PROBABILISTIC UNFOLDING MODELS FOR THE ANALYSIS OF CHANGE

Mark de Rooij*and Willem J. Heiser

Leiden University

In the present paper we present distance-association models for the representation of association in the analysis of cross-classified data. More specifically we present models for the analysis of transition tables, i.e. cross-classifications of one variable measured at two time points. Log-linear and log-bilinear models have become useful models for such data. However, the interpretation of log-bilinear models is not direct. A graphical plot resulting from a log-bilinear model can only be interpreted by projection. A reparametrization results in an unfolding model with a graphical representation in which the distances between points referring to the categories in the rows and the columns can be interpreted directly. A large distance corresponds to a small association; a small distance corresponds to a large association.

Key words: Euclidean Distances, $RC(M)$-association model, cross-classified data

*E-mail: rooijm@fsw.leidenuniv.nl. Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was conducted while the first author was supported by a grant of the Foundation for Behavioural and Educational Sciences of this organization (575-30-006).

1 INTRODUCTION

The last two decennia a wide variety of models for the analysis of contingency tables have been proposed. Especially the work of Goodman (1972, 1979, 1981, 1985, 1986), Andersen (1980), Gilula and Haberman (1986), Becker (1990), and Haberman (1974, 1978, 1979, 1995) on log-linear and $RC(M)$-association- and correlation models are important developments. The latter models give an intermediate between the model of independence and the saturated model in the traditional log-linear analysis approach, by various restrictions on the association terms. In the present paper we will propose another model in which the interpretation of the graphical representation is more straightforward than the usual $RC(M)$-model parameterization. The topic is related to the difference between Correspondence Analysis (CA; Nishisato, 1980; Gifi, 1990) and Multidimensional Scaling (MDS; Borg and Groenen, 1997), and the generalization of the latter to rectangular matrices, Multidimensional Unfolding (MDU; Heiser, 1981). In CA only the distances within the set of row-points, or within the set of column-points can be interpreted directly, provided that certain coherent normalizations are chosen. The relation between row- and column-points can only be assessed by projection (cf. Greenacre, 1984, 1993). In MDS and MDU all distances between points can be interpreted directly. The inner product rule in CA is replaced by a distance rule in MDU for the relationship between row-points and column points. For a detailed discussion on the relationships between MDS, MDU, and CA the reader is referred to Heiser and Meulman (1983).

2 THE $RC(M)$-ASSOCIATION MODEL

Let us start with a multiplicative model for the expected frequencies of a two-way contingency table (π_{ij}), that is

$$\pi_{ij} = \alpha_i \beta_j \eta_{ij}, \tag{1}$$

where $\alpha_i, i = 1, \ldots, I$, and $\beta_j, j = 1, \ldots, J$, are main effect parameters, and the η_{ij} are interaction parameters. If we set the interaction terms equal to unity we obtain the usual model of independence. In empirical data we often find that the independence model does not fit the data. Moreover, the saturated model fits the data but is of no use since there is not any reduction of data. One way to proceed is to restrict the η_{ij} parameters,

for example with a bilinear decomposition of the logarithm of the interaction term, i.e.

$$\eta_{ij} = \exp(\sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm}). \tag{2}$$

This decomposition was proposed by Goodman (1979, 1981, 1985, 1986, 1991), under the name $RC(M)$-association model. The complete model then is written as

$$\pi_{ij} = \alpha_i \beta_j \exp(\sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm}). \tag{3}$$

The ϕ_m are called intrinsic association parameters, the μ_{im} are called row-scores and the ν_{jm} are called column scores. If these scores originate from a singular value decomposition, they satisfy $\sum_i \mu_{im}^2 = \sum_j \nu_{jm}^2 = 1$, for each m, and $\sum_i \mu_{im}\mu_{im'} = \sum_j \nu_{jm}\nu_{jm'} = 0$ for any pair of dimensions m and m'. If $M = M^* = \min(I, J) - 1$, the $RC(M)$-association model is equal to the saturated model, if $M = 0$ the model reduces to the model of independence. For the interpretation of this model usually graphical displays are used. One often sees two single displays, one for the row scores and one for the column scores. Since only indirect relations between the rows and indirect relations between the columns are displayed, there is no way to really interpret the association. Only joint graphical displays of the rows and the columns can show how any one category of the row variable is associated with some category of the column variable.

