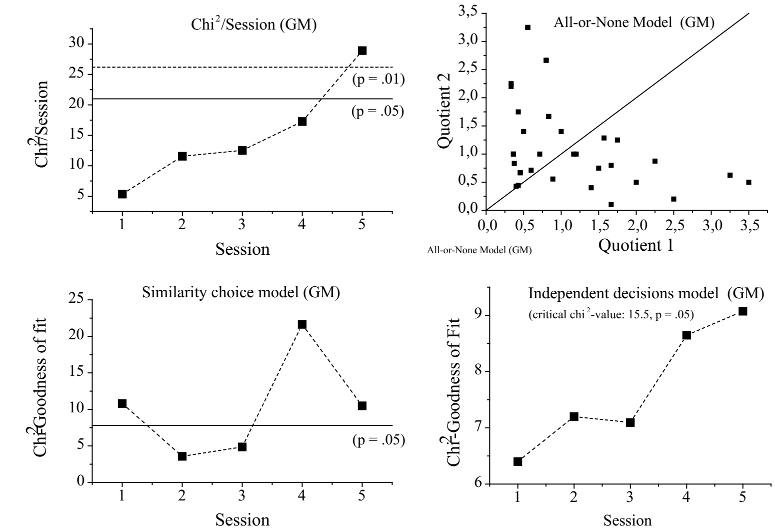
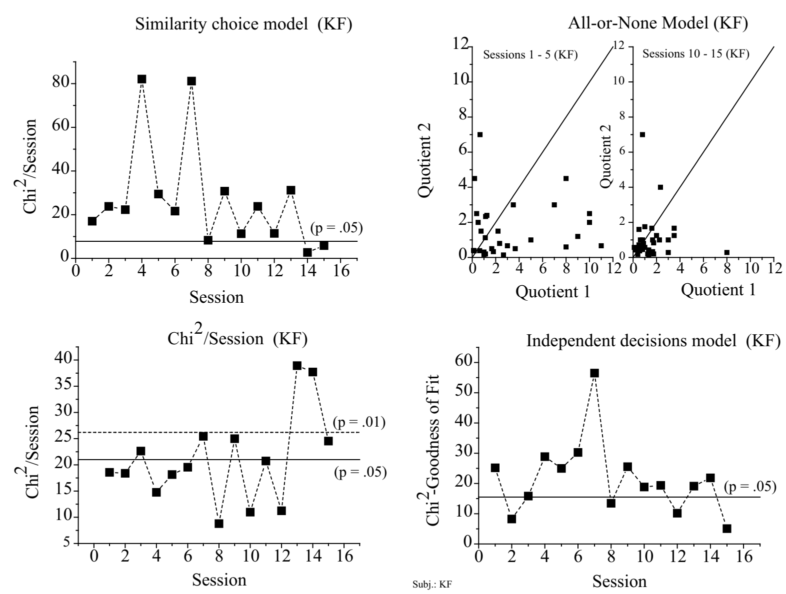


Figure 5: χ^2-values for confusion matrices, test independence, the biased similarity and the all-or-none model (KF)



independent decisions model, for the subjects KF and GM, together with the χ^2-values for the confusion matrices resulting from the individual experimental sessions.

For subject GM the χ^2-value representing the dependencies between stimuli and responses reaches a significant value only in the fifth session. The all-or none model cannot be fitted at all; the attempt to find the parameters p_i and b_j by minimizing the χ^2 of a confusion matrix yields negative estimates for some of the p_i[3]. An alternative, parameterfree test was devised computing the quotiens qq; they yould correspond to the straight line with slope 1 in Fig. 3 (b). The results illustrate the lack of fit. The similarity choice model is compatible with the data only for two sessions, where, however, the χ^2 representing the dependencies among stimuli and responses (Fig. 3 (a)) is not significant, so the model can be discarded for this subject. The independent decisions model is compatible with the data from all sessions (Fig. 3 (c)); note, however, that the χ^2 for the goodnes-of-fit is increasing with the overall χ^2. Although the model is compatible with the data from the fifth session - the only session with a significant overall χ^2 - the model may become inadequate with further learning; it seems that the increase of dependencies among stimuli and responses implies a worsening of the fit of the model.

With subject KF the situation is slightly different. The χ^2-values fluctuate between being significant and nonsignificant, and increase substantially for the 13-th and 14-th session; however, at the 15-th session there is again a considerable drop of the χ^2, which remains significant on the 1%-level, however.

As with subject GM, the all-on-none model is not compatible with the data; there do not even exist meaningful values for the parameters p_j, attempts to find them yield negative values. We present the plot Fig. 4, (b) to illustrate the lack of fit.

The biased similarity model becomes compatible with the data from the last two sessions. The independent decisions model is compatible with the data in the non trivial sense, i.e. when the overall χ^2 is significant, from the the sessions 8 and 12.

Now authors who have successfully fitted models like the ones considered here usually fitted the data after a large number of training sessions (Townsend and Ashby (1982), Nosofsky et al (1986)). So one may argue that the attempt to fit such models to our data is inadequate for a start since the subjects were still learning their task.

Let us look at the sessions 2, 7, 9, possibly 11, and 12, 13, 14. The overall χ^2 for these sessions are significant, meaning that the subject follows a certain decision strategy in a systematic way, - a change of strategies would result in confusion frequencies looking like randomly generated decisions and consequently in low χ^2-values. Inspection of the results of the CAs yields further insight into the nature of the dependencies.

4.2.3 Correspondence analyses

Fig 6 shows the biplots of subject GM, and Figs 7, 8 and 9 show the biplots of subject

[3]The Mathematica-routine FindMinimum was used.

Figure 6: Biplots, Subj. GM. See text for explanation.

