Explanatory Unfolding Models for Prediction of Judgment and Choice

Willem J. Heiser
Leiden University

With great help of Frank Busing & Mark de Rooij

(1) Two paradigms in single-peaked measurement: PRT and IRT

(2) Explanatory unfolding models

(3) Single-peaked PRT example: Taste and smell judgment data

(4) Single-peaked IRT example: Longitudinal multinomial choices

"Nature Has Placed Mankind Under the Governance of Two Sovereign Masters: Pain and Pleasure" (Bentham, 1789)

**People are driven by their interests and fears.
In the abundant world of today, they may even develop *Fear Of Missing Out* (FOMO), or *Choice Anxiety***

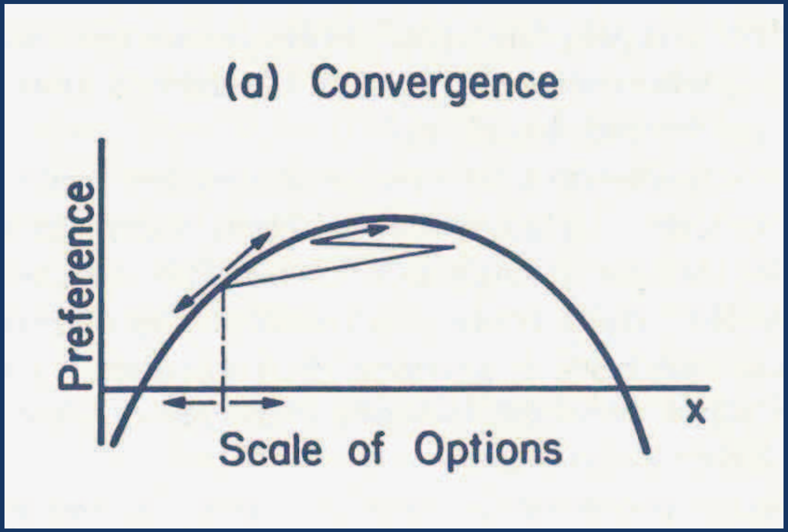
Dilemmas are the habitat of unfolding: How to measure and predict individual differences in preference & choice?

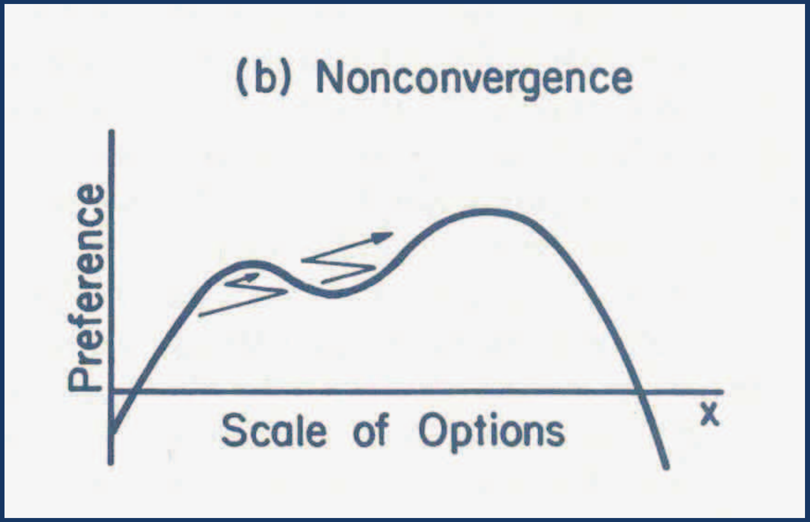
Active research fields are Marketing, Sensometrics & Political science

Individual Preference Tends to Be Single-Peaked

According to Coombs' (1977, 1988) theory of conflict, *interesting* or *relevant* preferences arise from different trade-offs made in conflicts between incompatible goals. Under those circumstances, individual preference may often be described by a single-peaked function (SPF).

SPFs allow a simple search process converging to optimal choice.





