A STOCHASTIC MODEL FOR PAIRED COMPARISONS OF SOCIAL STIMULI

by

Axel Mattenklott
Psychologisches Institut, University of Mainz

Klaus-J. Miescke*
Department of Statistics, Purdue University

Joachim Sehr
Fachbereich Rechts-und Wirtschaftswissenschaften
University of Mainz

Technical Report #82-20

Department of Statistics
Purdue University

July 1982

*The research of the second mentioned author was supported partly by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University.

A Stochastic Model for Paired Comparisons of Social Stimuli

by

Axel Mattenklott
Psychologisches Institut, University of Mainz

Klaus-J. Miescke*
Department of Statistics, Purdue University

Joachim Sehr
Fachbereich Rechts-und Wirtschaftswissenschaften
University of Mainz

Summary

A stochastic model for paired comparisons of multi-attribute social stimuli is proposed where one objective is to find the relative importance of the attributes for a judge. The model can be conceived as a special strict binary utility model, i.e. a BTL-model, and is related to the factorial model of Abelson and Bradley. The scale values of the stimuli are linear combinations of functions of the stimuli's attributes. The model does neither assume that the functions are fixed in advance nor that different judges have the same set of functions. The choice among such functions, however, is admitted only within a finite scope. Within the framework of exponential families, maximum likelihood estimators and tests are derived and applied to data coming from two psychological experiments.

Key words: Paired comparisons; social stimuli; importance of attributes; exponential families; maximum likelihood estimators.

*The research of the second mentioned author was supported partly by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University.

1. Introduction

Almost all models of multi-attribute decision and judgment research consider the (subjective) utility (judgment, resp.) $U(G^{(i)})$ of an object (alternative, resp.) $G^{(i)}$ as a function of values $\mathcal{M}_k^{(i)}$, where $\mathcal{M}_k^{(i)}$ is the aspect of the attribute $\mathcal{M}_k$ for $G^{(i)}$, $k=1,\ldots,K$, and $G^{(i)}$ is characterized by the $\mathcal{M}_k^{(i)}$ with fixed $i \in \{1,\ldots,N\}$:

$$U(G^{(i)}) = F[a_1 f_1(\mathcal{M}_1^{(i)}),\ldots,a_K f_K(\mathcal{M}_K^{(i)})]. \tag{1}$$

Here, $f_1,\ldots,f_K$ are the functions which represent, cognitively, the K attributes of objects $G^{(1)},\ldots,G^{(N)}$, and $a_1,\ldots,a_K$ are the weights which indicate the importance of the attributes $\mathcal{M}_1,\ldots,\mathcal{M}_K$. The attributes jointly describe all objects by different aspects. F is a real valued monotone function in each of its arguments and specifies the composite effect of the attributes.

If the $f_k(\mathcal{M}_k^{(i)})$ have scale values already, e.g. the test scores of intelligence, experience, or neuroticism of an applicant, and if the judgment U gets a value on a (subjective) scale, e.g. the degree of the applicant's qualification, weights a_k's can be simply estimated from models, e.g. from the model of multiple linear regression. Clearly, in psychology these models frequently are formulated stochastically. In these cases an error random variable should be added to the right side of the equation (1).

In many situations, the decision between two objects $G^{(i)}$ and $G^{(j)}$ at a time, which one of them is more attractive, more qualified, or rather more threatened, is psychologically more meaningful and less difficult for the judge than to assign a number on a rating scale. Bradley (1976) has given an excellent overview over the stochastic methods in the area of paired comparisons. The models being discussed in this framework can be considered as variants of

$$p_{ij}: = \mathrm{pr}(G^{(i)} \text{ is preferred to } G^{(j)}) = \psi[U(G^{(i)}),U(G^{(j)})],$$

where U is the (subjective) utility and ψ is a combination $\varphi \circ d$ of a sign dependent distance d and a function φ from $\mathbb{R}$ to [0,1] which has the properties of a cumulative distribution function. In Bradley's (1976) paper, $U(G^{(i)})$ is most frequently represented by a single parameter π_i, while representation by a function F is given for factorial designs. In factorial designs, the (subjective) utility of an object is decomposed into contributions from the respective attributes.

In pairwise comparisons, the processing of the objects is assumed to follow mainly one of two broad strategies. According to the first strategy, each object $G^{(i)}$ is evaluated separately and independently from all other objects. It then gets a (subjective) overall value $U(G^{(i)})$ of attractiveness, for instance, $G^{(i)}$ then is judged to be more attractive than $G^{(j)}$, whenever $U(G^{(i)}) > U(G^{(j)})$. According to the second strategy, the objects are compared attribute-wise. In comparing $G^{(i)}$ and $G^{(j)}$, at first each attribute $\mathcal{M}_k$ gets a value that indicates whether, and how much, $G^{(i)}$ is more or less attractive than $G^{(j)}$ with respect to this individual attribute. $G^{(i)}$ then is preferred to $G^{(j)}$ if the overall value of all attribute-wise comparisons is assumed to favour object $G^{(i)}$.

If it is further assumed that, in the first strategy, the overall value $U(G^{(i)})$ is formed by summing up its corresponding attribute values or, in the second strategy, by summing up the attribute-wise comparisons, then the processing of the objects for a decision can be represented by an additive difference model (Tversky, 1969):

$$G^{(i)} \succsim G^{(j)} \text{ iff } \sum_{k=1}^{K} \phi_k[f_k(\mathcal{M}_k^{(i)})-f_k(\mathcal{M}_k^{(j)})] \geq 0,$$

where $\phi_k(-\delta) = -\phi_k(\delta)$, and ϕ_k as well as f_k are real valued functions, k=1,...,K.

Clearly, the additive difference model represents the second strategy. The first strategy, if represented by an additive model, is a special case of the additive difference model where all the difference functions have the form $\phi_k(\delta) = a_k\delta$. In this special case, the two models coincide, but not necessarily the two processing strategies.

If the $f_k(\mathfrak{m}_k^{(i)})$ have scale values while the judgments result from paired comparisons, Srinivasan and Shocker (1973) have proposed an estimation procedure for the weights $a_1,\ldots,a_K$ from a system of linear inequalities. This procedure is related to the approach of Krantz et al. (1971, chap. 9) for whom the main problem is to find necessary and sufficient conditions for the existence of a specific form F that indicates the specific integration of the attributes, if only a finite set of data coming from paired comparisons is available. The first answer is clear: It is necessary and sufficient that scale functions $a_1f_1,\ldots,a_Kf_K$ exist such that $U(G^{(i)}) > U(G^{(j)})$ implies that $G^{(i)}$ is preferred to $G^{(j)}$ for all $i < j$. Due to the stochastic nature of many evaluation processes, fully consistent behavior of the judges cannot be expected. Nevertheless it seems to make sense to estimate the scale functions or parts of them, i.e. the $a_1,\ldots,a_K$ so that only as few as possible of the implications above are wrong, especially if a particular form of F or some special forms of F and f_k, $k=1,\ldots,K$, are assumed to be valid.

However, if judgments are made with respect to paired comparisons of social objects, as to the attractiveness between two particular persons or holiday places, respectively, models for the estimation of importances $a_1,\ldots,a_K$ should satisfy the following two conditions.

[1] <u>It cannot be assumed that the attributes of the objects have already fixed functions</u> $f_1,\ldots,f_K$ <u>being available for the judge</u>.

Under this condition, the approach of Srinivasan and Shocker (1973) is not appropriate. Basically, the attributes could have got such functions from the judges, as it is assumed in several approaches (e.g. Gensch & Recker, 1979). However, this procedure requires two conflicting judgment situations within the same judge:

(a) the decision between two objects (stimuli) $G^{(i)}$ and $G^{(j)}$ at a time, for instance, two holiday places, and

(b) the decision for scale values for each aspect of each respective attribute, e.g. the attractiveness of the hotels.

It is intuitively clear that these two judgment situations will mutually influence each other and thus will inflate the validity of the model.

[2] It cannot be assumed that different judges decide on the basis of the same set of functions $f_1,\ldots,f_K$ for the attributes $\mathcal{M}_1,\ldots,\mathcal{M}_K$ of the objects.

The aspects within the attributes often are preferred quite differently by different judges. For instance, one sea-side resort will be preferred by one judge due to its much frequented beaches while the other resort will be preferred by another judge due to its tranquil beaches. Averaging over judges would lead to results which represent none of the judges. Thus an analysis of the judgments of the individual judge should be performed.

The objective of the present paper is to propose a stochastic model for the decision process as well as for the estimation of the weights a_k of the attributes $\mathcal{M}_k$, k=1,...,K. Its applicability will be demonstrated with paired comparisons in two psychological experiments. The final form of the model satisfies both conditions [1] and [2].