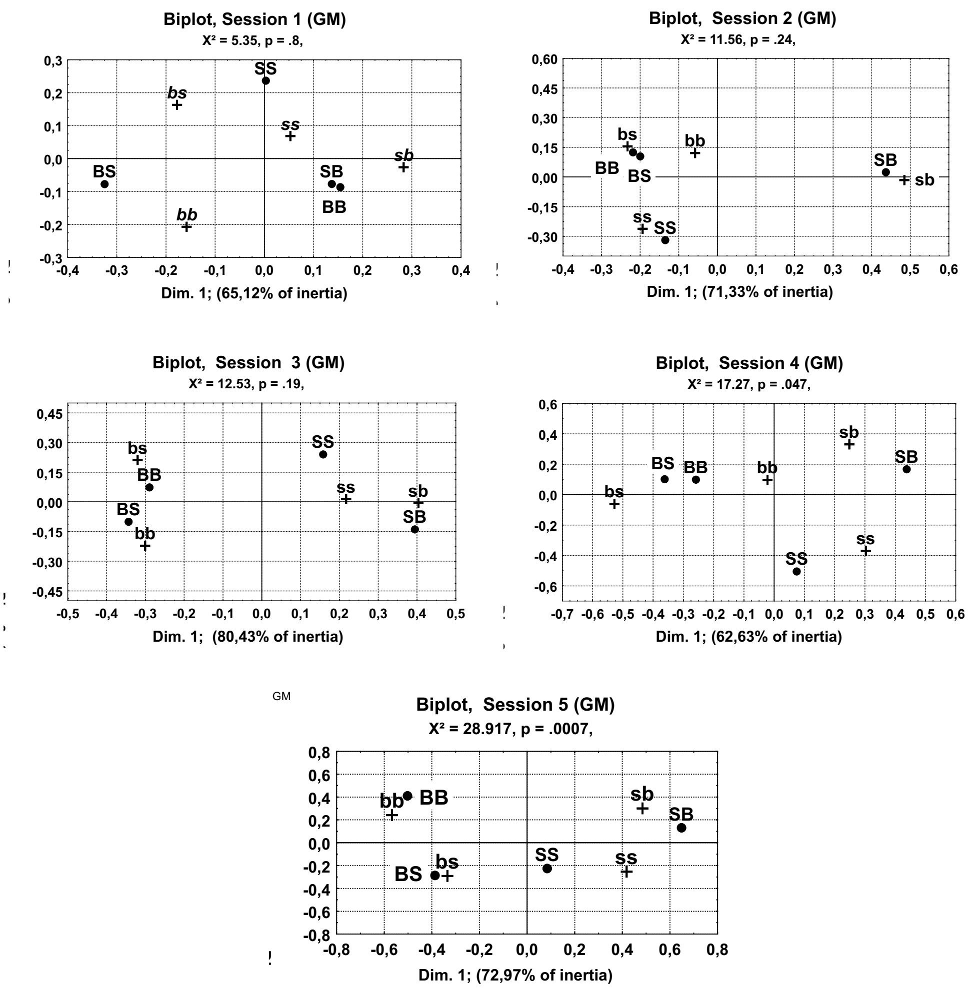
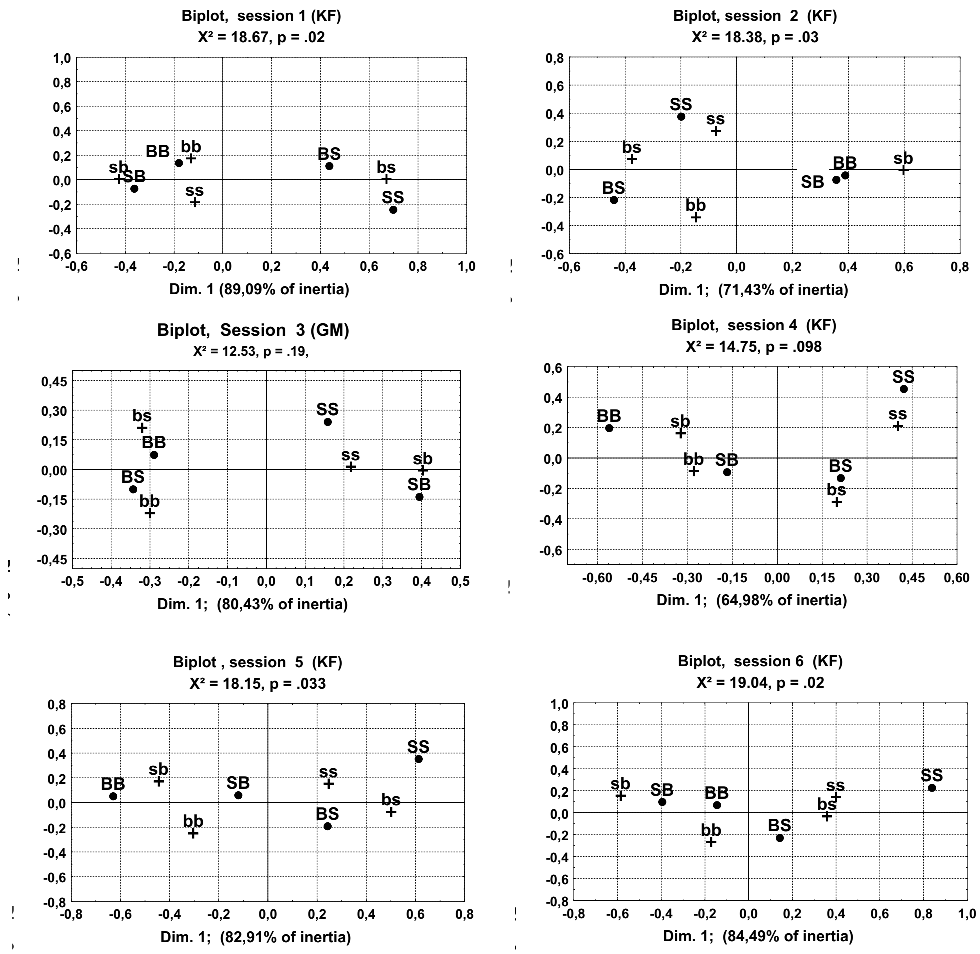


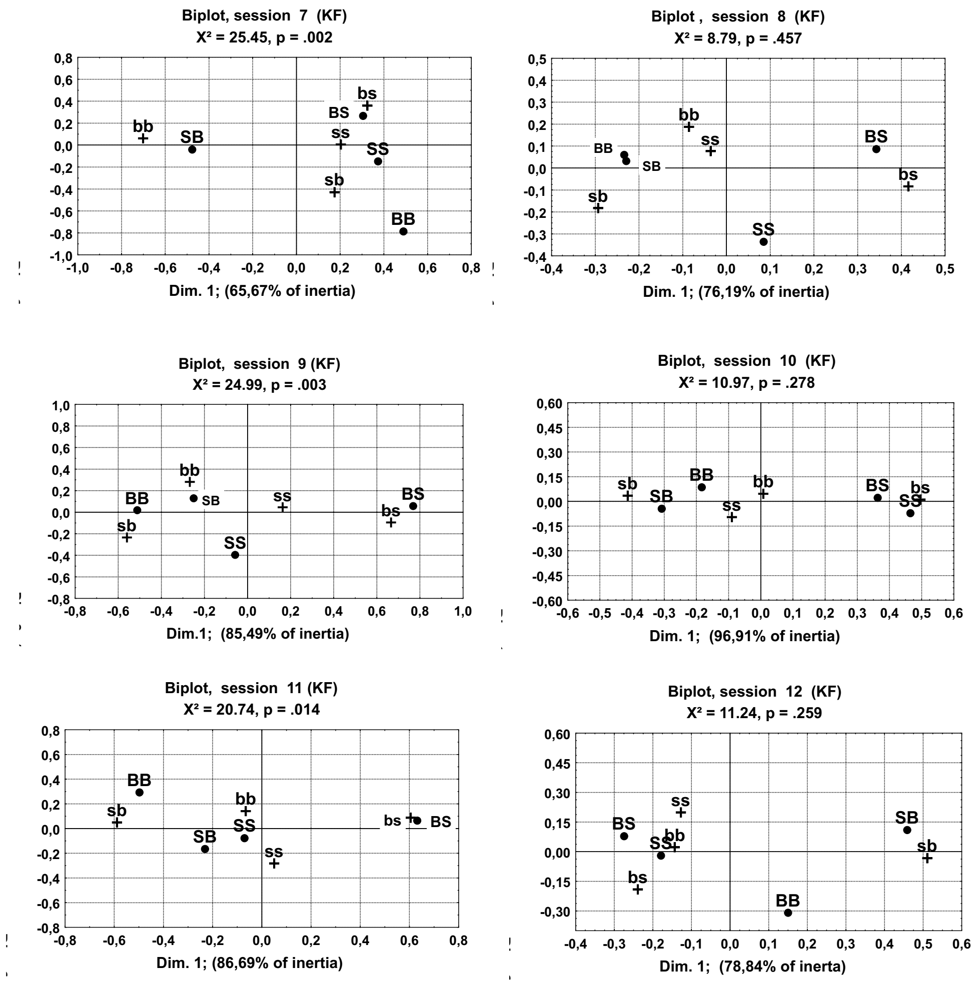
Figure 7: Biplots, Sessions 1 - 6 (KF). See text for explanation.



