Prediction of Sensory Preference and Choice: Recent Progress with the Unfolding Model

Willem J. Heiser
Leiden University

(1) From testing & description to prediction

(2) Prediction of preference rankings and ratings

(3) Prediction of categorical outcomes

People Are Different, Especially When Surrounded by Competition

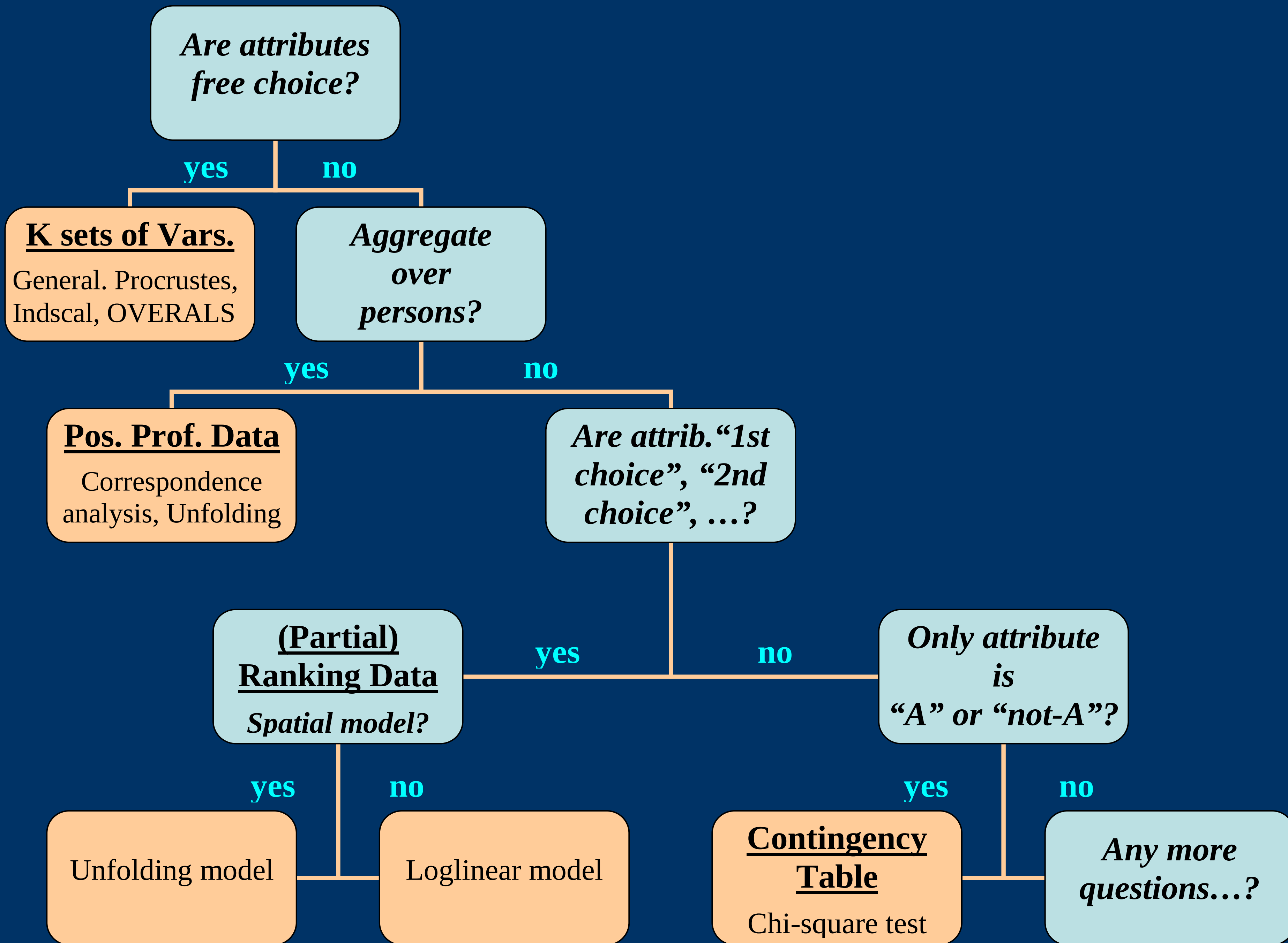
**Our focus will be:
How to deal with
individual differences?**

Framework for Sensory Response Data: Humans as Instruments

	Person 1 attributes 1, …, m_1	Person 2 attributes 1, … , m_2	…………	Explanatory attributes 1, … , p
1 … sensory stimulus … n				By design or measured
Explanatory Person characteristics	*constant*	*across*	*attributes*	

- **Usually we have more variables than cases;**
- **Multi-level (nested, hierarchical) observations**

Idea of Tree of Methods for Sensory Research Designs (unfinished)



Some Observations on Sensory Response Data

1. *Human as an instrument*: observations are likely to have dependencies and biases. Explanatory person characteristics lead to mixed level data.

2. In *descriptive sensory analysis*, there seems to be no consensus about methodology and vocabulary [we see Flavor Profile method, Texture Profile method, Quantitative Descriptive analysis™, Spectrum™ method, Free-choice profiling …]
   It is time to share databases and reach for the Big Five !
   (cf. Zarzo and Stanton, 2009, who suggest standard sensory maps are possible, at least for odor descriptors).

3. *Process, process, process*. As is evident from Time-Intensity studies, aroma and flavor change over time. In psychology, we have learned to view many phenomena as processes rather than as traits or states [emotion, memory, personality, psychopathology…].

Why is Growing Interest in Prediction to be Expected?

- Presence of substantial and/or changing individual differences needs an ***explanation***, and being able to statistically predict effects is better than offering post-hoc after-thoughts;
- In development of new products, it is good to know their sensory profile in advance on the basis of ***product components*** or variations in production processes.
- Because of the data mining revolution, new ***statistical learning*** methods become available every day! Predictability is the new standard for model selection, variable selection, and much more
- Allows for more complex theoretical relations to be tested.
- …

What is the Difficulty of Predicting Variables?

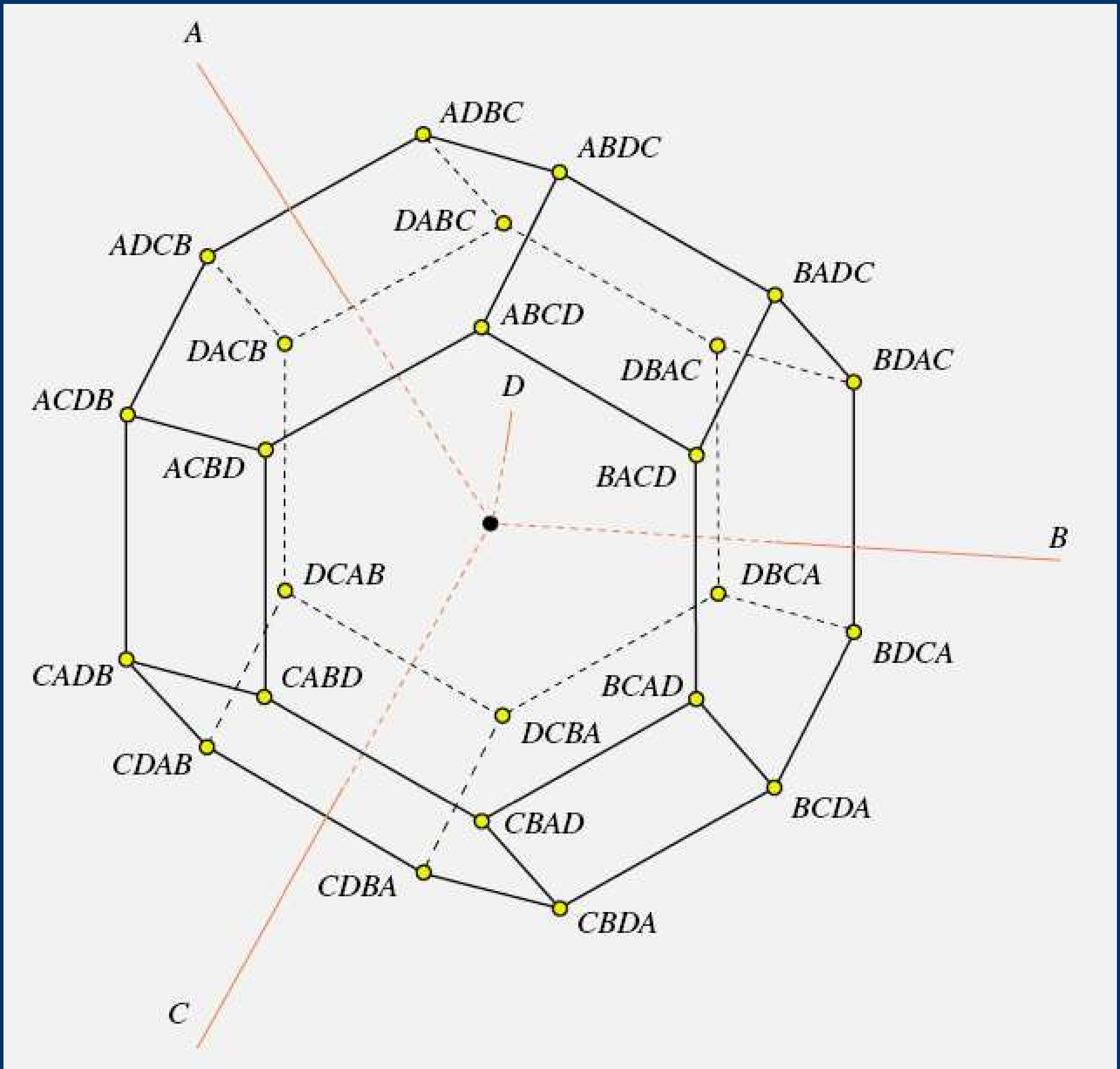
It is useful to call the assessor/panelist/expert/consumer in a sensory study a ***variable***, because he or she assigns a value from some range of values to each element in a domain of sensory objects.