In the second section, normal equations for the maximum likelihood estimators, and a necessary and sufficient condition for their existence, will be derived as well as tests for hypotheses with respect to the a_k's. In the third section, the results of two experiments will be analysed by the proposed procedures. The results corroborate the importance of condition [2].

2. The Stochastic Model

2.1. Presentation of the Model in its Preliminary Form and Likelihood-Equations

Let $G^{(1)},\ldots,G^{(N)}$ be N stimuli (objects) which can be described by K attributes $\mathcal{M}_k \in \{1,2,\ldots,m_k\}$ with fixed m_k, $k=1,\ldots,K$. Thus they may be represented by

$$G^{(n)} = (\mathcal{M}_1^{(n)},\ldots,\mathcal{M}_K^{(n)}),\ n = 1,\ldots,N,\ N = \prod_{k=1}^{K} m_k. \qquad (2)$$

Each of the N(N-1)/2 pairs of stimuli are presented in a random succession to a subject (judge) who has to decide in each case which one of the two stimuli is preferred to the other one. These preferences are assumed to be given independently. Thus we have a full design in a double sense: All objects which can be constructed from the attributes and all pairs of objects are at hand.

Let us further assume that a judge prefers object $G^{(i)}$ to $G^{(j)}$ with probability

$$\mathrm{pr}(G^{(i)} \text{ is preferred to } G^{(j)}) := p_{ij}(\underline{a}) :=$$

$$:= (1+\exp\{-(\sum_{k=1}^{K} a_k M_k^{(i)} - \sum_{k=1}^{K} a_k M_k^{(j)})\})^{-1}, \qquad (3)$$

$1 \leq i < j \leq N$, where $\underline{a} = (a_1,\ldots,a_K)$ are K unknown real parameters, $M_k^{(i)} = f_k(\mathcal{M}_k^{(i)})$, $i = 1,\ldots,N$, $k = 1,\ldots,K$, and $f_1,\ldots,f_K$ are (for a while) fixed known real valued functions.

It is easy to see that

$$p_{ij}(\underline{a}) = \frac{\exp\{\sum_{k=1}^{K} a_k[M_k^{(i)}-M_k^{(j)}]\}}{1+\exp\{\sum_{k=1}^{K} a_k[M_k^{(i)}-M_k^{(j)}]\}} \tag{3'}$$

$$= \frac{\exp\{\sum_{k=1}^{K} a_k M_k^{(i)}\}}{\exp\{\sum_{k=1}^{K} a_k M_k^{(i)}\}+\exp\{\sum_{k=1}^{K} a_k M_k^{(j)}\}}$$

$$= \frac{\prod_{k=1}^{K} \exp\{a_k M_k^{(i)}\}}{\prod_{k=1}^{K} \exp\{a_k M_k^{(i)}\}+\prod_{k=1}^{K} \exp\{a_k M_k^{(j)}\}},$$

and that

$$\text{logit } p_{ij}(\underline{a}) := \log \frac{p_{ij}(\underline{a})}{1-p_{ij}(\underline{a})} = \sum_{k=1}^{K} a_k[M_k^{(i)}-M_k^{(j)}], \quad 1 \leq i < j \leq N.$$

Thus the model may be considered as a strict binary utility model (Luce & Suppes, 1965), as a special version of the Bradley-Terry-model (1952) (cf. Zermelo, 1929), or as a particular regression model (cf. Grizzle, Starmer & Koch, 1969). If the $f_1,\ldots,f_K$ can vary only with respect to the constraint

$$\sum_{\{\mathcal{m}_k^{(i)}|1\leq i\leq N\}} \exp\{a_k f_k(\mathcal{m}_k^{(i)})\} = 1 \quad \left(\text{or} \sum_{\{\mathcal{m}_k^{(i)}|1\leq i\leq N\}} f_k(\mathcal{m}_k^{(i)}) = 0\right), \quad k = 1,\ldots,K,$$

we get the factorial model of Abelson and Bradley (1954)(cf. Bradley & El-Helbawy 1976). In this model, we can fix the $a_1,\ldots,a_K$ to nonzero real numbers (e.g. $a_k = 1$, $k = 1,\ldots,K$) or put other constraints on $f_1,\ldots,f_K$.

Let the outcomes be random variables $X_{ij} \in \{0,1\}$, where $X_{ij} = 1(0)$ has the meaning that $G^{(i)}$ is preferrred to $G^{(j)}$ ($G^{(j)}$ is preferred to $G^{(i)}$), $1 \leq i < j \leq N$. Thus the likelihood function is given by

$$\mathrm{pr}(X_{ij}=x_{ij}, 1\leq i<j\leq N) = \prod_{i<j} p_{ij}(\underline{a})^{x_{ij}}(1-p_{ij}(\underline{a}))^{1-x_{ij}}$$

$$= \frac{\exp\{\sum_{k=1}^{K} a_k \sum_{i<j} x_{ij}[M_k^{(i)}-M_k^{(j)}]\}}{\prod_{i<j} (1+\exp\{\sum_{k=1}^{K} a_k[M_k^{(i)}-M_k^{(j)}]\})} ,$$

$$x_{ij} \in \{0,1\},\ 1 \leq i < j \leq N. \qquad (4)$$

Since the likelihood function is a continuous function from $\mathbb{R}^K$ to [0,1], it has a global maximum, at least asymptotically at $a_k = +\infty$ or $a_k = -\infty$ for some of the a_k, $k = 1,\ldots,K$. Moreover, since the likelihood function is differentiable, the partial derivations have to be zero for a maximum likelihood estimator from $\mathbb{R}^K$.

It can be seen immediately that the following random vector $\underline{Z}$ is a sufficient statistic for $\underline{a} \in \mathbb{R}^K$:

$$\underline{Z} = (Z_1,\ldots,Z_K) ,$$
$$\text{where } Z_k = \sum_{i<j} X_{ij}[M_k^{(i)}-M_k^{(j)}],\ k = 1,\ldots,K . \qquad (5)$$

Moreover, the distributions of $\underline{Z}$ constitute a K-parametric exponential family, since for every fixed $\underline{a} \in \mathbb{R}^K$

$$\mathrm{pr}(\underline{Z}=\underline{z}) = \exp\{\sum_{k=1}^{K} a_k z_k + c(\underline{a}) + d(\underline{z})\},\ \underline{z} = (z_1,\ldots,z_K) \in \mathcal{Z} , \qquad (6)$$

where $\mathcal{Z}$, $c(\underline{a})$ and $d(\underline{z})$ depend on $f_1,\ldots,f_K$ in an obvious manner.

The parametrization is identifiable if and only if for every $k \in \{1,\ldots,K\}$, $m_k \geq 2$ and the values $f_k(1),\ldots,f_k(m_k)$ are not all identical. Under this condition, an unique maximum likelihood estimator $\hat{\underline{a}} := (\hat{a}_1,\ldots,\hat{a}_K) \in \mathbb{R}^K$ for $\underline{a}$ exists if the following system of equations has a solution:

$$\sum_{i<j} [x_{ij}-p_{ij}(\underline{a})][M_k^{(i)}-M_k^{(j)}] = 0,\ k = 1,\ldots,K, \qquad (7)$$

which then is the unique solution of (7) and equals to $\hat{\underline{a}}$.

System (7) constitutes a system of normal equations which result from putting all partial derivations of the likelihood function (4) equal to zero and some additional standard analysis. That (7) in fact leads to unique maximum likelihood estimators can be seen more quickly by realizing that $E_{\hat{\underline{a}}(\underline{z})}(Z_k) = \sum_{i<j} p_{ij}(\underline{a})[M_k^{(i)}-M_k^{(j)}]$, $k = 1,\ldots,K$, and thus a solution $\hat{\underline{a}}(\underline{z})$ of (7) satisfies the canonical conditions for exponential families:

$$E_{\hat{\underline{a}}(\underline{z})}(Z_k) = z_k,\ k = 1,\ldots,K, \tag{8}$$

(cf. Bickel & Doksum, 1977, p. 106).

2.2. Existence of Real Maximum Likelihood Estimators

Now, it may happen that a maximum likelihood estimator $\hat{\underline{a}} \in \mathbb{R}^K$ does not exist. For example, consider the case $x_{ij} = 1(0)$ if $M_1^{(i)} > (<)\ M_1^{(j)}$ for all $i < j$. In this instance, a judge decides (lexicographically) with respect to the first attribute only, provided that $M_1^{(i)} \neq M_1^{(j)}$. The likelihood function, then, increases with a_1 for any fixed $a_2,\ldots,a_K \in \mathbb{R}$. For a fixed pair (i,j) with $i < j$, three cases can be distinguished.