KF.

One may conclude from Fig. 3 that subjects do learn over the sessions so that the number of correct decisions increases. This means that in the biplots the points representing a stimulus on the one hand and the correct response on the other should be closer together the more sessions the subject has served. We have presented the biplots also for sessions with a nonsignificant overall χ^2 in order to show the variation of the plots between sessions. Fig. 9 shows also the biplot predicted by the independent decisions model for the 15-th session of subject KF. Recall that for this session this model was compatible with the data. The predicted biplot was found by estimating the frequencies in a confusion matrix on he basis of the estimated probilities for identifying the components correctly. The comparison of the actual biplot and the predicted my give an impression of the distribution of scale values (the sampling distribution of the

Figure 8: Biplots, Sessions 7 - 12 (KF). See text for explanation.



scale values does not seem to be known).

The question is now whether anything systematic exists in the biplots. Further inspection of them reveals that the first latent dimension (the one explaining the larger proportion of the total inertia (χ^2/N) is often defined by the pair *bs* and *sb* of the stimulus pattern. The second latent dimension is less clear, but there appears to be a trend that it is specified by the pair *bb* and *ss*. To see more clearly what sort of information the biplots contain the coordinates of the patterns *bb*, *bs*, *sb* and *ss* as well as of the responses *BB*, *BS*, *SB* and *SS* on the first latent dimension have plotted and are presented superimposed in Figures 10 and 11. For both subjects there seems to exist a rather clear pattern: on the first latent dimension the coordinates for the pattern *sb* tend to be maximal for subject GM and minimal for subject KF; for the pattern *bs* the

Figure 9: Biplots, Sessions 13 - 15, and prediction of the independent decisions model (KF). See text for explanation.

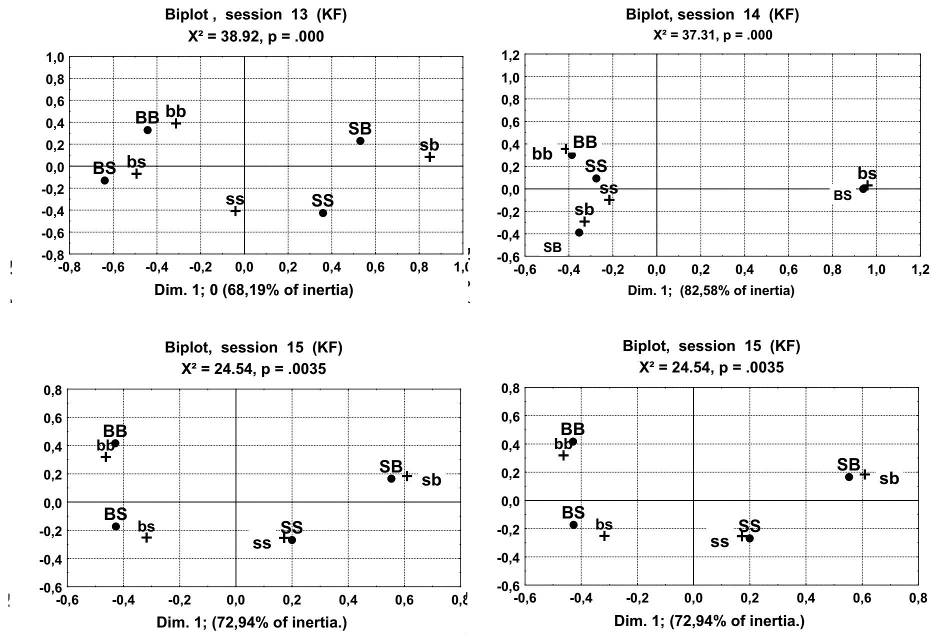
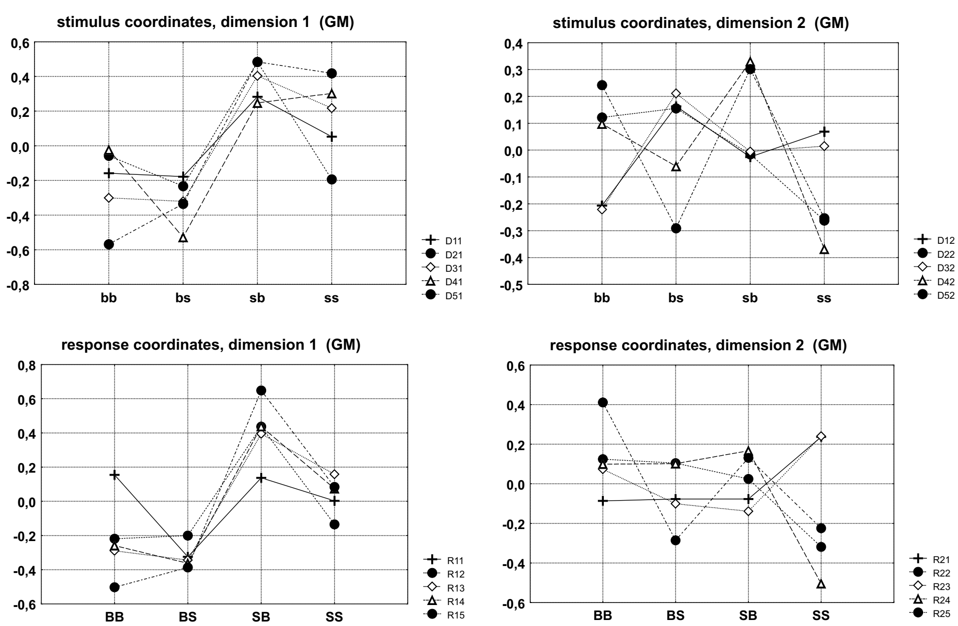