There is no difficulty in predicting variables from other variables. Usually done with some Structural Equation Model (SEM) or with Partial Least Squares (PLS) procedures.

Consider a variable as a direction in some high-dimensional space. [think of a correlation matrix measuring the angles between vectors]

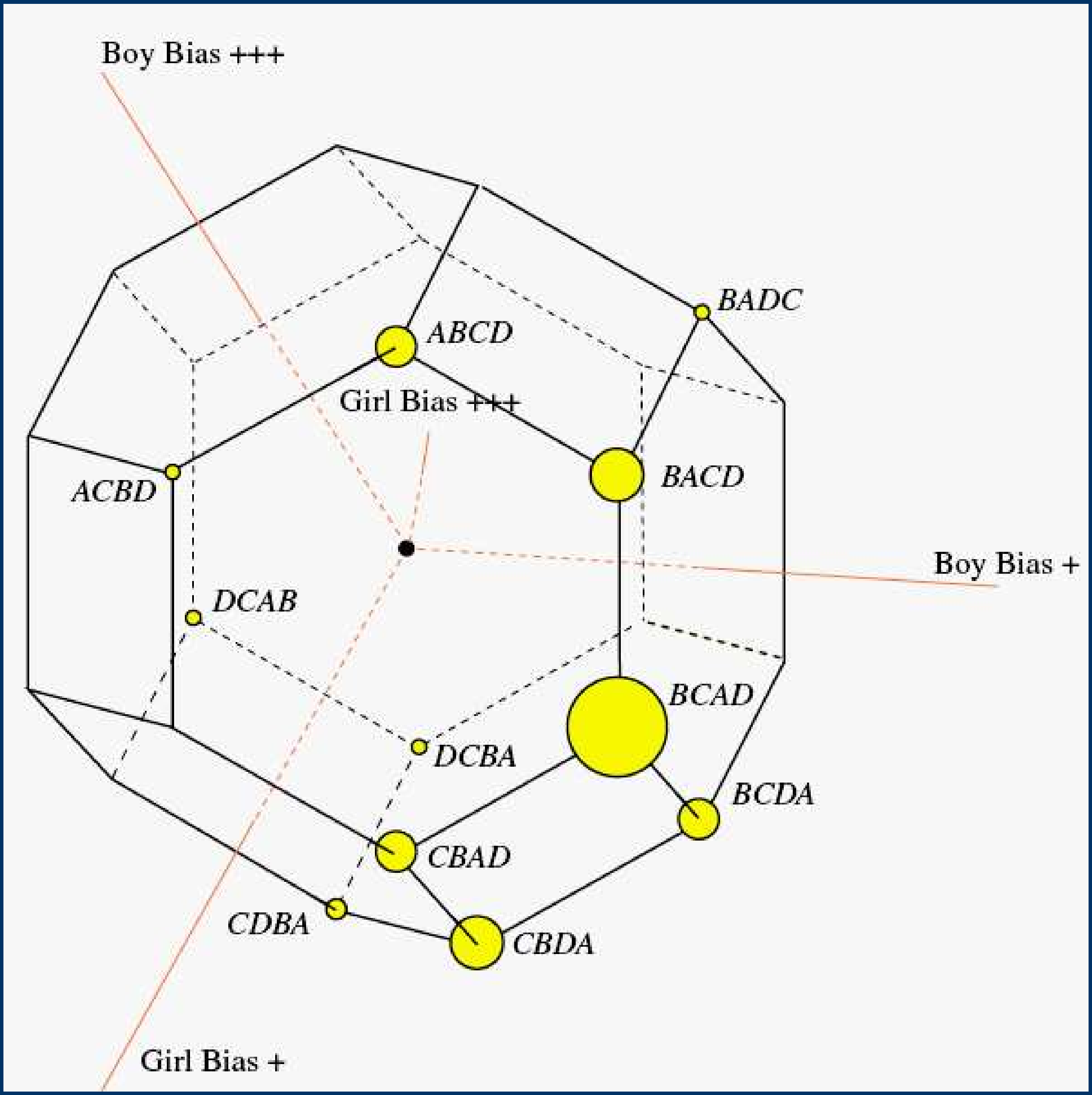
> Then it is much less obvious
> how to predict a direction in high-dimensional space
> from some score on a line, or
> from some category label

Prediction of Rankings: Sample Space is a Polytope



- There are $4 \times 3 \times 2 \times 1 = 24$ possible rankings in the case of 4 options.
- Distance along edges between two rankings is related to ***Kendall's tau***: number of adjacent transpositions
- Data = distribution over the vertices of the polytope.

Preferences for Family Types with Equal Number of Children



Example

Selection of Delbeke's (1968) family types, total of 3 children $N = 102$.

A = (3 boys, 0 girls)
B = (2 boys, 1 girl)
C = (1 boy, 2 girls)
D = (0 boys, 3 girls)

Yellow circles have radius proportional to square root of their frequency. *BCAD* is most prevalent order.

How to Predict a Ranking from Explanatory Person Characteristics?

There seem to be two major possibilities:

a) ***Simple prediction with a categorical variable***. For example, if we have gender, we could calculate two average or median rankings. The median ranking is one node of the ranking polytope that has least Kendall distance towards all rankings present.

b) ***Building a Classification Tree for Rankings***. If we have more predictors, perhaps both numerical and categorical, we can do optimal prediction using a new kind of CART methodology (d'Ambrosio and Heiser, 2009, work in progress).

Example: Sex Differences in Rankings of Odorants

Source:
Moncrieff (1966),
data on 132 odorants,
16 rankings
by preference.

Time to produce
the orderings varied
from 2 to 3 hours
to one and a half day.