(a) $M_1^{(i)} > M_1^{(j)}$: The associated factor in the likelihood function is $p_{ij}(\underline{a})$, which is increasing in a_1,

(b) $M_1^{(i)} < M_1^{(j)}$: The factor is $1 - p_{ij}(\underline{a})$, which is increasing in a_1, and

(c) $M_1^{(i)} = M_1^{(j)}$: The factor is constant in a_1.

Thus, if we want to speak about a maximum likelihood estimator in this case, the first component has to be ∞ .

To handle this problem in general, let us state the following.

Definition. A set $\{k_1,\ldots,k_L\} \subseteq \{1,\ldots,K\}$ is called a determining set (of indices of attributes), if there are nonzero real numbers $\alpha_1,\ldots,\alpha_L$, such that for all $i < j$,

$$\sum_{\ell=1}^{L} \alpha_\ell M_{k_\ell}^{(i)} > (<) \sum_{\ell=1}^{L} \alpha_\ell M_{k_\ell}^{(j)} \quad \text{implies} \quad x_{ij} = 1(0).$$

A determining set is called minimal if no proper determining subset exists, and it is called maximal, if it is not a proper subset of a determining set.

Let us now state three properties of those determining sets:

First, if a determining set exists then there is a unique minimal and a unique maximal one. The existence is obvious. To demonstrate the uniqueness, let us consider two different minimal determining sets, say, $\mathcal{K} = \{k_1,\dots,k_L\}$ and $\tilde{\mathcal{K}} = \{\tilde{k}_1,\dots,\tilde{k}_{\tilde{L}}\}$. If $\mathcal{K} \subset \tilde{\mathcal{K}}$ or $\tilde{\mathcal{K}} \subset \mathcal{K}$, minimality is contradicted. If $\mathcal{K} \setminus \tilde{\mathcal{K}} \neq \emptyset$ and $\tilde{\mathcal{K}} \setminus \mathcal{K} \neq \emptyset$, a pair (i,j) of objects can be found (full design) with $M_k^{(i)} = M_k^{(j)}$ for $k \in \mathcal{K} \cap \tilde{\mathcal{K}}$, and with $\sum_{\ell=1}^{L} \alpha_\ell [M_{k_\ell}^{(i)} - M_{k_\ell}^{(j)}] > 0$ and $\sum_{\ell=1}^{\tilde{L}} \tilde{\alpha}_\ell [M_{\tilde{k}_\ell}^{(i)} - M_{\tilde{k}_\ell}^{(j)}] < 0$, so that $x_{ij} = 1$ and $x_{ij} = 0$, a contradiction. The proof for maximal sets is the same.

Secondly, there is a (unique) hierarchy of those minimal determining sets. For if $\mathcal{K}_1 = \{k_1^{(1)},\dots,k_{L_1}^{(1)}\}$ is the unique minimal determining set, let us eliminate all pairs (i,j) with $\sum_{\ell=1}^{L_1} \alpha_\ell^{(1)} [M_{k_\ell^{(1)}}^{(i)} - M_{k_\ell^{(1)}}^{(j)}] \neq 0$. The remaining pairs include, what can be considered as a full design from the attributes not from $\mathcal{K}_1$ with $\prod_{k \in \mathcal{K}_1} m_k$ observations per pair: the set of all pairs (i,j) with $M_k^{(i)} = M_k^{(j)}$ for every $k \in \mathcal{K}_1$. Clearly, again a unique minimal determining set $\mathcal{K}_2 = \{k_1^{(2)},\dots,k_{L_2}^{(2)}\}$ may exist with $\mathcal{K}_2 \cap \mathcal{K}_1 \neq \emptyset$ and

$$\sum_{\ell=1}^{L_1} \alpha_\ell^{(1)} [M_{k_\ell^{(1)}}^{(i)} - M_{k_\ell^{(1)}}^{(j)}] = 0 \text{ and } \sum_{\ell=1}^{L_2} \alpha_\ell^{(2)} [M_{k_\ell^{(2)}}^{(i)} - M_{k_\ell^{(2)}}^{(j)}] > (<) \ 0$$

imply $x_{ij} = 1(0)$.

This process may be continued until the last minimal determining set $\mathcal{K}_p$ has been found.

Thirdly, $\{1,2,\dots,K\} \setminus \bigcup_{r=1}^{p} \mathcal{K}_r$ is the non determining remainder, denoted stochastic remainder, and $\bigcup_{r=1}^{p} \mathcal{K}_r$ is the unique maximal determining set.

Both sets may be empty.

Let $\bigcup_{r=1}^{p} \mathcal{K}_r = \{\gamma_1,\gamma_2,\ldots,\gamma_q\}$ so that $\mathcal{K}_1 = \{\gamma_1,\ldots,\gamma_{L_1}\}$, $\mathcal{K}_2 = \{\gamma_{L_1+1},\ldots,\gamma_{L_1+L_2}\}$ and so on. It is possible to select $\alpha_1,\ldots,\alpha_{L_1}$, $\alpha_{L_1+1},\ldots,\alpha_{L_1+L_2},\ldots,\alpha_q$, where the α_ℓ from the next following minimal determining set are smaller than those from the preceding one and where $\sum_{\ell=1}^{q} \alpha_\ell[M^{(i)}_{\gamma_\ell}-M^{(j)}_{\gamma_\ell}] > (<)\ 0$ implies $x_{ij} = 1(0)$. It is maximal since, otherwise, a further minimal determining set would have been found.

__The Main Result__. __The maximum likelihood estimators__ $\hat{a}_1,\ldots,\hat{a}_K$ __exist (as real numbers) if and only if there is no determining set__.

__Proof.__ If there is a determining set, then there is a minimal one, say, $\mathcal{K} = \{k_1,\ldots,k_L\}$ with $\alpha_1,\ldots,\alpha_L$. The part of the likelihood function (4) associated with pairs (i,j) with i < j and (*): $\sum_{\ell=1}^{L} \alpha_\ell[M^{(i)}_{k_\ell}-M^{(j)}_{k_\ell}] \neq 0$, can get close to 1, but only if the $|a_{k_\ell}|$ are sufficiently large, $\ell = 1,\ldots,L$. For instance, let $a_{k_\ell} = \beta\alpha_\ell$, $\ell = 1,\ldots,L$, and a_k, $k \notin \mathcal{K}$, be arbitrary real numbers. Then, $\prod_{i<j(*)} p_{ij}(\underline{a})^{x_{ij}}(1-p_{ij}(\underline{a}))^{1-x_{ij}} \xrightarrow{\beta\to\infty} 1$, and the product is unequal to 1 for every $\underline{a} \in \mathbb{R}^K$. The remainder of the likelihood function is not affected by a change of β.

To prove the opposite direction, let us begin with the assumption of non-existence of $\hat{\underline{a}} \in \mathbb{R}^K$. Let $(a_1^{(n)},\ldots,a_L^{(n)},\hat{a}_{L+1},\ldots,\hat{a}_K)$, n = 1,2,..., with $|a_\ell^{(n)}| \xrightarrow{n\to\infty} \infty$, $\ell = 1,\ldots,L$, be a sequence which tends to the location of the maximum, which exists at least in $\bar{\mathbb{R}}^K$. In this case for i < j,

$$(**):\ \lim_{n\to\infty} \sum_{\ell=1}^{L} a_\ell^{(n)}[M_\ell^{(i)}-M_\ell^{(j)}] = \infty\ (-\infty) \text{ implies } x_{ij} = 1(0),$$

since, otherwise, one factor of the likelihood function tends to zero.

There exists a $k \in \{1,\dots,L\}$ such that $\lim_{n\to\infty} a_\ell^{(n)}/|a_k^{(n)}| = \alpha_\ell \in \mathbb{R}$, say, $\ell = 1,\dots,L$. Then the first equation in (**) is implied by $\sum_{\ell=1}^{L} \alpha_\ell [M_\ell^{(i)} - M_\ell^{(j)}] > (<)\ 0$, i.e., the subset $\mathcal{K} = \{\ell | \alpha_\ell \neq 0, \ell=1,\dots,L\}$ is a determining set, and therefore the proof is completed.

If there is a determining set, then there is a maximal one: $\{\gamma_1,\dots,\gamma_q\}$, say. In this case the a_k, $k \notin \{\gamma_1,\dots,\gamma_q\}$, may be estimated from the paired comparisons at those (i,j) with $\sum_{\ell=1}^{q} \alpha_\ell [M_{\gamma_\ell}^{(i)} - M_{\gamma_\ell}^{(j)}] = 0$, i.e., in a reduced model. The $\hat{a}_k$, $k \notin \{\gamma_1,\dots,\gamma_q\}$, which are found in this way, together with $\alpha_\ell \cdot \infty$, $\ell \in \{\gamma_1,\dots,\gamma_q\}$, may be interpreted as a solution of the problem where α_ℓ, $\ell \in \{\gamma_1,\dots,\gamma_q\}$ are taken from the maximal determining set. In this way, the model is also suitable if a judge at first decides lexicographically with respect to certain attributes, and only in cases where no conclusion can be made in this way, decides in a non-deterministic manner.