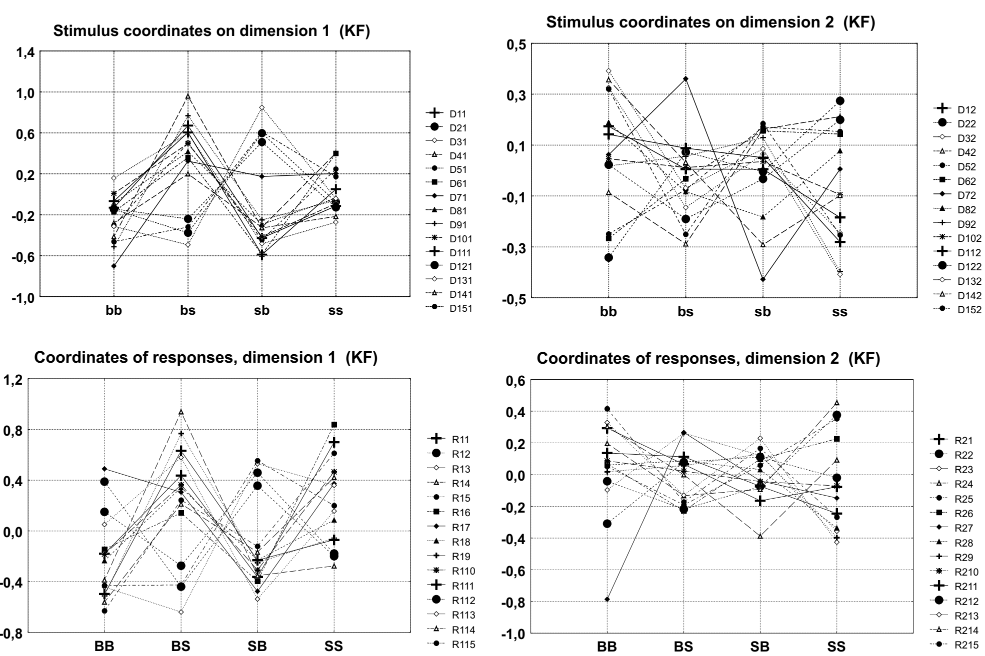
Figure 10: Coordinates for sessions 1- 5 (GM)



reverse holds; note that the coordinates follow the general pattern also in sessions with non-significant χ^2-values. For subject KF there are a few sessions where the coordinates of *bs* tend to be minimal and those for *sb* are maximal. Such reversals are irrelevant in so far as the meaning of the latent dimension appears to remain invariant. The coordinates of the responses on the first latent dimension follow the pattern of the corresponding stimulus patterns. The fact that the patterns *bs* and *sb* seem to define the main latent dimension corresponds to the finding that these patterns were easier to identify in all sessions (c.f. Fig. 3).

The plots of the coordinates on the second latent dimension show a less clear pattern: while the first latent dimension seems to govern the decisions in all sessions, the second latent dimension appears to be too fuzzy to play an invariant role in different sessions. For subject GM the proportions of the inertia χ^2/N of the confusion matrices due to the second dimension have values between 18.7% and 37.33% with a mean of 28%; these proportions seem to be sufficiently large to presume that the second dimension does

Figure 11: Coordinates for sessions 1 - 15 (KF)



indeed reflect some relevant aspect of the decisions; for subject Kf the situation is similar. We will offer an interpretation of the second dimension after the following considerations.

In particular for subject KF the first latent dimension appears to be characterised by the pair (*sb*, *bs*); this is clearly so for all sessions except sessions 4 and 7, i.e. for about 87% of the sessions. Note that the overall χ^2 needs not be significant for a general structure to become visible in the biplots. The proportion of inertia explained by this dimension is often very high compared to that explained by the second dimension: the proportion varies between 14% and at most 20% with the exception of session 7, where the second dimension accounts for about 30% of the inertia; in session 10 the second dimension accounts for only 3% of the inertia.

To pursue the question of the meaning of the dimensions a bit further let us start from the hypothesis that it is the *ratio of the diameter or radius* of the two superimposed discs that define the major criterion for identifiying the patterns. The ratio of the radii of the outer to the inner disc are, in ascending order:

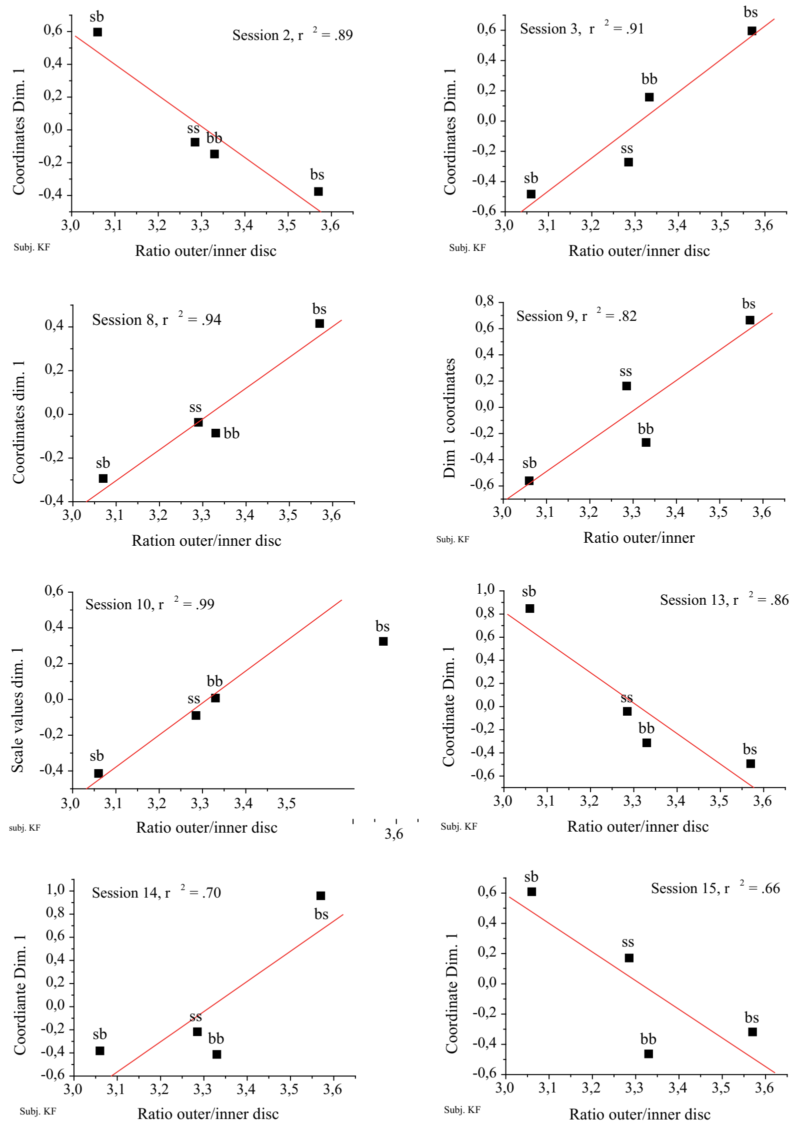
$$3.067\,(sb), \quad 3.289\,(ss), \quad 3.333\,(bb), \quad 3.571\,(bs), \tag{19}$$

Equivalently, one could consider the reprocal ratios; the relation between the patterns remains, of course, invariant. There is a unique correspondence between patterns and ratios, so it is possible at least in principle to identify the patterns by evaluating the value of the ratio; if the subject consistently follows the strategy to identify stimulus patterns with respect to this ratio correspondence analysis would represent stimuli and responses on a single dimension; however, the ratios for the patterns *bb* and *ss* are very similar, and this could mean a high confusion probability for these two patterns.