Odorants to which the reactions of men and women were considerably different

Odorant material	Placing of odorant by: Men 15 years and over	Placing of odorant by: Women	Odour preferred by	Ratio of placings of men and women
Honeysuckle flower	1·3	4·5	Men	1 : 3·5
Wild rose	7·5	22·2	Men	1 : 3·0
Heliotrope	16·0	29·4	Men	1 : 1·8
Mock orange	11·0	45·0	Men	1 : 4·1
Mint	19·0	34·5	Men	1 : 1·8
Vanilla essence	49·0	28·1	Women	1·7 : 1
Ylang-ylang	14·8	34·4	Men	1 : 2·3
Cinnamon leaf oil	28·3	43·8	Men	1 : 1·5
Citral	26·3	52·0	Men	1 : 2·0
Cinnamon bark oil	30·8	49·9	Men	1 : 1·6
Bay leaf	48·3	24·6	Women	2·0 : 1
Musk ambrette	19·5	58·8	Men	1 : 3·0
Geraniol ex Palma Rosa	27·0	54·6	Men	1 : 2·0
Lemongrass oil	25·5	54·3	Men	1 : 2·1
Onion, Raw	82·3	40·9	Women	2·0 : 1
Heliotropin	72·0	46·1	Women	1·6 : 1
Ethyl alcohol	59·8	39·0	Women	1·5 : 1
Alpine violet perfume	75·5	36·4	Women	2·1 : 1
Allyl caproate	54·3	87·5	Men	1 : 1·6
Naphthalene	103·0	69·6	Women	1·5 : 1
Toluene	105·3	69·3	Women	1·5 : 1
Skatole	107·8	87·1	Women	1·2 : 1
Cade oil	62·3	99·8	Men	1 : 1·6

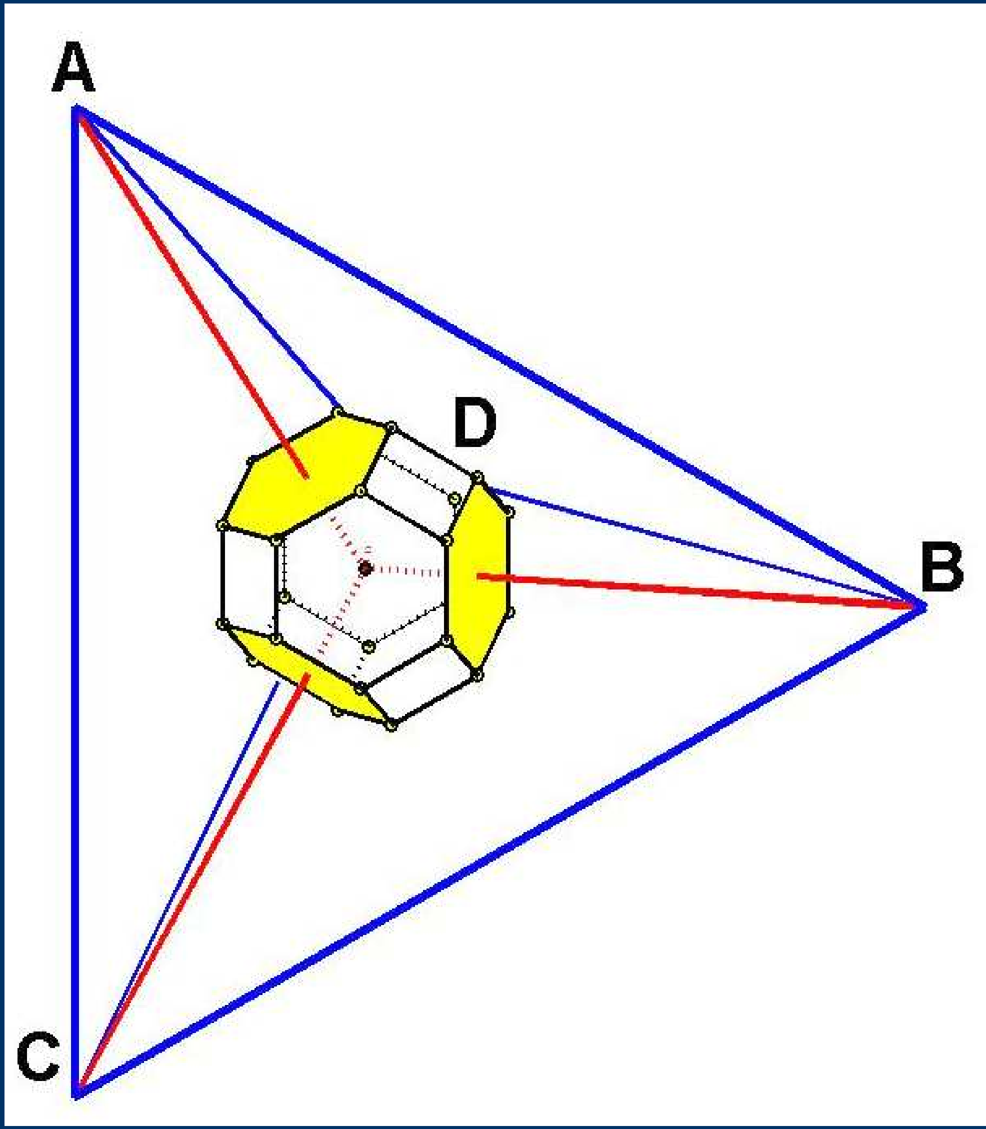
How to Predict Rank of Stimulus From Explanatory Attributes?

Note that sample space of rankings does not contain points for stimuli or choice objects.

SO IT APPEARS WE ARE STUCK

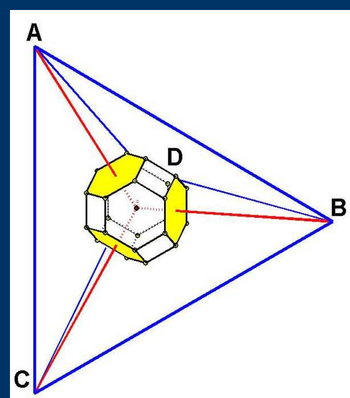
Fortunately, Heiser (2004) showed that sample space can be exactly reconstructed by determining ***centers of gravity*** obtained by using the rank numbers as weights in averaging with respect to a ***simplex of stimuli***

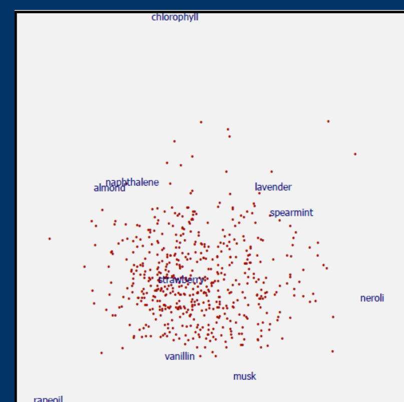
[*cf*. correspondence analysis]



Common Idea of Unfolding Models

In all unfolding models, the idea is to map closeness of a ranking to a stimulus into another space that contains two sets of points, one called ***ideal points*** (person points), and the other called ***stimulus points*** (object or option points).



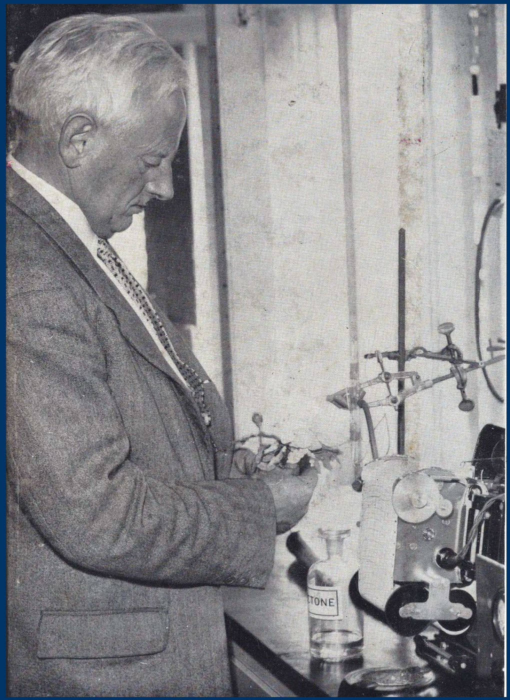


Both ideal points and stimulus points are in the same space, and can be restricted in function of a number of explanatory variables.
[in ecology, a pioneer of rectricted unfolding is Cajo ter Braak, with his CANOCO program]

Example: Unfolding of Odour Rankings (Moncrieff, 1966)

In his pioneering monograph entitled *Odour Preferences*, Moncrieff reported an empirical study in which he tried to assess dependence of olfactory preferences on sex, age, and temperament. Stimuli used:

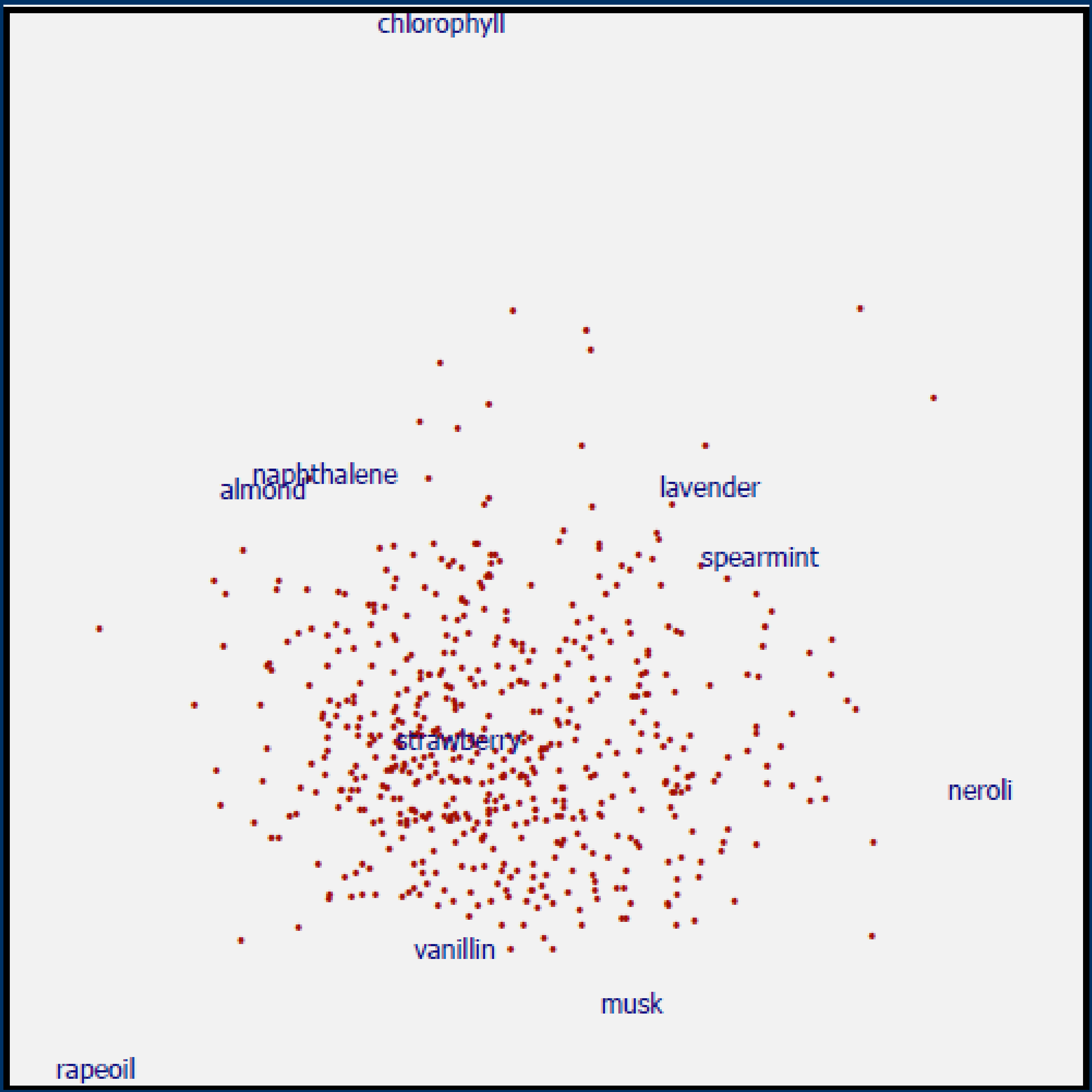
1. **Strawberry** flavoring essence;
2. **Spearmint** oil;
3. French **lavender** oil (with high ester content);
4. **Musk** lactone (100%);
5. **Vanillin** (essential odorant of vanilla pod);
6. **Neroli** oil;
7. **Almond** flavoring essence;
8. **Naphthalene** (smell of moth-balls & fire-lighters);
9. **Rape** oil (nutty, oily)
10. Oil-soluble **chlorophyll** (strong flavor, ☹)



Moncrieff placed the odorant materials in 6 oz. glass bottles with wide necks and ground glass stoppers and asked subjects to sniff them successively and then arrange the bottles in order of liking.

N = 559 here, and m= 10.

PREFSCAL Mapping of Odour Preferences (Distance Biplot)



Major features:

Chlorophyll and rape oil are disliked by most persons, and end up on the edge.

Strawberry flavoring is liked by many, located in center.

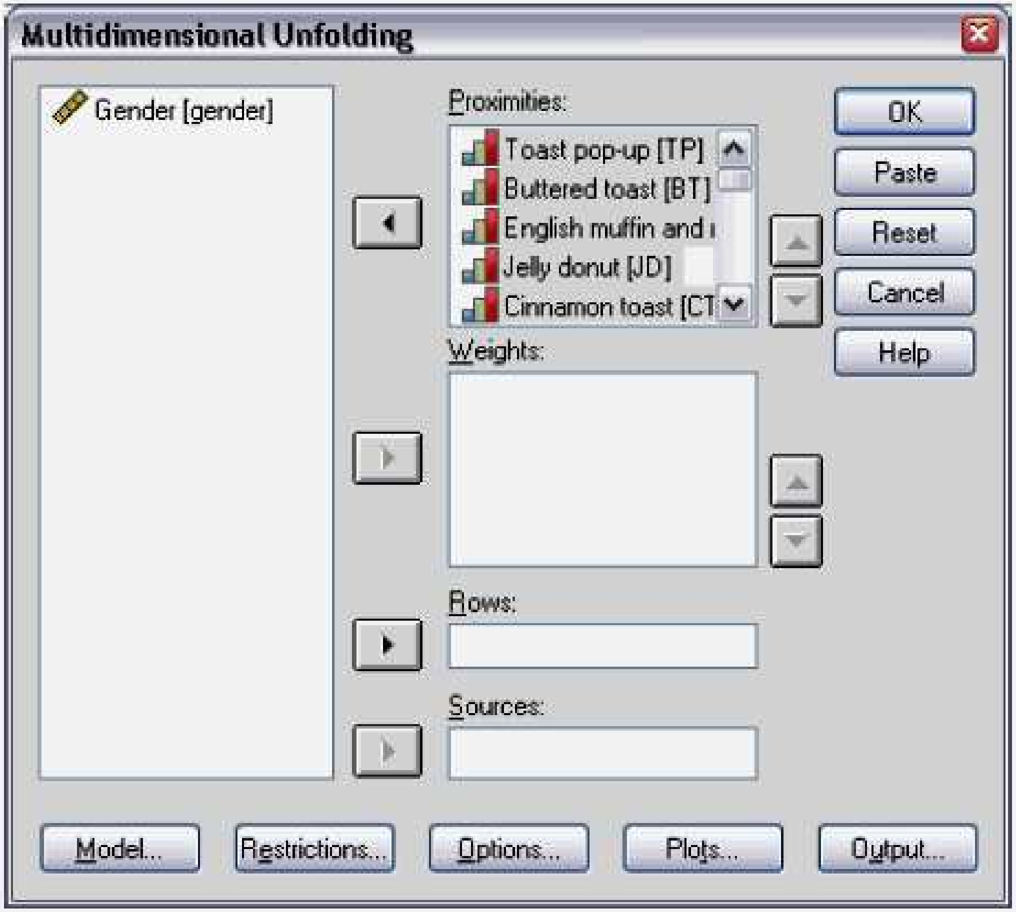
Stress-I = 0.21,
VAF = 0.81,
Rho = 0.83.

For 3-dim. solution:
Stress-I = 0.0294

The PREFSCAL program in SPSS Categories®

PREFSCAL is an unfolding program developed in Leiden that starts from one or more rectangular tables with *proximities*. It attempts to find a common quantitative scale (or space) that accounts for the *relationships between two sets of objects* (persons and stimuli).

Multidimensional Unfolding main dialog box



PREFSCAL Minimizes Stress With Penalty Term

PREFSCAL calculates a configuration **X** for row objects, another one **V** for column objects, and determines inter-point distances $d(\mathbf{X}, \mathbf{V})$.