2.3. Remarks.

(a) Obviously, the form of a solution of system (7) depends heavily on the form of the given $f_1,\dots,f_K$. It cannot be given explicitly in general but has to be evaluated in every concrete situation numerically.

(b) If not all N(N-1)/2 pairs of stimuli but only a fixed subset of them are presented to the subject and judged by her or him similar results can be derived. The only thing that changes in the formulas is that all sums and products in (4), (5), and (7) have to be restricted to this subset. However, the identifiability of the parametrization as well as the results concerning the existence of $\hat{\underline{a}} \in \mathbb{R}^K$ depend now additionaly on this subset.

(c) If the N(N-1)/2 judgments are collected independently from R subjects who can be assumed to judge according to (3) with a common $\underline{a}$, or are collected from one subject who is assumed to give her of his judgments independently R times, respectively, our results derived so far hold analogously. One only has to replace the X_{ij}'s and x_{ij}'s throughout by the arithmetic means of the corresponding responses and to pay attention to the fact that then (4) is no longer the likelihood function, but now its R-th root. Especially, the model can be examined with generalized (maximum) likelihood ratio tests, in two directions, which can be established by use of the well known χ^2-approximation. The first tests the hypothesis $\underline{H}$: "$a_1=\ldots=a_K=0$" (pure randomness) versus $\underline{K}$: "model (3)", and significance supports our model. The second may test the hypothesis $\underline{H}'$: "model (3)" versus $\underline{K}'$: "Bradley-Terry model (1952)", and non-significance then supports our model. Of course, one should clearly distinguish between the meaning of "supports" in the two tests due to the well known unsymmetry in the theory of testing hypotheses. It should be pointed out also that in the case of R = 1, both tests are not at hand and even the parameters in the Bradley-Terry model (1952) are then mostly not estimable within the maximum likelihood approach.

2.4. The Final Form of the Model

Instead of proceeding along the lines as indicated in Remark (c), we prefer to generalize our model, admitting a more individual behavior of the single subject. By this we do not only mean that parameters $\underline{a}$ may now differ from subject to subject but also that, instead of having $\underline{f} := (f_1,\ldots,f_K)$ fixed known, we assume now that only $f_k \in \mathfrak{F}_k$ holds, where $\mathfrak{F}_k$ is a given finite set of real-valued functions, $k = 1,\ldots,K$. Maximum likelihood estimates $\hat{\underline{f}} := (\hat{f}_1,\ldots,\hat{f}_K)$ and $\hat{\underline{a}}$ for a single person are determined now in the following way: For every fixed $\underline{f}$ with $f_k \in \mathfrak{F}_k$, $k = 1,\ldots,K$, a solution of (7) has to be found and, together with $\underline{f}$, to be inserted into the likelihood function (4). The largest value of (4) then determines the maximum likelihood estimates $\hat{\underline{f}}$ and $\hat{\underline{a}}$.

Now let us assume that there is a determining set $\mathcal{K}$ with respect to a special choice $f_k \in \mathfrak{F}_k$, $k = 1,\dots,K$. To give meaningful maximum likelihood estimates in this case we propose the following way. Let us define the maximal determining set $\mathcal{K}_{max} = \{\gamma_1,\dots,\gamma_q\}$ in this case as to be the largest possible determining set under all possible choices of $f_k \in \mathfrak{F}_k$, $k = 1,\dots,K$. Before we prove the existence and uniqueness of $\mathcal{K}_{max}$, which requires a comparison of all maximal determining sets associated with $\underline{f} = (f_1,\dots,f_K)$, $f_k \in \mathfrak{F}_k$, $k = 1,\dots,K$, it should be pointed out (as will be demonstrated later in Section 3) that in practice this work needs not to be done for all sets of data.

To prove that $\mathcal{K}_{max}$ exists uniquely, let us consider two different candidates $\mathcal{K}_{max} = \{\gamma_1,\dots,\gamma_q\}$ and $\tilde{\mathcal{K}}_{max} = \{\tilde{\gamma}_1,\dots,\tilde{\gamma}_{\tilde{q}}\}$, say. Let us assume that $\mathcal{K}_{max}(\tilde{\mathcal{K}}_{max})$ belongs to a choice $g_k(\tilde{g}_k) \in \mathfrak{F}_k$, $k = 1,\dots,K$. If $\mathcal{K}_{max} \subset \tilde{\mathcal{K}}_{max}$ or $\tilde{\mathcal{K}}_{max} \subset \mathcal{K}_{max}$, we can discard the smaller one. If $\mathcal{K}_{max} \backslash \tilde{\mathcal{K}}_{max} \neq \emptyset$ and $\tilde{\mathcal{K}}_{max} \backslash \mathcal{K}_{max} \neq \emptyset$, we can find a pair (i,j) of objects (full design) with $\mathfrak{m}_k^{(i)} = \mathfrak{m}_k^{(j)}$ for $k \in \mathcal{K}_{max} \cap \tilde{\mathcal{K}}_{max}$, and with

$$\sum_{p=1}^{q} \alpha_p[g_{\gamma_p}(\mathfrak{m}_{\gamma_p}^{(i)}) - g_{\gamma_p}(\mathfrak{m}_{\gamma_p}^{(j)})] > 0 \text{ and } \sum_{p=1}^{\tilde{q}} \tilde{\alpha}_p[\tilde{g}_{\tilde{\gamma}_p}(\mathfrak{m}_{\tilde{\gamma}_p}^{(i)}) - \tilde{g}_{\tilde{\gamma}_p}(\mathfrak{m}_{\tilde{\gamma}_p}^{(j)})] < 0,$$

so that both $x_{ij} = 1$ and $x_{ij} = 0$ must hold for this pair, which forms a contradiction. Thereby $\alpha_p \in \mathbb{R}$, $p = 1,\dots,q$, and $\tilde{\alpha}_p \in \mathbb{R}$, $p = 1,\dots,\tilde{q}$, are constants as described in the third property of the determining sets.

So let $\mathcal{K}_{max} = \{\gamma_1,\dots,\gamma_q\}$ be the maximal determining set, being constructed from all possible choices of $f_k \in \mathfrak{F}_k$, where g_k, $k = 1,\dots,K$, and α_p, $p = 1,\dots,q$ are the associated choices of functions and constants as given above. In this case it makes sense to put $\hat{f}_{\gamma_p} = g_{\gamma_p}$ and $\hat{a}_{\gamma_p} = \alpha_p \cdot \infty$, $p = 1,\dots,q$, since for this choice of $\hat{f}_{\gamma_p}$, $p = 1,\dots,q$, all factors of the likelihood function which are related to comparisons of objects (i,j) associated with deterministic decisions, i.e. with $\sum_{\ell=1}^{q} \alpha_\ell[M_{\gamma_\ell}^{(i)}-M_{\gamma_\ell}^{(j)}] \neq 0$, tend to 1 if $a_{\gamma_p} \to \infty$ and $a_{\gamma_p}/a_{\gamma_r} = \alpha_p/\alpha_r$, p, r = 1,...,q. $\hat{f}_k$ and $\hat{a}_k$, $k \notin \mathcal{K}_{max}$, may now be estimated from all comparisons (i,j) with $\sum_{\ell=1}^{q} \alpha_\ell[M_{\gamma_\ell}^{(i)}-M_{\gamma_\ell}^{(j)}] = 0$. These estimations are unaffected by f_k and a_k, $k \in \mathcal{K}_{max}$.

It is possible to think about infinite sets $\mathfrak{F}_k$, which in general will lead to a better fit of the model to real data. One may take $\mathfrak{F}_k = \{f_k^{(\underline{\theta}_k)} \mid \underline{\theta}_k \in \Theta_k \subseteq \mathbb{R}^n\}$, with a sufficiently differentiable parametrization, and try to estimate not only the weights $\underline{a}$ but also the parameters $\underline{\theta}_k$, k = 1, ...,K, at a time. We have found two negative effects from this way. First, the distribution of the total observation $(X_{ij}, 1 \leq i<j \leq N)$ or of the statistic $\underline{Z}$ does not longer form an exponential family in general. Secondly, the numerical evaluations become much more complicated (which is in part an effect of the first point). As the applications will reveal, finite $\mathfrak{F}_k$'s are not a very severe limitation.