Suppose that the first dimension reflects the evaluation of the stimuli with respect to this ratio (c.f. the remark 2. in section 2.2.4). One may therefore plot the coordinates of the four patterns on the first dimension versus the value of the ratio of the radius of the greater and the smaller disc ("outer/inner disc"). Fig. 12 shows a sample of the plots from one of the subjects (KF). The plots show the data conforming to the hypothesis that the first latent dimension is defined by the quotient. For the remaining sessions the value of r^2 is between .005 (session 7) and .77 (session 12). The point is that the sessions for which the hypothesis may be accepted do not form a block of consecutive sessions, indicating that the subject may change his strategy from one session to the next. Note that for session 10 the data conform extremely well with the hypothesis, although the overall χ^2 is not significant, indicating that the amount of "noise" in the data is large compared to the structure underlying the subject's decisions. Note also that the values of the quotients for the patterns *ss* and *bb* are similar, and that CA yields scale values for these patterns on the first dimension that are, for the data shown, also very similar; in the remaining sessions the scale values for *ss* and *bb* may differ considerably. Further, for the last two sessions (14 and 15) the data do not agree with the hypothesis; - in session 15 the independent decisions model fits, which is, of course, incompatible with the quotient-hypothesis. In so far the data are internally consistent.

It remains to consider what the second dimension could represent. If the first dimension is mainly defined by the patterns *bs* and *sb* (possibly via the ratio of the components) one could think of the patterns *bb* and *ss* as possibly defining the second dimension. If the subjects do indeed evaluate the patterns with respect to some function of the repre-

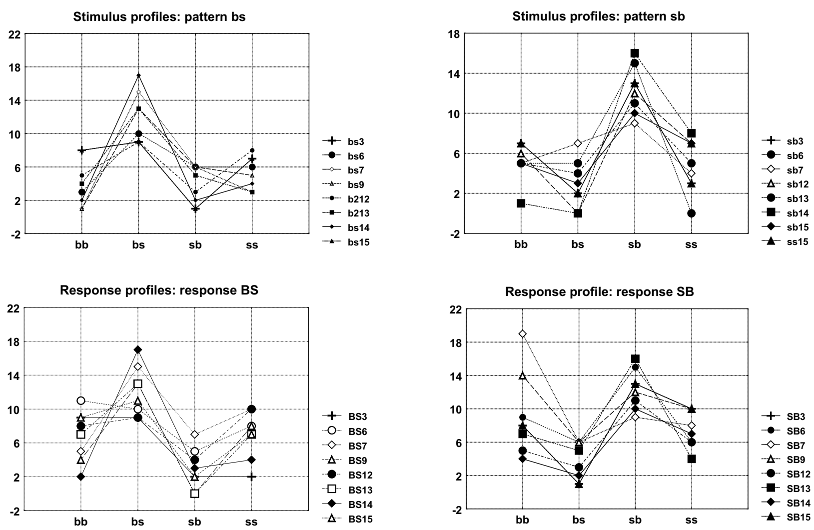
Figure 12: Stimulus dimension: quotient of components for sessions with significant χ^2, (KF)



sentations of the pattern components one has to suspect that either the subject employs two such functions, or that the second dimensions represents, in a way that has to be specified, just noise, i.e the effects of uncertainty how to identify certain patterns.

To clarify this point let us recall that the configuration of points representing stimuli and responses is determined by what is called the row- and column profiles in CA. In our case, a row profile corresponds to a stimulus, and a column profile to a response. A row or stimulus profile is given by the relative frequencies $n_{i1}/n_{i+}, \ldots, n_{im}/n_{i+}$, and a response profile by the relative frequencies $n_{1j}/n_{+j}, \ldots, n_{mj}/n_{+j}$. The distance between

Figure 13: Stimulus and response profiles for the patterns *bs* and *sb* (KF)



two stimulus points will be the smaller, the more similar the profiles of the stimuli are, and the analogous statement holds for the points representing the responses. The scalar product of the vectors pointing from the origin of the coordinate system to two stimulus points will be about zero, if their profiles have about zero covariance. Let us now look at the profiles of the patterns *bs* and *sb* on the one hand and the profiles of *bb* and *ss* on the other. The stimulus and the response profiles for the patterns *bs* and *sb* show a clear pattern, corresponding to the pattern of coordinate values on the first dimension. For the patterns *bs* and *sb* the maximum number of responses occurs for the responses BS and SB. For the patterns *bb* and *ss* the situation is different. For *bb* the maximum number of responses is not at BB, and similarly for the pattern *ss*. In Fig. 15 the mean number of responses and the corresponding variances for the four patterns have been plotted. So on average the response given to *bb* is SB, the response with second highest average is BS; the variances of the frequencies of responses is maximal for *bb*, with maximal variance for the response SB. For the pattern *ss* the situation is slightly different: the responses BS, SB and SS are chosen with about equal frequencies, and BB

Figure 14: Stimulus and response profiles for the patterns *bb* and *ss* (KF)

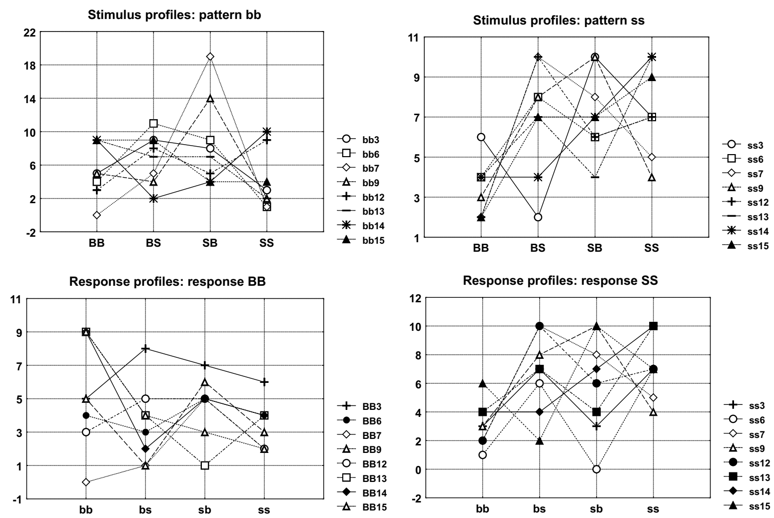
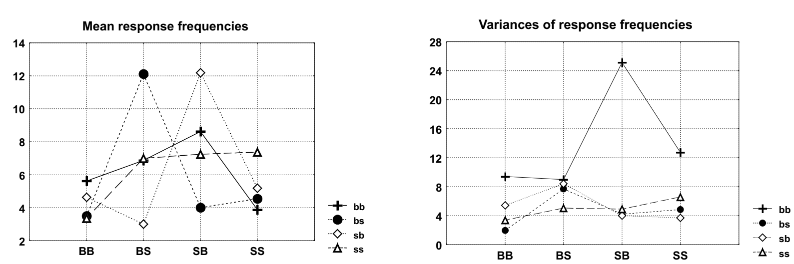