A number of joint graphical displays can be used. However, as Greenacre (1984, p.65) noted:

> *"There are advantages and disadvantages of the simultaneous display. Clearly an advantage is the very concise graphical display expressing a number of different features of the data in a single picture. The display of each set of points indicates the nature of similarities and dispersion within the set. Notice, however, that we should avoid the danger of interpreting the distances between the points of different sets, since no such distances have been explicitely defined*".

In general all plots with row coordinates $\mu_{im}^* = \phi_m^\tau \mu_{im}$ and column coordinates $\nu_{jm}^* = \phi_m^\kappa \nu_{jm}$, where $\tau + \kappa = 1$ are mathematically correct. In practice one often finds one of the following graphical displays (here the terminology of a major statistical package is used; cf. Meulman, Heiser and SPSS Inc., 1999):

1. *Row principal normalization:* Plot the row categories as points with coordinates $\mu'_{im} = \phi_m \mu_{im}$ and the column categories as vectors with coordinates ν_{jm}. In row

principal normalization, the Euclidean distances between the row points approximate (possibly weighted) Euclidean distances between the row entries of the contingency table, after they have been logarithmically transformed and corrected for the main effects. The column vectors have a direction and a length. The former can be used to reconstruct the association with the row categories by projection, the latter indicates how well a column fits the chosen dimensionality.

2. *Column principal normalization:* Plot the row categories as vectors with coordinates μ_{im} and the column categories as points with coordinates $\nu'_{jm} = \phi_m \nu_{jm}$. The interpretation of this display is analogous to (1), with the role of rows and columns reversed.

3. *Symmetric normalization:* Plot the row categories as points with coordinates $\mu''_{im} = \phi_m^{\frac{1}{2}} \mu_{im}$ and the column categories as points with coordinates $\nu''_{jm} = \phi_m^{\frac{1}{2}} \nu_{jm}$. This normalization spreads the intrinsic association terms symmetrically over the rows and over the columns. Note that neither the distances between the row points nor the distances between the column points are approximations to data related distances. This plot can only be interpreted by projecting the row (column) points onto the direction indicated by any column (row) point.

4. *Principal normalization:* Plot the row categories as points with coordinates $\phi_m \mu_{im}$ and the column categories as points with coordinates $\phi_m \nu_{jm}$. Here the intrinsic association terms are spread twice in the solution, once over the row scores and once over the column scores. This is basically a wrong graphical display since $\tau + \kappa \neq 1$. So this method of normalization can only be used for making separate plots of row categories and column categories, respectively.

In the next Section we will show the models in a longitudinal context, i.e. the situation where one categorical variable is measured at two time-points (here $I = J$). Such a table is called a transition frequency table. In this kind of data the intrinsic association terms of the $RC(M)$-association model are often relatively large. Therefore the first two kinds of graphical displays become problematic, since in the row principal normalization the column points cluster in the center, and in the column principal normalization the row-points cluster around the center.

Table 1: Wiepkema Data

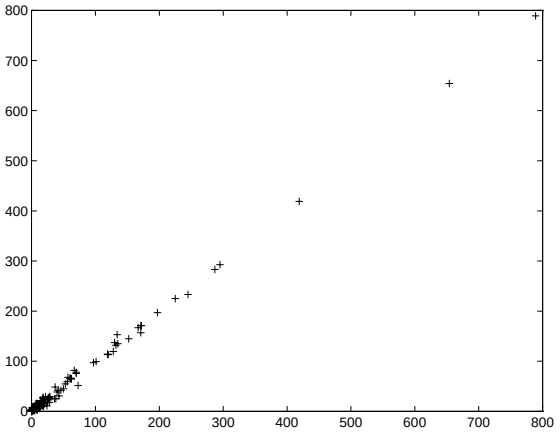
follow / *preceed*	JK	TU	HB	CHS	FL	QU	LE	HDP	SK	SN	CHF	FFL
jk	654	128	172	56	27	25	1	28	0	46	14	18
tu	101	132	62	27	5	1	1	11	0	8	5	9
hb	171	62	197	130	0	25	0	50	14	18	14	12
chs	60	22	152	135	0	8	0	43	16	15	12	4
fl	19	2	0	0	419	19	0	2	0	17	5	11
qu	36	1	18	5	12	789	119	295	26	70	1	14
le	4	0	0	0	0	57	167	73	0	8	0	0
hdp	22	9	40	37	5	245	7	171	287	53	8	13
sk	3	2	7	38	0	120	8	134	19	28	4	0
sn	42	2	17	16	20	70	11	67	9	225	12	12
chf	18	3	10	13	6	5	0	8	0	24	97	9
ffl	27	3	6	5	10	13	0	18	0	10	8	29