In this context, the relation between our model and the model of Abelson and Bradley (1954) can be seen more clearly though both objections do not hold for this model. If we put

$$\sum_{\{\mathcal{M}_k^{(i)}, 1 \leq i \leq N\}} \exp\{a_k M_k^{(i)}\} = 1 \quad \left(\text{or} \quad \sum_{\{\mathcal{M}_k^{(i)}, 1 \leq i \leq N\}} M_k^{(i)} = 0,\ a_k = 1\right),$$

k = 1,...,K, as the single restriction to the a_k and f_k, k = 1,...,K, we arrive at the model of Abelson and Bradley (1954). Generally, the $\mathfrak{F}_k$'s cannot be assumed to be finite, except for the instances where $m_k \leq 2$, $k \in \{1,...,K\}$. In these instances, the f_k's may be fixed ($\mathfrak{F}_k = \{f_k\}$) since the a_k's, k = 1,...,K, are still unconfined. For $m_k \geq 3$ this is no longer possible since, by letting vary a_k, one cannot change, e.g., the relation between the differences of the $a_k M_k^{(i)}$,

$i = 1,\ldots,N$, which are unrestricted in the model of Abelson and Bradley (1954). In our model, however, there is only a finite number of possible relations according to the finite $\mathfrak{F}_k$'s. If $m_k \leq 2$, $k = 1,\ldots,K$, each of the two models is evenly transformable into the other one.

Since in our model, $p_{ij}(\underline{a})$ depends on the $M_k^{(n)}$, $k = 1,\ldots,K$, $n = 1,\ldots,N$, only by their differences $M_k^{(i)} - M_k^{(j)}$, $i < j$, $k = 1,\ldots,K$, the restriction described above can be reached by adding a constant b_k to each $M_k^{(n)}$, $n = 1,\ldots,N$, $k = 1,\ldots,K$.

In their (1976) paper, Bradley and El-Helbawy suggested an algorithm to maximize the likelihood function even for factorial designs as in the Abelson-Bradley model. El-Helbawy and Bradley (1977) proved the convergence of this algorithm, provided the assumption of Ford (1957) holds:

> "In every possible partition of the N objects into two nonempty subsets, some objects in the second set has been preferred at least once to some object in the first set." (cf. Zermelo, 1929).

It is clear that the nonexistence of a determining set follows from the Ford assumption, but the converse is not true: In the first psychological experiment, to be discussed in Section 3, in only one of 36 cases a determining set exists (and can be handled by our model), but in 28 of the 36 cases the Ford assumption does not hold.

2.5. A Test for the Model.

As long as the sizes of the sets $\mathfrak{F}_1,\ldots,\mathfrak{F}_K$ are not too large, the following single tests, for every fixed $\underline{f}$, can be combined to a simultaneous test for $\underline{H}$: "$a_1=\ldots=a_K=0$". Its p-value can be bounded from above in the usual manner with the help of the single test's p-values and Bonferroni's inequality. Thus, let $\underline{f}$ be fixed for a moment and let us look for a test for $\underline{H}$ versus the general alternative that $\underline{H}$ is not true. For every $k \leq \{1,\dots,K\}$, for the testing problem $\underline{H_k}$: "$a_k=0$" versus $\underline{K_k}$: "$a_k \neq 0$" there exists an UMPU-test who rejects $\underline{H_k}$ if Z_k falls outside of an interval. But the boundaries of the interval depend not only on the level but also on the values of $Z_1,\dots,Z_{k-1},Z_{k+1},\dots,Z_K$ (cf. Lehmann (1959) p.134). Thus in view of the combinatorial difficulties arising from $\underline{f}$, these tests are practically not performable in general.

Alternatively, let us propose the following asymptotic tests for $\underline{H}$ (instead of $\underline{H_k}$) versus K_k, $k = 1,\dots,K$. The single tests hereby are based on statistics

$$\sum_{i<j} [X_{ij}-1/2][M_k^{(i)}-M_k^{(j)}] \Big/ \Big(1/4 \sum_{i<j} [M_k^{(i)}-M_k^{(j)}]^2\Big)^{1/2} \tag{9}$$

and acceptance regions $[\Phi^{-1}(\alpha/2K),\Phi^{-1}(1-\alpha/2K)]$, $k = 1,\dots,K$, where Φ denotes the standard normal cumulative distribution function and α an upper bound for the level of the corresponding test. A sufficient condition for the asymptotic normality of the test statistics in a sequence of models (2) and (3) (triangular array) is that for large N the following terms tend to 0:

$$\max_{r<s} [M_k^{(r)}-M_k^{(s)}]^2 \Big/ \sum_{i<j} [M_k^{(i)}-M_k^{(j)}]^2, \quad k = 1,\dots,K. \tag{10}$$

This since under $\underline{H}$, the X_{ij}, $1 \leq i < j \leq N$, are independently identically distributed Bernoulli-variables with parameter 1/2, and therefore the conditions are equivalent to the Lindeberg condition (cf. Feller (1971) p. 264(f)).

To reduce the loss of power induced by the use of Bonferroni's inequality, other testing procedures are thinkable, e.g. those which are based on the asymptotic joint normality of the K statistics given in (9).

3. Applications to Psychological Experiments

The model was applied to data from two psychological experiments, each generating a 3 × 3 × 2 - design.

The objects (stimuli) of the <u>first experiment</u> were schematic faces with $K = 3$ attributes: "Mouth" $\mathfrak{m}_1 \in \{1(\frown), 2(-), 3(\smile)\}$, "Hair" $\mathfrak{m}_2 \in \{1(\text{bald}), 2(\text{short}), 3(\text{full})\}$, and "Eyes" $\mathfrak{m}_3 \in \{1(- \ -), 2(\cdot\cdot)\}$, see Fig. 1.

Please insert Fig. 1 here.

Each pair of the $N = 18$ faces has been presented to the judge by projecting the two corresponding slides onto a screen in a random ordering. The judge decided spontaneously, which one of the two faces looked more likable, by crossing against one of two little boxes. In this manner $N(N-1)/2 = 153$ decisions of the judge have been recorded. 36 subjects participated in groups of 4 in this experiment.

To get the numerical analysis manageable, we restrict ourselves to functions f_1, f_2, f_3 which lead to equidistant scales. With other words, for every k, $\{M_k^{(i)} = f_k(\mathfrak{m}_k^{(i)}) \mid i=1,\ldots,N\}$ is a set of equidistant real numbers. Under this restriction, the choice of the admissible functions f_1, f_2, f_3 can be restricted to the following classes without further loss of generality (in a reasonable sense):

$$\begin{aligned} \mathfrak{F}_1 = \{p,q,r\} \text{ where } & p(1) = 1,\ p(2) = 2,\ p(3) = 3 \\ & q(1) = 2,\ q(2) = 1,\ q(3) = 3 \\ \text{and } & r(1) = 1,\ r(2) = 3,\ r(3) = 2, \\ \mathfrak{F}_2 = \mathfrak{F}_1 \text{ and } \mathfrak{F}_3 = \{s\} \text{ where } & s(1) = 1,\ s(2) = 3. \end{aligned} \tag{11}$$

For each fixed $k \in \{1,2,3\}$, the justification is as follows.

In the model (2), (3'), the $M_k^{(i)}$-values appear only in terms of differences. Thus their values can be shifted and their smallest value can be set equal to 1. A γ-fold of the differences can be compensated in the parameters a_k by a change to a_k/γ. Thus the largest value of the $M_k^{(i)}$, $i = 1,\ldots,N$, can be set arbitrarily to a value greater than 1. Here, we set it equal to 3. Now we arrive at scales $\{M_k^{(i)} | 1 \leq i \leq N\} = \{1,2,3\}$, for k = 1,2, and $\{M_3^{(i)} | 1 \leq i \leq N\} = \{1,3\}$. Thus, f_1 and f_2 can be only permutations of (1,2,3). Analogously, for f_3 there are two possibilities. Finally, a complete reverse of the values of a function f_k, e.g. (1,3,2) to (3,1,2), simply changes the signs of all differences of the $M_k^{(i)}$, $1 \leq i \leq N$, and therefore can be compensated by a change of the sign of a_k. The estimator $\underline{\hat{a}}$ has to be evaluated from computations for all triples (f_1,f_2,f_3). As the above considerations have shown, computations for 9 combinations do suffice.

The sole loss of generality results from the equidistantness of the numbers $M_k^{(i)}$, $1 \leq i \leq N$, in other words, from setting the middle value of these numbers equal to 2. If this value would be replaced by a parameter $\nu \in [1,3]$, say, this would lead to full generality and would reveal a direct parallelism to the model of Abelson and Bradley (1954). However, two more parameters, one for the first and one for the second attribute, would be necessary in this instance.