Figure 15: Means and variances of response frequencies (KF)



is chosen comparatively rarely. In any case, the mean response profiles for the patterns *bb* and *ss* are quite different from the mean response profiles of *bs* and *sb*. So on average the correlation between the profiles for *bb* and *ss* with those of *bs* and *sb* will be about zero, and therefore the CA will indicate a second dimension. But this dimension does not represent the values of some systematically evaluated function of the component representation, but just the uncertainty with respect to the response to be given when either *bb* or *ss* is presented. This is very plausible when one looks at the values of the ratios of the component measures: with respect to these values, *bb* and *ss* are hard to distinguish from each other: according to Table 3 the ratio of the radii of the "outer" to the "inner"disc is 3.333 for *bb* and 3.289 for *ss*; the reciprocal values of these quotients are .300 and .304, - the subjects may equally well judge according to the quotient of the radii of "inner" to "outer" disc. So if the quotient of the radii is the criterion variable *bb* and *ss* are very hard to distinguish. Further, the quotients "inner" to "outer" disc radius for these two patterns are just between those for the patterns *bs* and *sb*, which are for *bs* $1/3.571 = .280$ and $1/3.067 = .320$ for *sb*, implying that *bb* and *ss* are also hard to distinguish from *bs* and *sb* with respect to the quotient-criterion. However, this criterion does not seem to be the full story yet, since *ss* is not very often confused with *bb*, compared to the number of confusions with *bs* and *sb*. So the absolute sizes of the components may enter the evaluation of a pattern as well, possibly dependent on the size of the components: - for the smaller pattern such an evalution appears to be simpler than for the larger pattern.

5 Discussion

Let us summarize the main findings:

1. The identification experiment has no influence on the detection performance; moreover, the detection data are compatible with the hypothesis that the pattern is detected by a matched filter for this pattern.

2. While sensory learning seems to be fast - there is no difference between the threshold curves from the two detection experiments, and if one assumes that matched filters are formed during the experiment, they must be formed during the first trials - identification learning seems to require many more trials or sessions. Although the possibility that sensory learning may also play a role in identification processes, the main difficulty here appears to result from (i) gestalt-effects inherent in the sensory representation, and (ii) finding the optimal decision boundaries separating the possible sensory representations generated by the different patterns.

The sensory representation of the patterns seems to be such that the appearance of the individual components is in some way correlated. The patterns *bs* ans *sb* are easier to identify than the patterns *bb* and *ss*. If it justified to assume that the first latent dimension revealed by the CAs represents an aspect according to which the patterns are evaluated than the data suggest that this aspect is determined by the value of the ratio of the sizes of the components. Since *bb* and *ss* have very similar values with respect to this variable the are confused with other patterns far more often than the patterns *bs* and *sb*. It was argued that the second latent dimension suggested by the CAs represents not much more than the difficulty to identify the patterns *bb* and *ss* simply because the stimulus ("row"-) profiles for these two patterns are almost always uncorrelated with the profiles of *bs* and *sb*.

Now the subjects knew that the components of the pattern varied independently over trials, so they tried to optimize their decisions by evaluating each component independent of the other. That seems to be a hard task. Subject GM reported to have tried this independent evaluation from the start of the experiment on; however, to the extent the overall χ^2s for the confusion matrices increased, the χ^2 representing the goodness-of-fit for the independent decisions model also increased. It is as if the subjects had to work against what was perceptually given to them.

So one may deduce that the sensory representations generated by the patterns are characterized by strong gestalt-effects. The fact that the independent decision model fits the data of sessions[4] 2 and 15 of subject KF does not yet mean that the stimulus components - the discs - are now *perceived* independent of each other. To see this one may refer to the work of Ashby and Townsend (1986). These authors investigated the notion of perceptual independence and their findings seem to be of importance for the interpretation of our results. Let $f_i(x_1, x_2)$ be the density function characterizing the sensory representation of the stimulus pattern s_i. We may adopt Ashby and Townsend's definition of perceptual independence, namely that the components are perceptually independent if $f_i(x_1, x_2) = g_i(x_1)g_i(x_2)$, with $g_i(x_1)$ and $g_i(x_2)$ the marginal distributions of f_i. Further, Ashby and Townsend introduce the notions of *perceptual separability*, as defined by Garner and Morton (1969), and *decisional separability*: two components of a pattern are perceptually separable if the effect of one component does not depend upon the level of the other. The components are decisionally separable if the decision about one component does not depend upon the level of the other; so our model of independent decisions represents decisional separability. Decisional separability implies, as Ashby and Townsend point out, that the decision bounds are parallel to the coordinate axes defined by component representation X_1 and X_2. They prove (their Theorem 4) that if (i) $x = (x_1, x_2)$ is multivariate Gaussian, (ii) the expected values μ_{ij} are pairwise different, and (iii) the subject behaves optimally in the sense of maximizing the probability of answering correctly, i.e. if she or he operates according to (5), then *perceptual and decisional separability* imply *perceptual independence*.

[4] These are the sessions with a nontrivial fit of the model, since the overall χ^2 is significant for these sessions. IN session 8 the overall χ^2 is not significant.

Let us now assume in particular that the $f_i(x_1, x_2)$, $i = 1, \ldots, 4$ are 2-dimensional Gaussian with pairwise different mean values μ_{ij} and that the subjects behave optimally. If, in an identification experiment, the stimuli are coded in the same way as in a detection experiment (c.f. Thompson (1985) for an elaboration of arguments supporting this assumption), our results with respect to the matched-filter hypothesis suggest that the components are *not* perceptually independent; this assumption is implied by models describing the formation of matched filters as an adaptation of the receptive fields of cortical neurons (Mortensen and Nachtigall (1999)), so the stimulus patterns are encoded by highly correlated neurons, implying that the variance-covariance matrices Σ_i of the f_i are not diagonal. Let us further assume that the subjects behave optimally in the sense of maximizing the probability of correct responses. Ashby and Townsend's theorem then implies lack of perceptual separability or lack of decisional separability, or both. So if the independent decisions model holds we may assume that the patterns are *decisionally separable*, and we have to conclude that they are *not perceptually separable*.