In the third Section we will discuss a model proposed by De Rooij and Heiser (2000), that has a more direct interpretation of the joint graphical display in terms of distances. The models give a particularly nice representation of longitudinal data: given the marginal proportions, a large distance corresponds to a small transition frequency; a small distance corresponds to a large transition frequency. The result can be interpreted as a transition map, in which people travel from one category point to another. More transitions occur between categories that are close together; less transitions occur between categories that are far apart. Before we discuss this model in more detail, we first show an empirical example, and discuss the problems with the joint graphical displays obtained from the usual $RC(M)$-parameterization.

3 THE PROBLEM: AN EXAMPLE

We will discuss the problems noted above in more detail here using an empirical example. Table 1 gives data of reproductive behavior of male bitterlings, studied by Wiepkema (1961). The data are derived form 13 sequences using a moving time-window of size

Figure 1: Observed Versus Expected Frequencies For $RC(2)$-Association Model With Inheritance Terms.



two. The 12 categories of behaviors are jerking (jk), turning beats (tu), head butting (hb), chasing (chs), fleeing (fl), quivering (qu), leading (le), head down posture (hdp), skimming (sk), snapping (sn), chafing (chf), and finflickering (ffl). We use small letters for the rows and capitals for the columns.

We analyzed these data with the $RC(M)$-association model with two components. The program LEM was used to fit the model (Vermunt, 1997), but other programs can be used as well. As is usually done with this kind of data, an extra set of parameters (called 'inheritance' parameters or 'immobility' parameters) is used to equalize the expected values on the diagonal to their observed counterparts. The fit of this model is $X^2 = 294; LR = 252; df = 69$. This is a reasonable fit for such a large table. Compared to the quasi-independence model, 93.4 percent of the association is accounted for. As can be seen from Figure 1, the observed values are all well approximated.

The joint plots with row principal and column principal normalization for the $RC(M)$-association model are shown in Figure 2 and 3. As in Table 1 we use small letters for the row points and capitals for the column points. As can be seen in these two figures,

Figure 2: Row principal normalization For Wiepkema Data

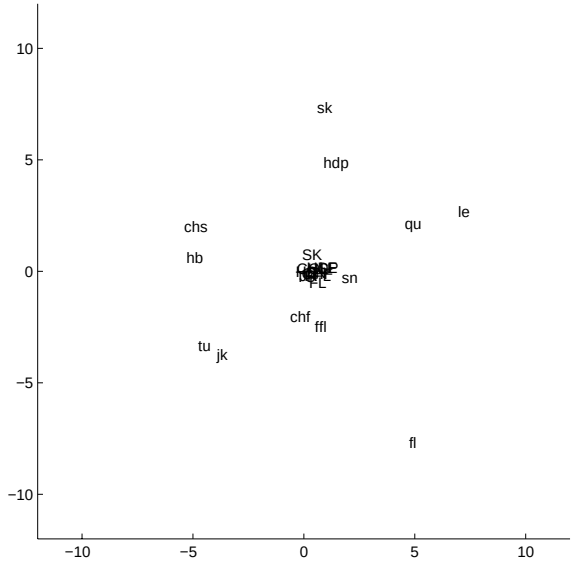
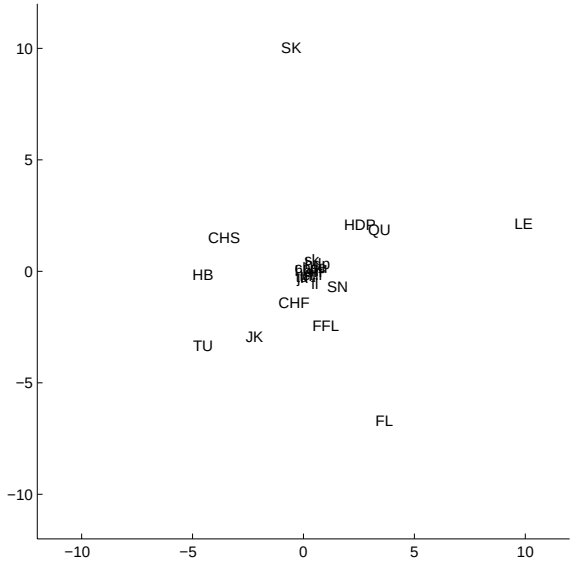