To evaluate the quality of the solution, we first find optimal transformations of the proximities $\gamma(\mathbf{P})$ or $\gamma_i(\mathbf{P})$ and then calculate

$$\min_{\Gamma,\mathbf{X},\mathbf{V}} \left\{ \sum_i \frac{\|\gamma_i(\mathbf{P}) - d_i(\mathbf{X},\mathbf{V})\|^2}{\|\gamma_i(\mathbf{P})\|^2} \right\}^{\lambda} \left\{ \frac{1}{I} \sum_i \left(1 + \frac{\omega}{\upsilon^2[\gamma_i(\mathbf{P})]} \right) \right\}.$$

Stress term *Penalty term*

- **Penalty term is necessary to avoid problem of *degeneration*.**
- Function $\nu[\,\cdot\,]$ is Pearson's *coefficient of variation*.
- Tuning parameter $0 \leq \lambda \leq 1$ controls balance between stress and regularization by the penalty term (default = 0.5). The parameter $\omega \geq 0$ (currently, $\omega\upsilon^2[\mathbf{p}_i]$) controls the *range* of the penalty term.

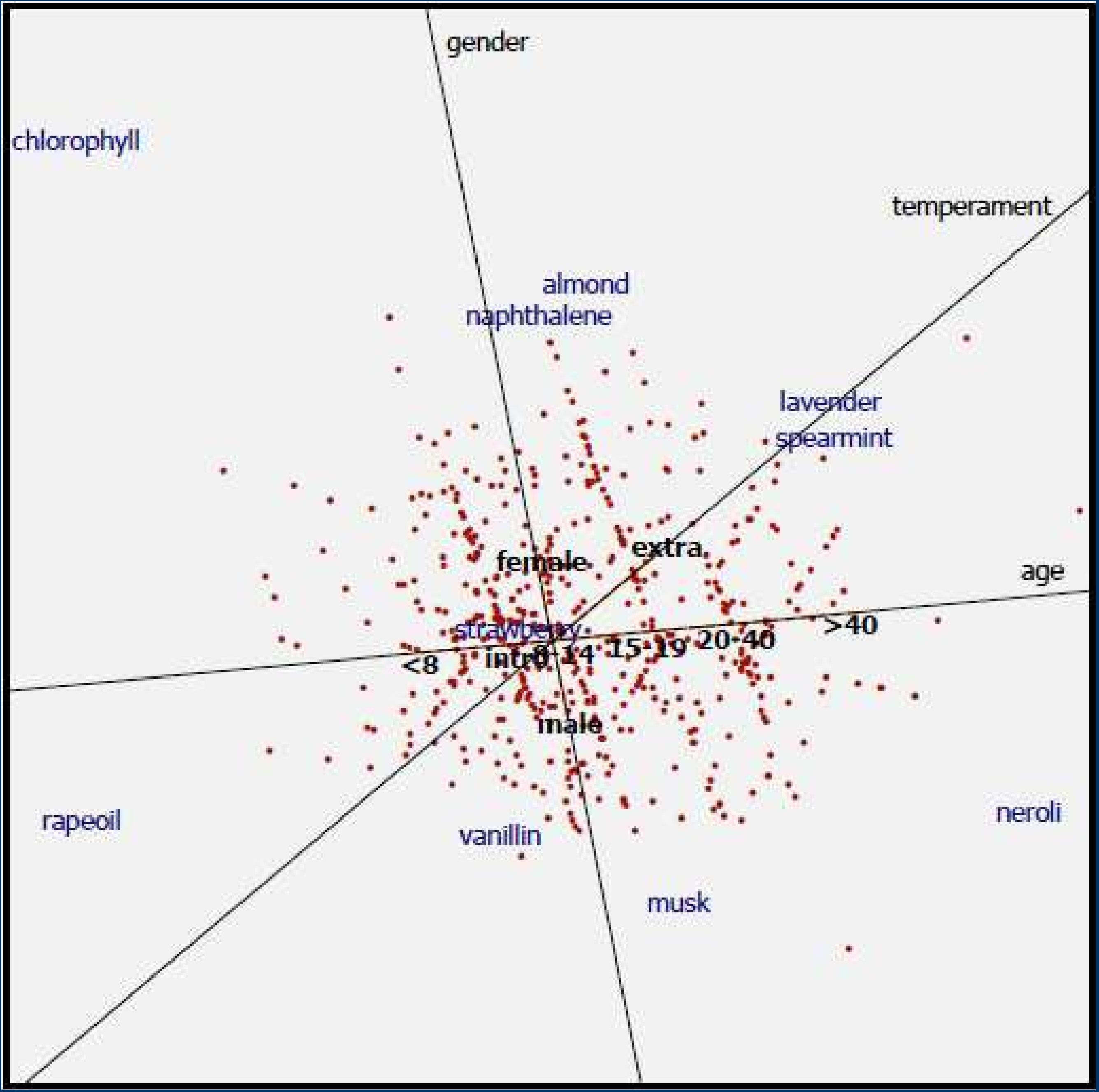
Prediction Through Restriction in Unfolding

All unfolding models have location parameters for persons and location parameters for stimuli. When $\mathbf{Q}_X$ and $\mathbf{Q}_V$ are the matrices of prediction variables for persons and stimuli, respectively, we add the restrictions:

$$\boxed{\begin{aligned}\mathbf{X} &= \mathbf{Q}_X \mathbf{B}_X, \\ \mathbf{V} &= \mathbf{Q}_V \mathbf{B}_V.\end{aligned}}$$

Depending on the specific model and optimization method used, usually estimation of the regression weights $\mathbf{B}_X$ and $\mathbf{B}_V$ is reasonably standard. In PREFSCAL, we use the projected gradient method and Alternating Least Squares for the restrictions.

Triplot of Restricted PREFSCAL on Odour Preferences



VAF of explanatory variables:
Gender = 0.78
Age = 0.70
Temperament = 0.87

(*contrary to Moncrieff's conclusion in Rule 62*)

Stress-I = 0.23,
VAF = 0.78,
Rho = 0.79.

Fit of rankings is still reasonable

Second example: Soup Rating Data (Busing et al., 2010, in *FQP*)

Flavour Intensity	Sourness	Median	Mean	Standard deviation	Share of choices (%)
Descriptive statistics					
Low	Low	6	6.01	2.08	14
	Medium	6	5.93	2.00	13
	High	5	4.97	2.14	5
Medium	Low	7	6.32	2.06	17
	Medium	6	5.95	1.92	11
	High	6	5.52	1.97	8
High	Low	6	6.01	2.14	15
	Medium	6	5.91	1.98	11
	High	5	5.25	2.02	6