The question behind our detection experiments refers to the way patterns are represented in the visual system. Even if the hypothesis of coding by matched filters or by cell assemblies where each neuron is a matched filter for that aspect of the pattern that covers the receptive field of the neuron, we still have to model in greater detail how the activity of such assemblies is evaluated by cognitive systems having this activity as input. Attempts to disentangle these activities will be the center of future work.

6 Appendix

Correspondence analysis

Correspondence analysis may be thought of as a method of finding the scale values u_i and v_j in the general, multidimensional case. If there are more than one underlying attribute or "latent dimensions", the stimulus pattern s_i will be represented by the scale values $u_{i1}, \ldots, u_{is}$, say, where s is the number of latent dimensions. In the same way the responses will be represented by scale values $v_{j1}, \ldots, v_{js}$.

Let n_{ij} be the number of confusions of stimulus pattern s_i with the pattern s_j, let $n_{i+} = \sum_j n_{ij}$ and $n_{\cdot j} = \sum_i n_{ij}$ and $N = \sum_{ij} n_{ij}$, and let

$$x_{ij} \stackrel{def}{=} \frac{n_{ij} - n_{i+}n_{+j}/N}{\sqrt{n_{i+}n_{+j}/N}}. \qquad (20)$$

so that

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}^2, \quad df = (I-1)(J-1) \qquad (21)$$

Let $X = (x_{ij})$ be the matrix of x_{ij}-values. CA consists essiantially of (i) the *Singular Value Decompositon* (SVD) of X, and (ii) a re-scaling of the resulting scores for the stimuli and the responses in such a way that the latent dimensions "explain" independent components of the total inertia χ^2/N. The SVD yields the decomposition

$$X = Q\Lambda^{1/2}T', \quad \Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}) \qquad (22)$$

where Q is an $I \times s$ matrix of eigenvectors of XX', T is a $J \times s$-matrix of eigenvectors of $X'X$ and Λ is an $s \times s$-diagonal matrix containing the *singular values*, i.e. the roots $\sqrt{\lambda_k}$, $k = 1, \ldots, s$ of nonzero eigenvalues λ_k of XX' and $X'X$. The element q_{ik} in the i-th row and k-th column of Q is a score of stimulus pattern s_i on the k-latent dimension, and the element t_{kj} in the k-th row and j-th column of T is a score of the response r_j on the k-th latent dimension.

The re-scaling of the q_{ik} and t_{jk} acoording to

$$u_{ik} = q_{ik}\frac{\sqrt{\lambda_k}}{\sqrt{n_{i+}}}, \quad v_{jk} = t_{jk}\frac{\sqrt{\lambda_k}}{\sqrt{n_{+j}}} \qquad (23)$$

implies that the k-th axis "explains" that proportion of inertia that is generated by the k-th latent dimension in the same sense the coordinates in an ordinary PCA are chosen as to correspond to the proportion of variance explained by a principal component. Although $n_{i+} = n$ for all i we have written n_{i+} instead of n in order to stress the point that $\sqrt{\lambda_k}$ has to be divided by the row sums. u_{ik} corresponds to u_i and v_{jk} to v_j in (3).

An important notion for the interpretation of the results of a CA is that of a χ^2-*distance* between two row categories, i.e. stimuli in our case, or between two column categories, i.e. responses in our case; there is no such distance between a row and a column category, though, i.e. the concept of a distance between a stimulus and a response is not defined. Let s_i and $s_{i'}$ be two stimulus patterns. The χ^2-distance between the corresponding rows of the confusion matrix is defined by

$$\delta_{ii'} = \sum_{j=1}^{J} \frac{1}{n_{+j}} \left(\frac{n_{ij} - n_{i'j}}{n_{i+}} \right)^2, \tag{24}$$

since the stimuli are presented equally often $n_{i+} = n_{i'+}$ for all i and i' . The χ^2-distance between the responses is defined analogously, namely

$$\delta_{jj'} = \sum_{i=1}^{I} \frac{1}{n_{i+}} \left(\frac{n_{ij}}{n_{+j}} - \frac{n_{ij'}}{n_{+j'}} \right)^2. \tag{25}$$

Generally, $n_{+j} \neq n_{+j'}$.

Since CA provides a representation of the stimulus patterns s_i by points σ_i in some coordinate system it is possible to define the Euclidian[5] distance $d_{ii'}$ between two points σ_i and $\sigma_{i'}$ representing the patterns s_i and $s_{i'}$:

$$d(\sigma_i, \sigma_{i'}) = d_{ii'} = \sum_{k=1}^{s} (u_{ik} - u_{i'k})^2. \tag{26}$$

Analogously, the Eucledian distance between tow response points ρ_j and $\rho_{j'}$ is be defined by

$$d(\rho_j, \rho_{j'}) = d_{jj'} = \sum_{k=1}^{s} (v_{jk} - v_{j'k})^2. \tag{27}$$

It may be shown that the coordinates u_{ik} and v_{jk} have the following properties:

1. The dimensions represent independent latent variables, each of which accounts for a certain proportion of the total inertia χ^2/N.

2. Generally,

$$d_{ii'} = \delta_{ii'}, \quad d_{jj'} = \delta_{jj'}, \tag{28}$$

   i.e. the Eucledian distance between any two stimulus points σ_i and $\sigma_{i'}$ equals the corresponding χ^2-distance between the two points, and analogously the Eucledian distance between any two response points ρ_j and $\rho_{j'}$ equals the χ^2-distance between the two responses. In particular, the Eucledian distance $d(\sigma_i, 0)$ between the point σ_i, representing the stimulus s_i, and the origin 0 of the coordinate system reflects the contribution of s_i to the χ^2 of the confusion matrix,

3. The Eucledian distance between any row and column point cannot be related as directly to the χ^2 of the confusion matrix. However, the x_{ij}-values depend upon the scalarproduct of the coordinates u_{ik} and v_{jk}, is maximised if $u_{ik} = v_{jk}$ for all

[5] or Pythagorean

k, i.e. if the points σ_i and ρ_j coincide. Then x_{ij} assumes a maximum value. It follows that the nearer the points σ_i and ρ_j, the greater the value of x_{ij} will be, i.e. close points σ_i and ρ_j signal a high confusion probability for s_i with s_j (where, however, "nearer" should not be related to distance; the distance between a "row" (= stimulus) and a "column" (= response) point ist not defined!). If the subjects learn to identify the stimuli correctly we would therefore expect in particular the points σ_i and ρ_i to be close each other.

A comparison with (3) shows that dual scaling will, under the specified conditions, even yield estimates of the $E(x|s_i)$. Let $E(x|s_i) = \mu_i$ and suppose further that $\mu_i = \alpha\phi_i$, ϕ_i a measure of the physical component with respect to which the stimuli differ, and α a proportionality constant. For instance, let the patterns be Gabor patches differing only with respect to the value of the carrier frequency f, then $\phi_i = f_i$ for the i-the pattern. If the distribution is Gaussian and the value of σ is of the right order, we should then expect a linear relation between the u_i and the f_i. This is precisely what we observed.