Figure 3: Column principal normalization For Wiepkema Data



either the row points or the column points are all clustered in the center, which makes these plots close to useless for interpretation.

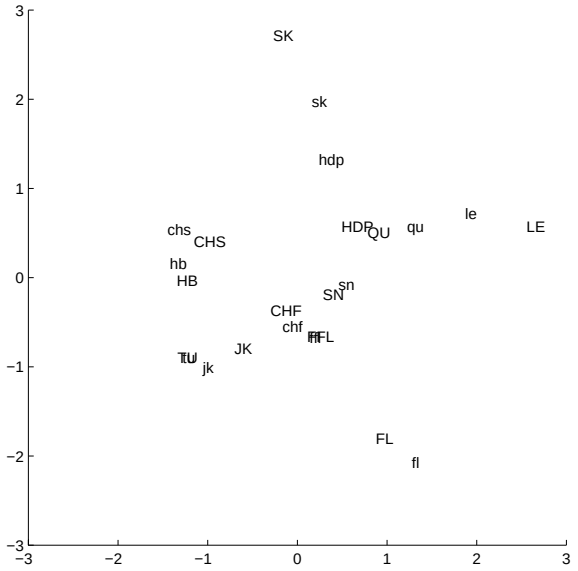
A plot with symmetrical normalization is given in Figure 4. In this plot neither the row points nor the column points are in a Euclidean metric. The plot can only be interpreted by projection of the row points onto the vectors of column points or the other way around. To reiterate, distances between row and column points should not be interpreted, nor can distances between the row (column) points be interpreted.

4 A REPARAMETRIZATION

Instead of bilinear decomposition we will make use here of unfolding distances. As it will turn out the model is a reparametrization of the $RC(M)$-association model. Our joint graphical display is better interpretable, however.

First, assume a Euclidean distance function on a set of row-coordinates collected in a $I \times M$-matrix $\mathbf{X}$, and a set of column coordinates collected in a $J \times M$-matrix $\mathbf{Y}$. The

Figure 4: Symmetric Normalization For Wiepkema Data.



squared distance is defined as

$$d_{ij}^2(\mathbf{X};\mathbf{Y}) = \sum_{m=1}^{M} (x_{im} - y_{jm})^2. \tag{4}$$

Using this squared distance instead of the bilinear part in the $RC(M)$-association model, that is, modeling the interaction parameters as

$$\eta_{ij} = \exp\left(-d_{ij}^2(\mathbf{X};\mathbf{Y})\right), \tag{5}$$

we obtain our distance-association model. The transformation (5) is called the Gaussian transformation and was used by Shepard (1957,1987; Nosofsky, 1985). Explicitely, the complete model is

$$\pi_{ij} = \alpha_i \beta_j \exp\left(-d_{ij}^2(\mathbf{X};\mathbf{Y})\right). \tag{6}$$

Like in the $RC(M)$ model, when $M = 0$ the model reduces to the independence model. If $M = M^* = \min(I, J) - 1$, the model becomes the saturated model. In an unfolding plot the distances between the row points and the column points are genuine distances, and can be interpreted as such. The distances between the row points (column points) are not related to observations; they can, however, be interpreted as the distances between the row-scores of the $RC(M)$-association model. Row points (column points) that are in the same position have the same profile.