Mixed model analysis of variance

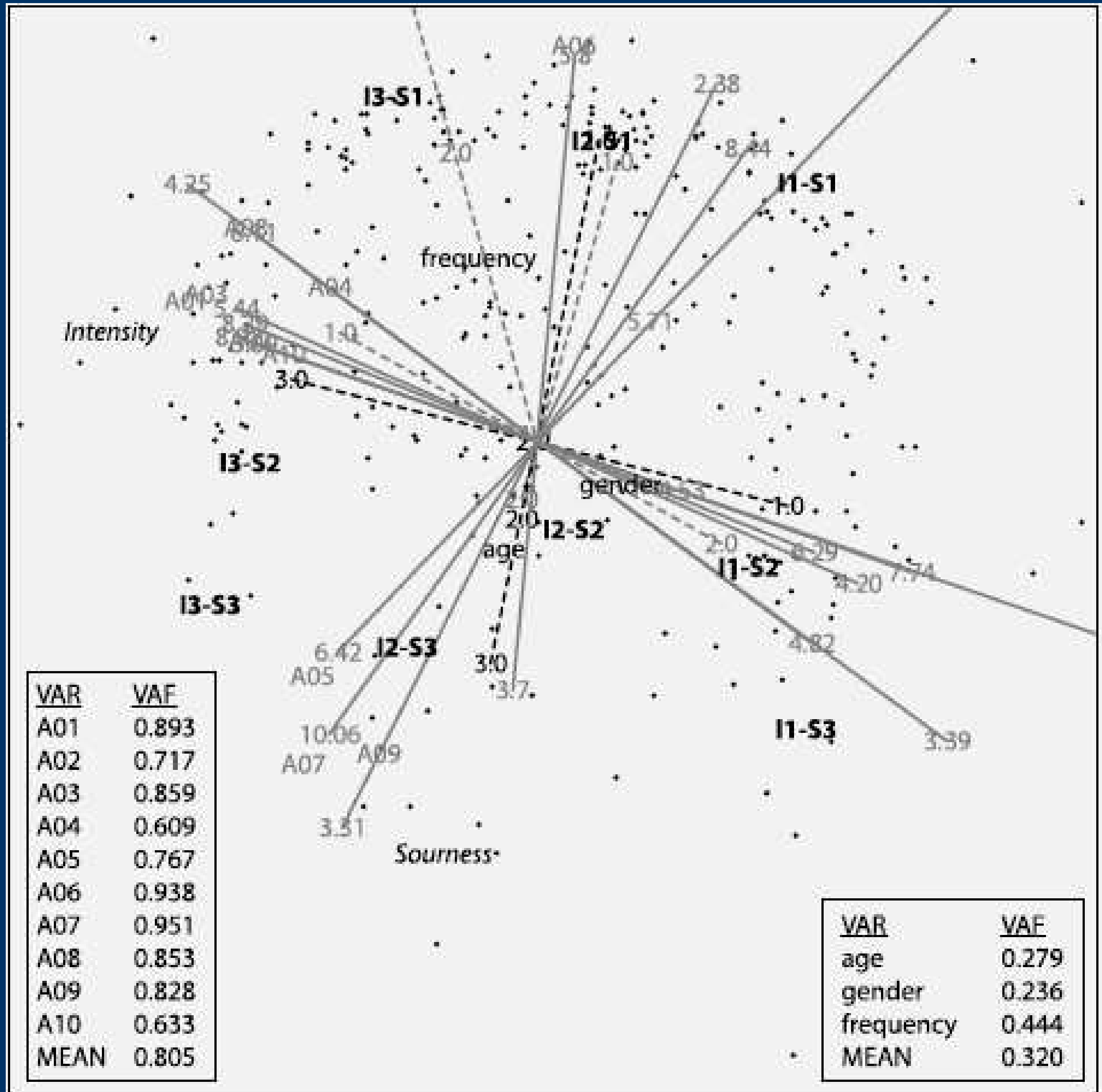
Fixed effects	Numerator DF	Denominator DF	F	Sig.
Flavour Intensity	2	594	5.25	.005
Sourness	2	594	57.40	.000
Flavour Intensity × Sourness	4	1188	2.41	.048
Random effects	**Estimate**	**Standard Error**	**Wald Z**	**Sig.**
Respondents	0.94	0.12	7.50	.000
Respondents × Flavour Intensity	0.50	0.08	5.99	.000
Respondents × Sourness	0.27	0.07	3.76	.000
Residuals	2.44	0.10	24.37	.000

$N = 298$, assessments on a nine-point liking scale.

Notes.

- Largest effect in the ANOVA is Respondents (level effect);
- Importantly, interaction effects of respondents with flavor intensity and sourness are significant and moderately large.

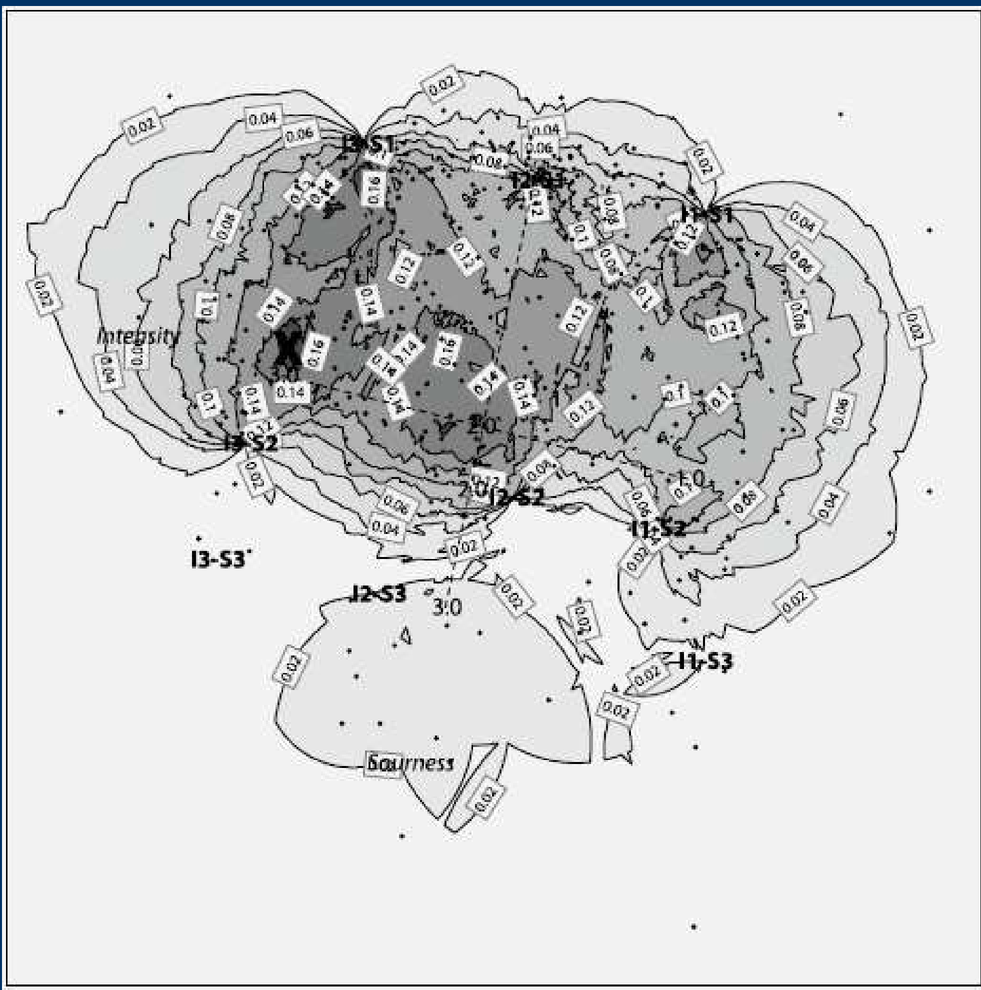
Triplot of Soup Rating Data (*soup locations restricted*)



⮤ Passive variables (fitted in later) ⮥

Finding the Optimal Product Facing Competition

Suppose we overlay the unfolding space with a fine grid, like pixels on a television screen. Each grid point is a potential product. We calculate grid values as the proportion of respondents with 1st choice for that potential soup. ***We are looking for the most dense region***.



X indicates location with highest grid value (16%)