So dual scaling or correspondence analysis can, under suitable conditions, offer more than just a graphical summary of the data in a confusion matrix. There is good reason to assume that the conditions under which (11) holds are met also in the experiment considered in this paper, since the differences between the radii of the discs are rather small. However, the stimulus patterns differ with respect to more than a single physical dimension since they are composed of two discs. Since CA may be considered a generalised version of dual scaling, automatically catering for the case of more than a single scale with respect to which the stimuli and the responses have to be characterised, CA will indicate to what extend a second dimension will be required to cater for the data. The question is which stimulus dimensions the latent dimensions provided by CA actually represent.

Let us assume that $x = (x_1, x_2)'$, i.e. that the sensory representation of the patterns can be described by two stimulus dimensions. In the simplest case, x_1 will represent the size (radius) of one disc, and x_2 the size of the other. Since the pattern and therefore the combination of the components (the discs) follow each other in an experimental session independent of each other, the x_1 and x_2 may also vary independent of each other. The sizes of the greater (the "outer") disc will be easier to differentiate than the sizes of the smaller (the "inner") disc, so we may expect that the first latent dimensions represents x_1, accounting for the larger proportion of the inertia χ^2/N of the matrix, and the second will represent x_2.

This need not be so, however. x_1 and x_2 may indeed be the representations of the individual discs, but the subject may judged according to the value of certain functions $y(x_1, x_2)$ of x_k, $k = 1, 2$. For instance, the subject may respond according to the value of $y(x_1, x_2) = x_1/x_2$. This is not intended to mean that the subject actually computes these quotients, there may exist some neural mechanism with an output representable by x_1/x_2. Another aspect may be the impression of brightness of the pattern. The brightness will be a function of the radius of the "inner" disc; yet another variable upon which the decisions may be based is the energy of the stimuli.

There are many functions of the immediate sensory representation upon which the decisions depend. To demand that methods like dual scaling or CA provide an immediate access to the variables subject employ in order to arrive at their decisions would be asking too much. However, as long as the u_i and the $E(y|s_i)$ can assumed to be proportional to each other, where x has been replaced by y in order to indicate that the scale values may depend upon some function of the immediate sensory representation, one can test a number of hypotheses about the y.

References

[1] Ashby, G.F. & Townsend, J.T. Varieties of perceptual independence. *Psychological Review*, 1986, 93, 154-179

[2] Ashby, F.G. & Gott, R.E. Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1988, 14, 33-53

[3] Bauer R, Dicke P (1997) Fast cortical selection: a principle of neuronal self-organization for perception? Biol Cybern 77:207-215

[4] Beard B.L., Levi D.M., Reich L.N. (1995) Perceptual learning in Parafoveal Vision. Vision Res 35:1679-1690

[5] Bienenstock EL, Cooper LN, Munro PW (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J. of Neuroscience 2: 32-48

[6] Bishop, C.M., Neural networks for pattern recognition. Clarendon Press, Oxford 1995

[7] Caelli, T., Rentschler, I., Scheidler, W. (1987) Visual pattern recognition in humans. I. Evidence for adaptive filtering. Biol. Cybernetics, 57, 233-240

[8] Fahle, M., Edelman, S., Poggio, T.: Fast perceptual learning in visual hyperacuity. Vision Res., 35, 3003 - 3013

[9] Falmagne, J. C. Biscalability of error matrices and all-or-none Reaction time theories. Jounal of Mathematical Psychology, 1872, 9, 206 - 224

[10] Garner, W.R. & Morton, J. Perceptual independence: definitions, models and experimental paradigms. *Psychological Bulletin*, 1969, 72, 233-259

[11] Greenacre, M.J. Theory and applications of correspondence analysis. *Academic Press*, London, 1983

[12] Hauske, G., Wolf, W. & Lupp, U. Matched filters in human vision. *Biological Cybernetics*, 1976, 22, 181-188

[13] Herzog MH, Fahle M (1998) Modeling perceptual learning: difficulties and how they can be overcome. Biol. Cybern. 78:107-117

[14] Jüttner, M., Caelli, T. Rentschler, I. (1997) Evidence-based pattren classification: a strucutal approach to human perceptual learning and generalization. Journal of Mathematical Psychology, 41, 144-259

[15] King-Smith PE, Kulikowski JJ (1975) The detection of gratings by independent activation of line detectors. J Physiol 247:237-271

[16] Luce, R.D., Detection and recognition. In: Luce, R.D., Bush, R.R. and Galanter, E. (eds.) Handbook of Mathematical Psychology, Vol. 1, Wiley, New York 1963

[17] Meinhardt, G. & Mortensen, U., Detection of aperiodic test patterns by pattern specific detectors revealed by subthreshold summation. *Biological Cybernetics*, 1998, 79, 413-425

[18] Morrone, M.C., Owens, R.A. (1987) Feature detection from local enenrgy. Pattern Recognition Letters, 6, 303 - 313

[19] Mortensen, U. & Nachtigall, C. Visual channels, Hebbian assemblies and the effect of Hebb's rule. *Submitted for publication*

[20] Nosofsky, R.M. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 1986, 115, 39-57

[21] Oja E (1982) A simplified neuron model as a pricipal component analyzer. J. Math. Biol. 15: 267-273

[22] Papoulis, A. Systems and transforms with applications in optics. *Robert F. Krieger Publishing Company, Malabar*, 1981

[23] Rentschler, I., Jüttner, M., Caelli, T. (1993) Probabilistic analysis of human supervised learning and classification. Vision Research, ??, 669-687

[24] Rentschler, I., Barth, E., Caelli, T., Zetsche, C., Jüttner, M.. (1996) Generalization of form in visual pattern classification. Spatial Vision, 10, 59-85

[25] Ripley, B. D., Pattern recognition and neural networks. *Cambridge University Press*, 1996

[26] Thomas, J.P., Detection and identification: how are they related? *Journal of the Optical Society of America, A*, 1985, 9, 1457-1467

[27] Townsend, J.T., Theoretical analysis of an alphabetic confusion matrix. Perception and Psychoph., 1971 (a), 9, 40 - 50

[28] Townsend, J.T., Alphabetic confusion: A test of models for individuals. Perception and Psychophys., 1971 (b), 9, 449-454

[29] Townsend, J.T., & Ashby, G.F. Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, 8, 834-864

[30] Townsend, J.T. & Landon, D.E. Mathematical models of recognition and confusion in psychology. *Mathematical Social Sciences*,1983, 4, 25-71

[31] Townsend, J.T., Hu, G.G., & Ashby, F.G. A test of visual feature sampling independence with orthogonal straight lines. *Bulletin of the Psychonomic Society*, 1980, 15, 163-166

Index