Let's go back to the definition of the $RC(M)$-association model

$$\begin{aligned}\pi_{ij} &= \alpha_i\beta_j \exp\left(\sum_{m=1}^{M}\phi_m\mu_{im}\nu_{jm}\right)\\ &= \alpha_i\beta_j \exp\left(\sum_{m=1}^{M}2u_{im}v_{jm}\right),\end{aligned} \tag{7}$$

where $u_{im} = \sqrt{\frac{1}{2}}\phi^{\frac{1}{2}}\mu_{im}$, and $v_{jm} = \sqrt{\frac{1}{2}}\phi^{\frac{1}{2}}\nu_{jm}$. If we now define $a_i = \alpha_i \times \exp(\sum_{m=1}^{M} u_{im}^2)$, and $b_j = \beta_j \times \exp(\sum_{m=1}^{M} v_{jm}^2)$, we obtain

$$\pi_{ij} = \frac{a_i b_j \exp\left(\sum_{m=1}^{M} 2u_{im}v_{jm}\right)}{\exp(\sum_{m=1}^{M} u_{im}^2) \times \exp(\sum_{m=1}^{M} v_{jm}^2)}. \tag{8}$$

By reordering the expressions, we obtain

$$\begin{aligned}\pi_{ij} &= a_i b_j \exp\left(-\sum_{m=1}^{M}u_{im}^2 - \sum_{m=1}^{M}v_{jm}^2 + \sum_{m=1}^{M}2u_{im}v_{jm}\right)\\ &= a_i b_j \exp\left(-\sum_{m=1}^{M}(u_{im}-v_{jm})^2\right),\end{aligned} \tag{9}$$

which brings us back at our distance-association model. The distance-association model is a reparametrization of the $RC(M)$-association model that allows for a distance interpretation.

For a further comparison, the likelihood function under Poisson sampling for both models is written

$$L = \sum_{ij} f_{ij} \log \pi_{ij} - \sum_{ij} \pi_{ij}, \tag{10}$$

where f_{ij} are the observed frequencies. Let's consider the first derivatives of this likelihood function with respect to the score parameters of the $RC(M)$-model and the coordinate parameters of the distance-association model. For the $RC(M)$-association model the first derivative with respect to μ_{im} is given by

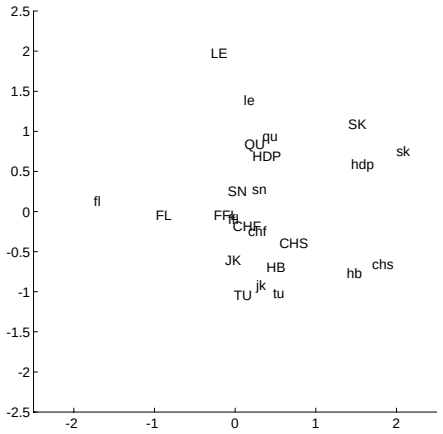
$$\frac{\partial L}{\partial \mu_{im}} = \sum_{j}(\pi_{ij} - f_{ij})\phi_m\nu_{jm}. \tag{11}$$

For the distance-association model the first derivative with respect to x_{im} is given by

$$\frac{\partial L}{\partial x_{im}} = \sum_{j}(\pi_{ij} - f_{ij})(x_{im} - y_{jm}). \tag{12}$$

The derivatives are different, and this has the important implication that generally the estimates are different. Although the expected frequencies for both models are the same

Figure 5: Distance Plot For Wiepkema Data Obtained With Algorithm



after convergence, the parameter estimates of the distance-association model and the $RC(M)$-association model with reparametrization given in (7) to (9) are not the same. De Rooij and Heiser (2000) proposed this distance-association model and also present an algorithm to fit the model directly.

5 THE EXAMPLE REVISITED

In Figure 5 the result of the distance-association model is given. Figure 5 is obtained with the algorithm presented by De Rooij and Heiser (2000).