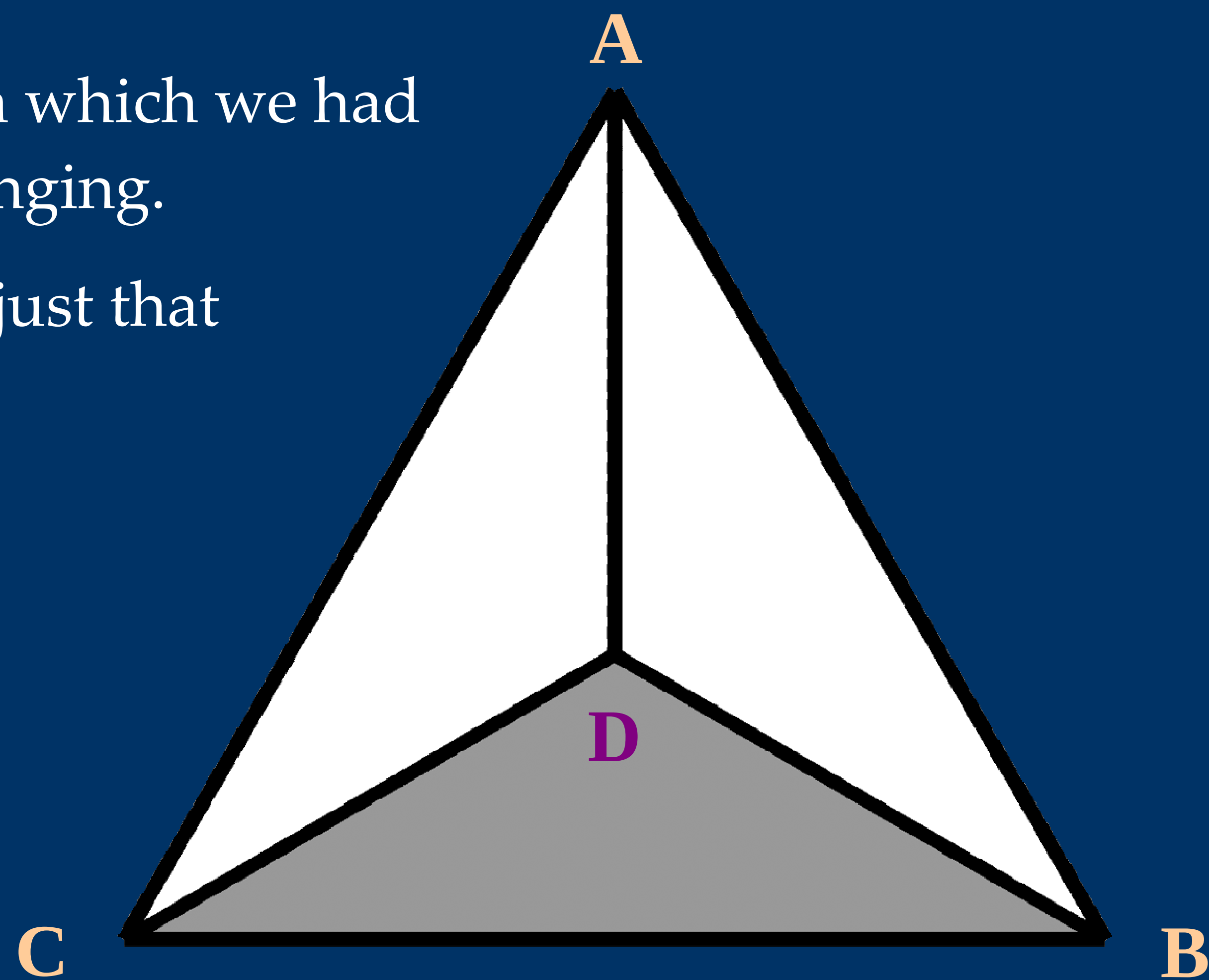
1. Now project this point on the lines for flavor intensity and sourness.
2. Back-transform them into the original scale of the explanatory variables.
3. **Result** is:
   Flavor intensity = 2.22
   Sourness = 1.66

Prediction of Choice: Sample Space is a Simplex

In multinomial data, every observation is located in one and only one of the corners of a simplex (2-dim simplex is a triangle, 3-dim simplex is a tetrahedron, etc.).

Recall this is the skeleton in which we had the ranking polyhedron hanging.

Not very revealing to map just that in a lower-dimensional space!



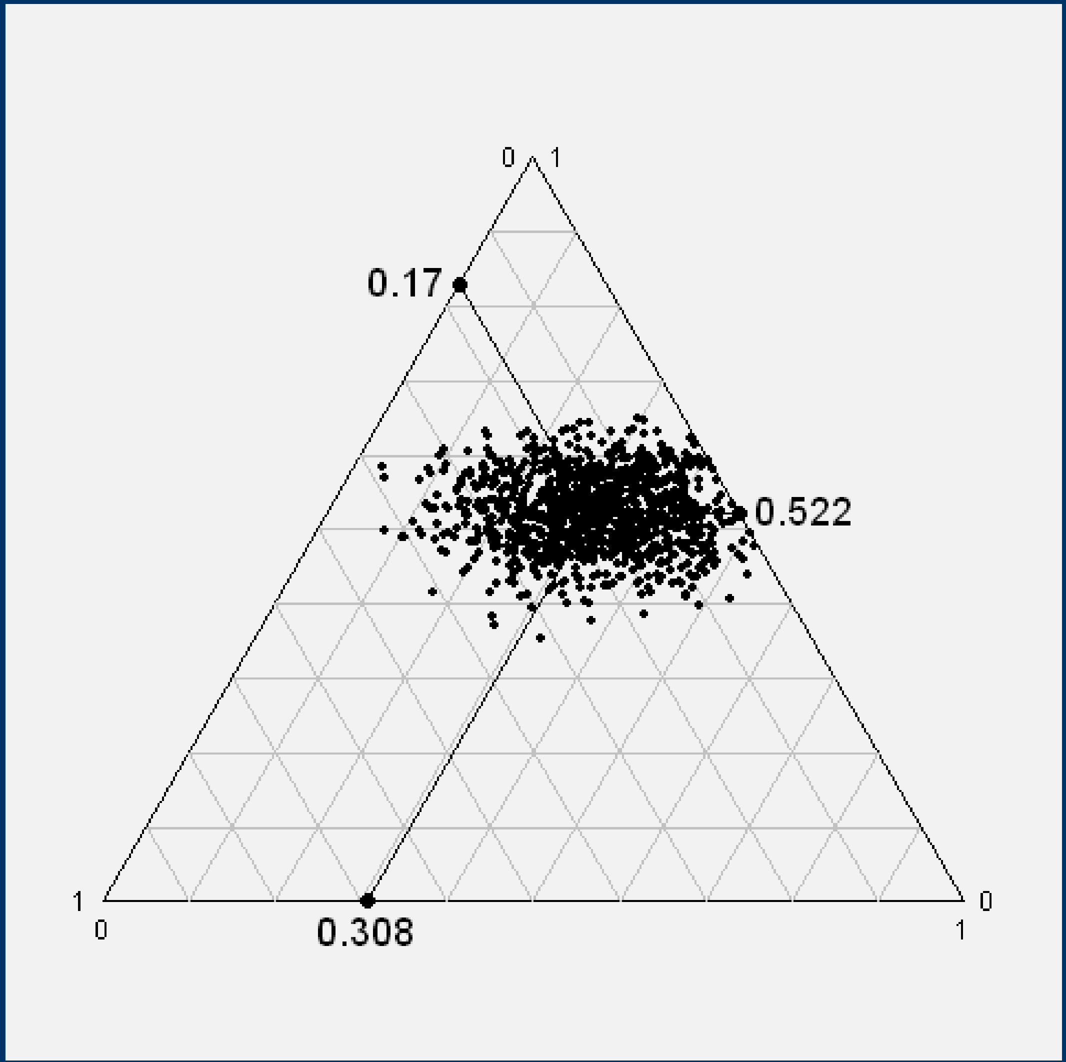
Multinomial Parameter Space

We are now going to use the full inside of the simplex. Probabilistic modeling implies an extra step: creating a parameter space for probability of choice.

On the right is the simplex for three categories.

Every point within the triangle is a different set of probabilities (p_A, p_B, p_C), positive numbers summing to one.

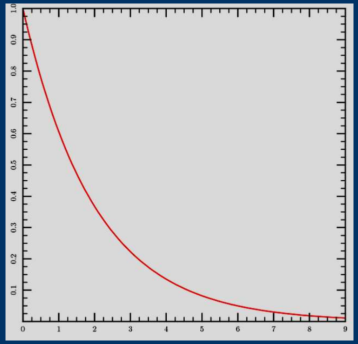
The triangle can also contain estimated probabilities, which in turn can be predicted by explanatory variables.



Probability is an instrument of the mind of the scientist!

Mixed Effect Ideal Point Model (De Rooij, *in prep.*)

First step is to link the probability π_{itc} that person i at time point t chooses choice option c to some distance-like quantity d^2_{itc}:



$$\pi_{itc} = \frac{\exp\left(-d^2_{itc}\right)}{\sum_l \exp\left(-d^2_{itl}\right)}$$

This link function is called the ***exponential decay*** function.

Next we use the reparametrization

$$d^2_{itc} = \sum_{m=1}^{M} \left(\eta_{itm} - \gamma_{cm}\right)^2$$

Where γ_{cm} is the location of the choice option point and η_{itm} is

$$\eta_{itm} = \mathbf{y}'_{it}\boldsymbol{\beta}_m + \mathbf{z}'_{it}\mathbf{u}_{im}$$

Here the ideal point is built up from fixed effects and random effects.

We assume MV normal distribution for the random effects $\mathbf{u}_{im}$.

Why Random ~~Parameters~~/Effects?