There are two possibilities to handle a lexicographic behavior of subjects. The first is to check the data for all possible combinations of f_k, k = 1,...,K, whether there is a determining set. This is rather laborious. The second is just to check whether the numerical procedure converges in the right way for all combinations of f_k, k = 1,...,K. The results are given in Table 1.

Please insert Table 1 here

The estimated functions $\hat{f}_1$ and $\hat{f}_2$ obviously confirm our condition [2]. A comparison of the $\underline{\hat{a}}$'s among subjects having the same functions $\hat{f}_1$ and $\hat{f}_2$ and signs in the $\hat{a}_k$'s demonstrates additional individual behavior. Since $\hat{f}_3 = s$ holds for each subject, $\hat{f}_3$ is not listed in the table.

A negative sign of an a_k means that the subject shows the opposite preference $\tilde{f}_k$ ($\tilde{f}_k = h \circ f_k$ with $h(1) = 3$, $h(2) = 2$, $h(3) = 1$) to f_k within the attribute $\mathcal{M}_k$, $k = 1,2,3$.

The absolute values of the parameters a_1, a_2, a_3 of a subject indicate which relative importance the three attributes exert on the decision process. This can be seen more clearly by considering the case where $|a_k| > |a_\ell|$,

$$0 \neq M_k^{(i)} - M_k^{(j)} = -(M_\ell^{(i)} - M_\ell^{(j)}), \text{ and } M_p^{(i)} = M_p^{(j)} \text{ otherwise,}$$

$\ell, k, p \in \{1,\ldots,K\}$, $\ell \neq p \neq k$, and $1 \leq i < j \leq N$.

It is easy to see now that here the answer to the question whether p_{ij} is greater or less than 1/2 depends only on the sign of $a_k[M_k^{(i)} - M_k^{(j)}]$.

A negative sign of an $\hat{a}_k$ and the absolute values of the $\hat{a}_1, \hat{a}_2, \hat{a}_3$ are thus to be interpreted analogously.

The data of subject 7 reflect a deterministic behavior with respect to $\mathcal{M}_1$ and neglect of $\mathcal{M}_2$ and $\mathcal{M}_3$ as long as the two faces in a pair differ with respect to $\mathcal{M}_1$. The order of preference for values in $\mathcal{M}_1$ is 3($\smile$) more likable than the other two aspects, and 2(—) more likable than 1($\frown$). Thus, if we set $f_1 = p$,{1} is a determining set. Clearly, it is a minimal one and further considerations show that it is also the maximal one for every choice of f_2. Since for $f_1 = p$ and $a_1 \to \infty$, all factors of the likelihood function tend to 1 which are associated with comparisons of faces differing in their type of mouth, it makes sense to have $\hat{f}_1 = p$ and $\hat{a}_1 = \infty$. $\hat{f}_2$ as well as $\hat{a}_2, \hat{a}_3$ have been computed from all comparisons of faces which have the same type of mouth.

For subject 17, both q and r are maximum likelihood estimates for f_2, and the sign of $\hat{a}_2$ depends on the choice among them. This result reflects a rejection of $\mathcal{M}_2 = 2$, the short hair, and indifference between the two aspects 3(full hair) and 1(bald hair). Similarly, for subject 19, both p and q are maximum likelihood estimates for f_2, leading to the interpretation that $\mathcal{M}_2$ = 3(full hair) is most likable, and no clear preference to either 2(short) or 1(bald) is made.

Column "Subj." gives the subjective rank order of the importance of the attributes for each subject. They have been requested in interviews following the experiment. For instance, (1,3,2) is to be understood that the subject rated $\mathcal{M}_2$ as most and $\mathcal{M}_1$ as least important for her or his judgments. If a subject rated all attributes as equally important this is denoted by a (-,-,-).

As a measure of the difference between the subjective rank order of a subject and her or his estimated rank order (based on $|\hat{a}_1|,|\hat{a}_2|,|\hat{a}_3|$), let us take the minimum number of inversions or transpositions needed to get from one rank order to the other one. For instance, from (1,2,3) to (3,1,2) 2 inversions and from (3,1,2) to (1,3,2) 3 inversions are necessary. There are 9 subjects without any inversion, 14 subjects with one, 4 subjects with two, and 3 subjects with three inversions, giving a total of 30 subjects who could give a subjective rank order. Among those subjects who could give a subjective rank order, a χ^2-test with 3 df can be performed for the hypothesis of "Randomness" versus the alternative that less inversions are more likely. The p-value for these 30 subjects is p = 0.0268.

The last column gives the upper bounds for the p-values of the proposed test ("Randomness" versus "Model") for each subject. Each value, by Bonferroni's inequality, is calculated as the 7-fold of the minimum of the p-values of the single tests, which are taken from Owen (1962). It should be noted that for only 8 of the 36 subjects (16,19,24,27,28,33,34,36) Bradley-Terry-estimates $\hat{\pi}_i$, i = 1,...,N, exist. That means that in most cases, there are partitions of the set of objects in two nonempty subsets where no object in the second set is preferred to any object from the first set and that the second set contains more than one object (one of the π_i's may be zero, $i \in \{1,...,N\}$).

In the second experiment, two variants were realized. The objects in both conditions were descriptions of ten year old boys showing learning difficulties, with K = 3 attributes:

(a) "Description of the student by his teacher" $\mathcal{M}_1 \in$ {1 (he frequently disturbs the lessons), 2(he uptakes slowly), 3(he is afraid of failing)}; "Assistance with the homework" $\mathcal{M}_2 \in$ {1(by day nurses), 2(by his siblings), 3(by his mother)}; "HAWIK-intelligence score (a German version of the WISC)" $\mathcal{M}_3 \in$ {1(103),2(91)}.

(b) In this condition, $\mathcal{M}_2$ was replaced by $\mathcal{M}_4$ "EEG-diagnosis of a neurologist" $\in$ {1(mild lability of the brain functions), 2(mild nervousness), 3(mild vegetative disorder)}. A total of 28 subjects (teachers and graduate students in education) participated individually in this experiment, 12 in condition (a), and 16 in condition (b). The pairs of objects were printed in a booklet in a random ordering. The subjects decided who of the two students in a pair would need more urgently support by a special education.

Due to its small diagnosticity, $\mathcal{M}_4$ was expected to be cognized as the least important attribute in condition (b). Since the importance of the attributes $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$, as conditions for learning disabilities, is viewed controversially among the experts, individual preferences were expected. The results are given in Table 2.

Please insert Table 2 here.

The suggested procedure for the numerical analysis of the estimates $\hat{f}_k$ and $\hat{a}_k$ in the first experiment was also applied to the judgments in the second experiment. As Table 2 shows, 4 subjects decide lexicographically with respect to $\mathcal{M}_3$, and further two subjects decide strictly lexicographically with respect to all three attributes. Thus, more frequently than in the first experiment, determining sets can be identified in the second experiment. The results again corroborate the justification of condition [2], that is to say, there are more different than common strategies among the subjects.

The expectation concerning the relative small importance of $\mathcal{M}_4$ in condition (b) was tested by assigning ranks to the estimated weights $|\hat{a}_k|$ for each subject. The squares of the sums of ranks for each k were jointly transformed into the Friedman-statistic, which is χ^2-distributed with 2 df under the null hypothesis: "Equal importance of the attributes". The resulting $\chi^2 = 13.625$ with a p-value $p < 0.005$ confirms our expectation.

Numerical calculations

The numerical calculations have been performed on the HB 66/80 at the University of Mainz. The solutions of the system of equations (7) are based on a procedure of Werner (1979), which converges somewhat faster than the classical, and also suitable, Newton-Raphson procedure.

4. Discussion and Summary

The present paper proposed a model for paired comparisons of multiattribute social stimuli which can be considered as a stochastic special case of the additive difference model (Tversky, 1969), and has relations to some other models, for instance to the BTL-model (Luce, 1959). It satisfies the conditions, given in the introduction, that scale values for the attributes are not available, and that each judge should be considered individually, i.e. aggregating across judges would misrepresent the individual judgment strategies.

As the results of the two experiments show, judges should in fact be regarded individually. Our condition [2] is especially important if the aspects of the attributes have no natural order. (We prefer to speak of attributes instead of features since all the schematic faces having been evaluated in the first experiment, for instance, have hair. On the other hand, if the schematic faces could be separated into two classes whose members either have or have no hair, the term feature would be more appropriate.)