In Figure 5 distances between row and column points may be interpreted, since the model is explicitely defined in terms of these distances. A small distance corresponds to a large value for the association, i.e. a large number of transitions between the two categories. For example, jerking (jk) is close to turning beats (TU), and head butting (HB), indicating many transitions occur from jerking to either turning beats and head butting. A large distance corresponds to a small numerical value for the association, i.e. a small

number of transitions. For example, jerking (jk) is far away from skimming (SK), and leading (LE), indicating few transitions occur from jerking towards skimming or leading. A row category on the boundary of the joint graphical display (for example fleeing (fl)) indicates not many transitions occur from this particular category towards other categories. A column point on the boundary of the joint graphical display (for example leading (LE)) indicates not many transitions occur towards this category. Categories in the center of the joint graphical display are often visited or left.

In the $RC(M)$-association model one often tries to find categories with equal profiles, that is the same pattern of frequencies in the corresponding rows or columns of the table. In the $RC(M)$-association model such points have the same row- or column-scores. The same reasoning can be applied to our distance-association model.

6 CONCLUSIONS

In cases where either the intrinsic association terms of the $RC(M)$-association model are large or small, the joint plots obtained with the row principal normalization or the column principal normalization are close to useless for interpretation. The symmetric normalization has the major drawback that neither the distances between the row points, nor the distances between the column points are approximations to data related distances.

We think that a joint plot in terms of distances is better understood then a joint plot based on projections. Therefore, we defined a model in terms of distances explicitely. We show that the joint graphical display is indeed easily interpreted: A small distance corresponds to a large association; a large distance corresponds to a small association.

The expected frequencies under the $RC(M)$-association model and the distance-association model are equal. The reasons to choose one model above the other can only be on substantive grounds. We think that whenever one looks for an easy interpretation of the data one should choose for the distance-association models presented here, because of the more direct interpretation of the results.

References

Andersen, E. B. (1980). *Discrete statistical models with social science applications.* Amsterdam: North-Holland.

Becker, M. P. (1990). Maximum likelihood estimation of the RC(M) association model. *Applied statistics*, *39*, 152-167.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling; theory and applications.* New York: Springer.

De Rooij, M., & Heiser, W. J. (2000, submitted). Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: John Wiley and Sons.

Gilula, Z., & Haberman, S. J. (1986). Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association*, *81*, 780-788.

Goodman, L. A. (1972). Some multiplicative models for the analysis of cross-classified data. *Sixth Berkely Symposium*, *1*, 649-696.

Goodman, L. A. (1979). Simple models for the analysis of association in cross classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537-552.

Goodman, L. A. (1981). Association models and cannonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, *76*, 320-334.

Goodman, L. A. (1985). The analysis of cross-classified data having ordered and/or unordered catergories: Association models, correlation models, and asymmetric models for contingency tables with or without missing entries. *The Annals of Statistics*, *13*, 10-69.

Goodman, L. A. (1986). Some useful extensions to the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review*, *54*, 243-309.

Goodman, L. A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data. *Journal of the American Statistical Association*, *86*, 1085-1111.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Greenacre, M. J. (1993). *Correspondence analysis in practice.* London: Academic Press.

Haberman, S. J. (1974). *The analysis of frequency data.* Chicago: university of Chicago Press.

Haberman, S. J. (1978). *Analysis of qualitative data, vol. 1.* New York: Academic Press.

Haberman, S. J. (1979). *Analysis of qualitative data, vol. 2.* New York: Academic Press.

Haberman, S. J. (1995). Computation of maximum likelihood estimates in association models. *Journal of the American Statistical Association*, *90*, 1438-1446.

Heiser, W. J. (1981). *Unfolding analysis of proximity data.* Unpublished doctoral dissertation, Leiden University.

Heiser, W. J., & Meulman, J. (1983). Analyzing rectangular tables by joint and constrained multidimensional scaling. *Journal of Econometrics*, *22*, 139-167.

Meulman, J. J., Heiser, W. J., & SPSS Inc. (1999). *Spss categories 10.0.* Chicago, IL: SPSS Inc.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto press.

Nosofsky, R. (1985). Overall similarity and the identification of seperable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, *38*, 415-432.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325-345.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Vermunt, J. K. (1997). *ℓEM: A general program for the analysis of categorical data* (Tech. Rep.).

Wiepkema, P. R. (1961). An ethological analysis of the reproductive behavior of the bitterling (rhodeus amarus bloch). *Archives Neerlandais Zoologique*, *14*, 103-199.