De Rooij's model is similar to that of Kamakura and Srivastava (1986), and others. Probabilistic unfolding was pioneered by Zinnes and Griggs (1974), Zinnes and MacKay (1987), MacKay and Zinnes (1995), but these models were for ratio judgments. For choice data pioneering work was done by De Sarbo and Hoffman (1986). In sensometrics, Daniel Ennis' work stands out (*cf*. Ennis and Mullen, 1992).

- We would like to model longitudinal choice data with person-specific models, so that we have a mechanism for the dependency among the responses. ***Going from person-specific parameters to random person effects avoids proliferation of parameters***;
- Random effects allow conclusions that can be generalized;
- It might be good idea to follow the Item Response Theory community in Psychometrics, who works with random effects since decades.

Example: TV Program Choices by Youngsters (Adachi, 2000)

First part of the data looks like this (N =100):

Table 1. The TV programme categories preferred by participants at time points $t = 1$ (first year of elementary school) to $t = 5$ (freshman year at university) and the frequencies of the series of preferred categories. Categories are abbreviated as follows: A = animation, C = cinema, D = drama, M = music, S = sport and V = variety

Males						Females					
$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	Freq.	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	Freq.
A	A	V	V	V	3	A	A	D	D	D	4
A	A	V	D	M	3	A	A	D	D	M	3
S	S	S	S	S	2	A	A	V	V	M	3
A	V	V	M	M	2	A	V	D	D	M	2
V	V	D	D	C	2	A	A	A	M	M	2
A	A	M	D	D	2	A	A	M	V	V	2
A	A	M	M	M	2	A	V	D	M	C	2
V	V	V	V	V	2	A	A	D	D	S	2
A	A	V	C	C	2	A	D	D	M	C	2
A	S	C	S	C	1	A	A	A	M	C	1
V	A	V	C	C	1	A	M	D	C	C	1
A	V	M	D	S	1	A	D	C	S	C	1
M	M	M	M	M	1	V	V	V	V	M	1
A	V	D	M	S	1	A	V	D	D	D	1

Estimation

- It is assumed that conditional upon the random effects the responses are independent (*cf*. ***local independence*** in IRT models given the person "parameter" θ).
- To obtain Maximum Likelihood estimates, we use ***marginal maximum likelihood estimation***; the likelihood can be approximated using Gauss-Hermite quadrature, where the integral is replaced by a weighted summation over a set of nodes.
- Prediction of the random effects can be done using ***expected a posteriori estimation***.

Model Selection

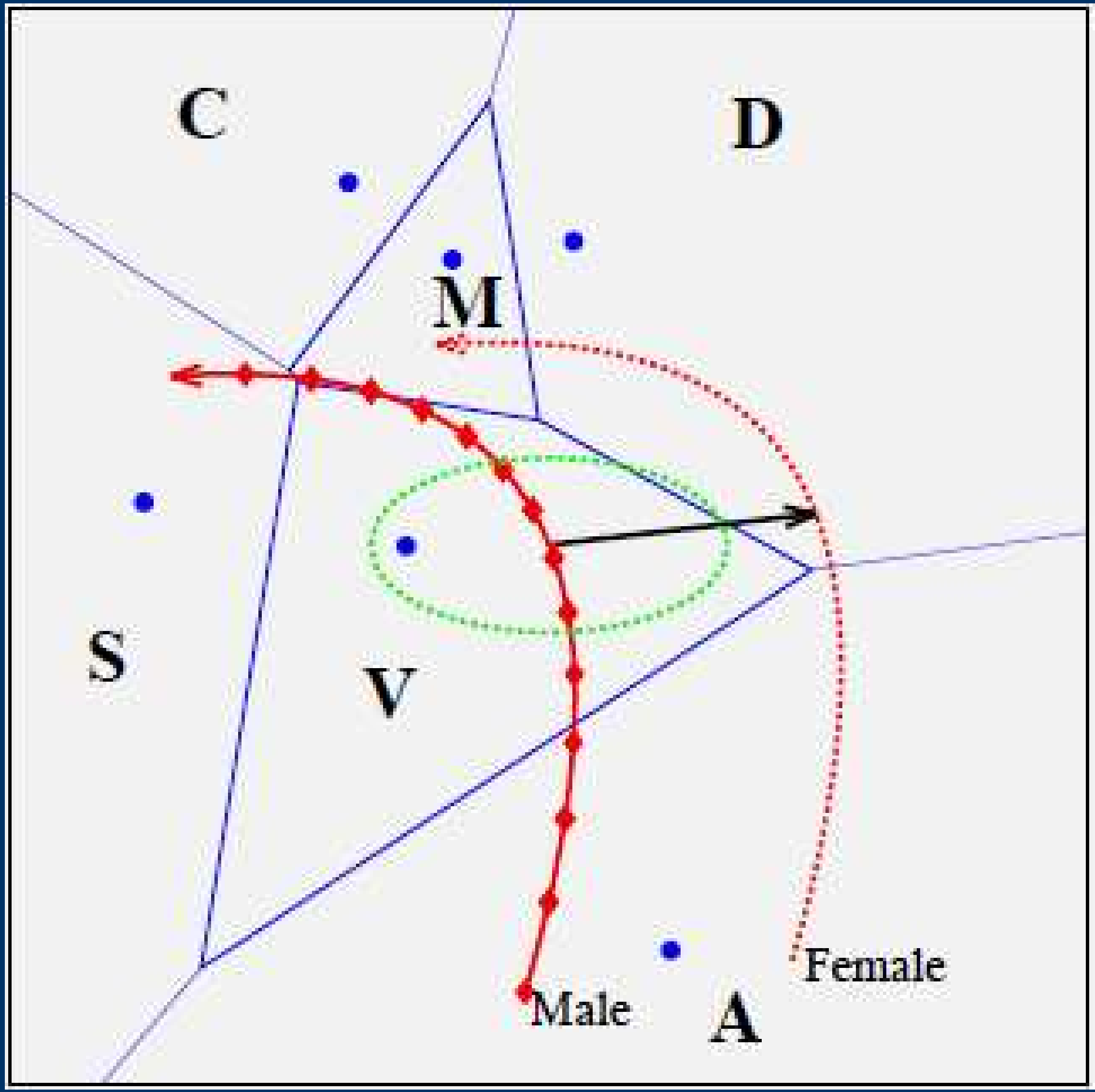
Fit statistics for several models are:

Dimensionality	Random	Fixed	npar	-2LL	BIC
2	I	$G+T$	18	1230.2	1313.1
		$G+T+T^2$	20	1212.2	**1304.3**
	$I+T$	$G+T$	22	1212.6	1313.9
		$G+T+T^2$	24	1197.7	1308.2
2	I	$G+T\|G$	20	1225.9	1318.0
		$G+T\|G+T^2\|G$	24	1207.7	1318.2
	$I+T$	$G+T\|G$	24	1202.9	1313.4
		$G+T\|G+T^2\|G$	28	1187.8	1316.7

- The second model has best BIC; it has a quadratic fixed *Time* effect in addition to the fixed *Gender* effect;
- *I* means a *random intercept* per dimension;
- *Random Time* effect would imply a different randomly chosen time function for each child around the fixed effect;
- $T|G$ indicates a different time function for males and females.

Solution TV Program Choice Data

A = *Animation*
C = *Cinema*
D = *Drama*
M = *Music*
S = *Sport*
V = *Variety*



All start with *Animation,* and have same trend except for different start.

Females tend more to *Drama,* males to *Sport*.

Age brings them from a preference for *Animation* and *Variety* to *Music, Cinema* and *Sport*.

Green ellipse is random intercept effect. Regions indicate first choice.

Concluding remarks

Sensometrics is the ideal domain for unfolding

Although unfolding ideas have been around for more than forty years, software development has been slow and difficult.

1. **PREFSCAL** is the first program that avoids *degeneracies* in ordinal (nonmetric) unfolding, using an effective penalty function. It can fit ***three-way models***, too.
2. *Prediction of preference* can be done in **PREFSCAL** by including predictor variables as constraints, either for options, or for actors, or for both.
3. *Prediction of choice* can be done by **GLMM modeling**

If you have applications for unfolding, you are most welcome to contact us for advice (mailto: **Heiser@Fsw.Leidenuniv.nl**, or **Busing@Fsw.Leidenuniv.nl**).