Intuitively, representing the processing of schematic faces by an additive model seems to be not conclusive. Recent approaches on processing information in social perception assume a partition of the set of objects into classes which are best represented by prototypes (cf. Mervis & Rosch, 1981). If the judgments $U(G^{(i)})$ on the schematic faces of the first experiment would have been guided by such a processing strategy, classes of more or less likable schematic faces should have led to clear preferences between the members of different classes and to indifferences between the members within classes. However, the empirical evidence does not support such interpretation. It should be quoted that Mervis and Rosch (1981) found in the present literature

> "considerable disagreement as to whether faces should be considered special holistically perceived objects" (p. 105).

First, the judgment situation of requesting paired comparisons should be taken into account since it favours a comparison of the objects by their common attributes (e.g. Tversky & Krantz, 1969). Even if a task consists of assigning schematic faces to one of two selected classes of faces which are represented by some typical members, the similarity between the schematic faces and the classes was well represented by a weighting feature model (Reed, 1972).

That means, a process of abstraction requires the attention mainly to the individual attributes that best discriminate between the two classes.

Secondly, when having subjects judge pairs of schematic faces coming from sets in which the aspects for some attributes differ only quantitatively, for instance when the eyes differ only in their distance, then the conditions of transitivity and independence are less likely satisfied than in our first experiment with its qualitative material. Additionally, subjective reports on the judgment strategies revealed more attention to the individual attributes than under the conditions where the schematic faces had only one or no attribute with qualitative aspects (Mattenklott, 1979). Qualitative aspects seem to increase the independence of the individual attributes and thus to increase their importance when processing schematic faces into impressions of likability.

Thirdly, for nearly all of the subjects, the judgments from the two experiments did not deviate severely from the necessary independence conditions. Rather, the deviations can be explained by the stochastic component of the evaluation process. This result neither suggests another simple polynomial nor a classification of the schematic faces into likable and unlikable objects as a suitable model.

The arguments for a processing of objects with special respect to their attributes do not intend to challenge a coding of information that can be described as a construction of classes which are represented by prototypes. However, as our arguments should demonstrate, task characteristics as well as the type of objects require a process of abstraction which favours the integration of schematic faces by its attributes.

Obviously, the process of abstraction should be more important when the attributes describing a student with learning difficulties should be integrated to a judgment of how severe his learning disability is. This process of abstraction together with the easily separated attributes of verbal information should lead more distinctly to an attribute-based judgment. The results of the second experiment corroborated this conjecture since 6 of 28 subjects decided strictly lexicographically with respect to at least one attribute. The results likewise reflected the minor importance of the attribute $\mathcal{M}_4$ as supposed in the second variant of the second experiment. This result would have been less probable if the subjects had processed the objects configurally.

On the other hand, empirical evidence for the attribute-based judgments do not allow strict implications on the particular integration rule F. Different integration rules are compatible with our model, for instance a weighted attribute-wise comparison, an additive combination of the attributes that leads to a separate evaluation of the respective objects, or a lexicographic rule as performed by some of the subjects, especially in the second experiment. Following the line of identifying the particular integration rule F, an axiomatic analysis in the sense of Krantz et al. (1971) could have been performed. However, this approach contains the problem of how to handle the error (stochastic remainder).

Acknowledgement. The authors are grateful to Professor P. H. Schönemann for several helpful comments.

References

ABELSON, R. M. & BRADLEY, R.A. A 2 x 2 factorial with paired comparisons. Biometrics, 1954, 10, 487-502.

BICKEL, P. J. & DOKSUM, K. A. Mathematical Statistics: Basic Ideas and Selected Topics. San Francisco: Holden-Day, 1977.

BRADLEY, R. A. Science, statistics and paired comparisons. Biometrics, 1976, 32, 213-232.

BRADLEY, R. A. & EL-HELBAWY, A. T. Treatment contrasts in paired comparisons: Basic procedures with application to factorials. Biometrika, 1976, 63, 255-262.

BRADLEY, R. A. & TERRY, M. E. The rank analysis of incomplete block designs, I. The method of paired comparisons. Biometrika, 1952, 39, 324-345.

EL-HELBAWY, A. T. & BRADLEY, R. A. Treatment contrasts in paired comparisons: Convergence of a basic iterative scheme for estimation. Communications in Statistics - Theory and Methods, 1977, A6 197-207.

FELLER, W. An Introduction to Probability Theory and Its Applications. Vol. II (2nd ed.). New York: Wiley, 1971.

FORD, L. R. Solution of a ranking problem from binary comparisons. American Mathematical Monthly, 1957, 64 II.

GENSCH, H. D. & RECKER, W. W. The multinomial, multiattribute logit choice model. Journal of Marketing Research, 1979, 16, 124-132.

GRIZZLE, J. E., STARMER, C. F. & KOCH, G. G. Analysis of categorical data by linear models. Biometrics, 1969, 25, 490-504.

KRANTZ, D. H., LUCE, R. D., SUPPES, P. & TVERSKY, A. Foundations of Measurement, Vol. I. New York: Academic Press, 1971.

LEHMANN, E. L. Testing Statistical Hypotheses. New York: Wiley, 1959.

LUCE, R. D. Individual Choice Behavior. New York: Wiley, 1959.

LUCE, R. D. & SUPPES, P. Preference, utility, and subjective probability. In R. D. LUCE, R.R. BUSH & E. GALANTER (Eds.), Handbook of Mathematical Psychology, Vol. III. New York: Wiley, 1965.

MATTENKLOTT, A. Über die Bedeutsamkeit von Komponenten des Urteilsmaterials für die Angemessenheit einfacher Polynome der Informationsintegration. In L. H. ECKENSBERGER (Ed.) Bericht über den 31. Kongreß der DGfP, Vol. I. Göttingen: Verlag für Psychologie, 1979.

MERVIS, C. B. & ROSCH, E. Categorization of natural objects. Annual Review of Psychology, 1981, 32, 89-115.

OWEN, D. B. Handbook of Statistical Tables. Reading, Mass.: Addison-Wesley, 1962.

REED, S. K. Pattern recognition and categorization. Cognitive Psychology, 1972, 3, 382-407.

SRINIVASAN, V. & SHOCKER, A. D. Estimating the weights for multiple attributes in a composite criterion using pairwise judgments. Psychometrika, 1973, 38, 473-493.

TVERSKY, A. Intransitivity of preferences. Psychological Review, 1969, 76, 31-48.

TVERSKY, A. & KRANTZ, D. H. Similarity of schematic faces: A test of interdimensional additivity. Perception & Psychophysics, 1969, 5, 124-128.

WERNER, W. Über ein Verfahren der Ordnung $1 + \sqrt{2}$ zur Nullstellenbestimmung. Numerische Mathematik, 1979, 32, 333-342.

ZERMELO, E. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift, 1929, 29, 436-460.

F i g u r e 1

The 18 schematic faces of the first experiment

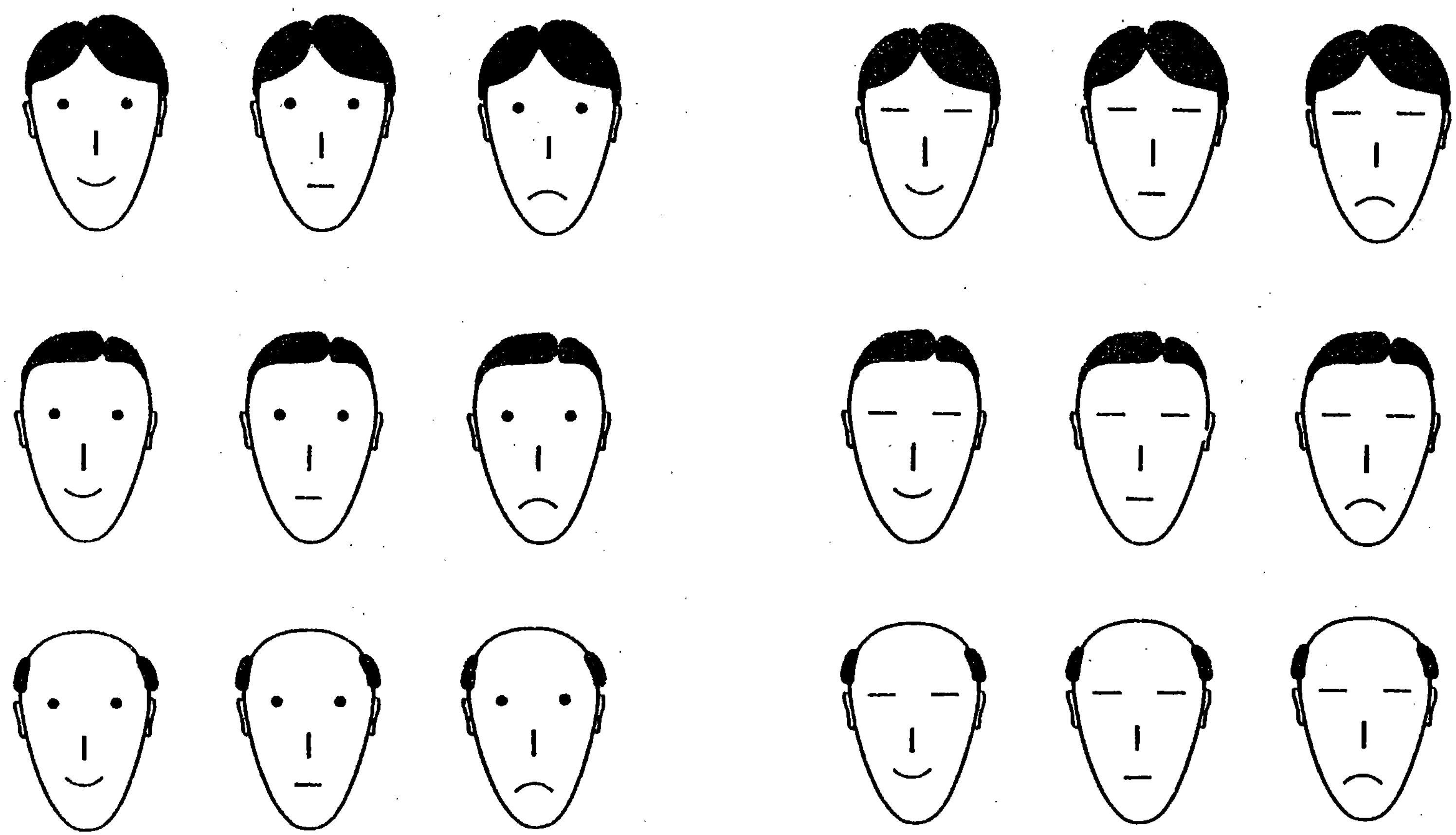

TABLE 1

Estimates, subjektive rank orders, and p-values of 36 subjects from the first experiment

Subject	$(\hat{f}_1,\hat{f}_2)$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	Subj.	Test
1	(p,p)	8.7230	2.7481	1.2790	(3,1,2)	$5.60 \cdot 10^{-21}$
2	(p,p)	5.5402	1.8673	1.7035	(2,1,3)	$1.59 \cdot 10^{-18}$
3	(p,q)	4.2700	1.5842	1.5627	(1,3,2)	$5.59 \cdot 10^{-17}$
4	(p,q)	2.8581	1.1590	1.6735	(3,1,2)	$9.54 \cdot 10^{-13}$
5	(p,p)	3.2006	1.1475	1.0484	(3,1,2)	$3.13 \cdot 10^{-16}$
6	(q,p)	1.8752	1.0757	0.6223	(3,2,1)	$4.22 \cdot 10^{-12}$
7	(p,r)	∞	0.8339	-0.7444	(3,1,2)	$5.60 \cdot 10^{-21}$
8	(q,p)	2.2223	0.7462	1.5745	(2,1,3)	$7.33 \cdot 10^{-11}$
9	(p,q)	4.6310	1.8901	0.4377	(3,2,1)	$2.51 \cdot 10^{-19}$
10	(p,p)	2.4133	0.7431	2.0370	(1,3,2)	$2.88 \cdot 10^{-10}$
11	(p,p)	6.3412	2.4086	0.4456	(-,-,-)	$3.82 \cdot 10^{-20}$
12	(p,p)	2.4624	2.5545	0.7648	(2,1,3)	$2.88 \cdot 10^{-10}$
13	(p,r)	1.6597	-1.9288	1.0031	(-,-,-)	$3.96 \cdot 10^{-9}$
14	(p,q)	3.8592	0.9703	1.6856	(2,1,3)	$3.13 \cdot 10^{-16}$
15	(p,q)	2.9308	-0.2548	1.4654	(3,1,2)	$8.71 \cdot 10^{-15}$
16	(q,r)	0.7244	-0.3808	1.0754	(2,1,3)	$1.09 \cdot 10^{-9}$
17	(p,q or r)	1.7691	±1.6961	0.8527	(-,-,-)	$2.55 \cdot 10^{-8}$
18	(q,r)	1.4558	-1.0516	1.3633	(1,3,2)	$1.39 \cdot 10^{-8}$
19	(p,p or q)	1.1351	0.9076	0.1316	(2,1,3)	$8.39 \cdot 10^{-8}$
20	(p,r)	1.7985	-0.6672	1.2310	(-,-,-)	$1.09 \cdot 10^{-9}$
21	(p,p)	1.8366	1.8763	1.0929	(3,2,1)	$4.65 \cdot 10^{-8}$
22	(p,p)	2.6216	2.3006	0.8293	(1,3,2)	$7.33 \cdot 10^{-11}$
23	(p,p)	2.9433	1.9670	0.6170	(2,3,1)	$4.34 \cdot 10^{-14}$
24	(p,p)	1.6562	-0.4739	0.2743	(3,1,2)	$2.08 \cdot 10^{-13}$
25	(p,p)	2.3590	0.6951	0.5175	(3,1,2)	$3.13 \cdot 10^{-16}$
26	(p,q)	2.1148	1.8883	0.3272	(2,3,1)	$7.33 \cdot 10^{-11}$
27	(q,q)	0.8842	1.1139	0.3190	(1,3,2)	$2.66 \cdot 10^{-7}$
28	(p,p)	1.2671	0.2191	0.3805	(3,1,2)	$2.88 \cdot 10^{-10}$
29	(p,p)	2.6836	-0.1918	1.3614	(2,1,3)	$8.71 \cdot 10^{-15}$
30	(p,p)	2.5130	0.3020	0.7664	(3,1,2)	$3.13 \cdot 10^{-16}$
31	(p,q)	0.6130	1.8514	2.5592	(1,2,3)	$1.69 \cdot 10^{-15}$
32	(p,r)	1.3106	-1.2606	2.4690	(1,2,3)	$1.69 \cdot 10^{-15}$
33	(p,q)	0.1412	0.4580	-0.2010	(-,-,-)	$1.36 \cdot 10^{-2}$
34	(r,r)	-0.8450	-1.4463	-0.1024	(-,-,-)	$7.33 \cdot 10^{-11}$
35	(p,r)	3.0601	0.4474	2.6794	(2,1,3)	$4.22 \cdot 10^{-12}$
36	(p,p)	1.6102	0.3132	0.4356	(3,1,2)	$9.54 \cdot 10^{-13}$

TABLE 2

Estimates from the second experiment

	Subject	$(\hat{f}_1,\hat{f}_2)$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$
Variant (a)	1	(p,p)	-3.0963	-2.1347	1.8576
	2	(p,p)	1.8510	1.8510	1.1574
	3	(p,p)	1.4581	-2.3901	0.8868
	4	(r,p)	-0.4136	2.5793	∞
	5	(q,p)	2.6174	0.9962	1.3898
	6	(r,r)	1.8205	1.0398	3.3550
	7	(p,p)	1.5986	2.6369	2.3595
	8	(p,p)	1.9963	3.7057	1.5508
	9	(p,q)	0.9250	1.2809	0.4629
	10	(q,q)	-2.3121	1.6188	0.4488
	11[1]	(q,p)	∞	$-\infty$	∞
	12	(q,p)	3.0755	0.9752	1.7471
		$(\hat{f}_1,\hat{f}_2)$	$\hat{a}_1$	$\hat{a}_4$	$\hat{a}_3$
Variant (b)	13	(p,r)	4.4130	-1.5565	2.1746
	14	(p,r)	-0.7567	0.2670	0.1928
	15	(q,r)	-1.0317	-2.0087	3.0843
	16	(q,p or r)	2.1162	±0.4851	1.0253
	17	(q,p)	-2.3604	0.9555	1.7578
	18	(p,p)	-1.5368	0.5693	1.1138
	19	(p,p)	1.6156	1.5282	1.7371
	20	(q,q)	3.0095	0.4570	∞
	21	(q,p)	-1.3605	0.5470	1.7768
	22[2]	(r,p)	∞	$-\infty$	∞
	23	(p,p)	-1.4268	3.0567	∞
	24	(p,q)	1.7389	0.6341	1.0820
	25	(p,q)	1.3159	0.8184	1.8267
	26	(p,p)	-3.3090	0.5771	1.8202
	27	(p,p)	0.4711	0.5947	∞
	28	(p,q or r)	-2.6879	±0.9802	4.0594

1,2: Subject 11 decided strictly lexicographically with $m_3 > m_2 > m_1$, and subject 22 with $m_1 > m_3 > m_4$.