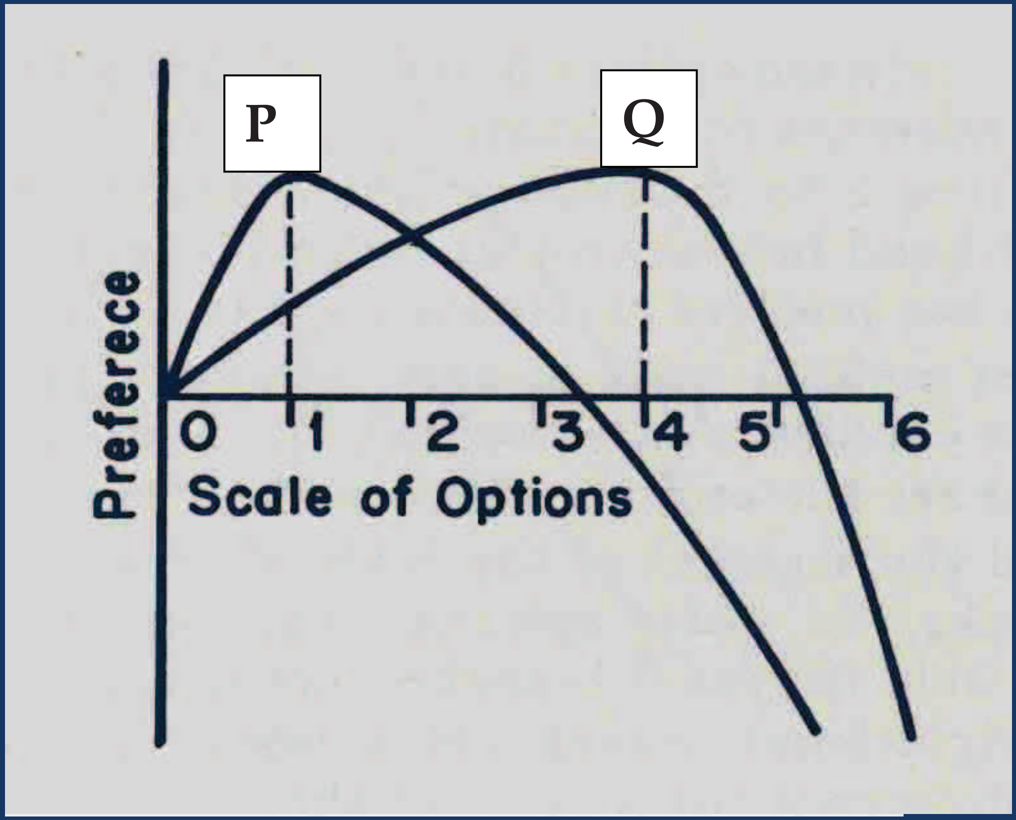
For instance, when you are under the shower, it is relatively easy to manipulate *temperature* as to reach your personal optimal value.

Modeling Heterogeneity of Preference and Choice

Apart from ***intra-personal*** conflict, single-peaked preference also clarifies ***inter-personal*** conflict, in which individuals want different things but must settle for the same thing.

Consider a family conflict over ***number of children***:

P's preference for number of children is (1, 2, 3, 0; 4, 5, 6), while **Q**'s preference is (4, 3, 2, 1, 5, 0; 6).

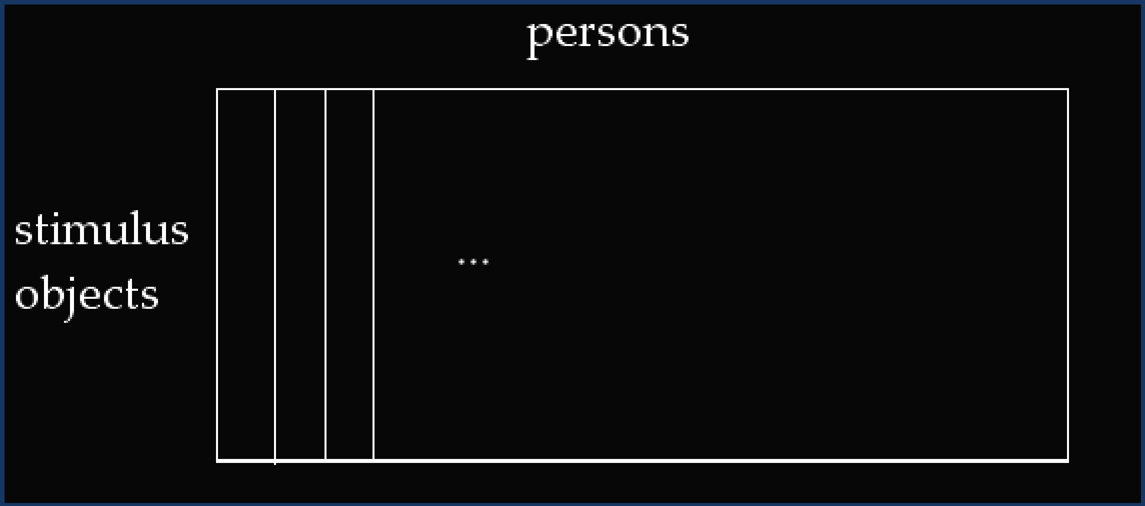


☞ Individual rank orders can be obtained by stretching the scale of options and *folding* it at the *ideal point*.

Single-Peaked Measurement in Person Response Theory (PRT)

In PRT, a person makes the distinction between stimuli or items.

- Single-peaked ***person curves*** (location parameter = ideal point);
- ***Stimulus*** gets a ***scale value***, which usually is a ***fixed effect***;
- Data are ***person conditional*** (persons attribute values to objects), and typically have this shape:



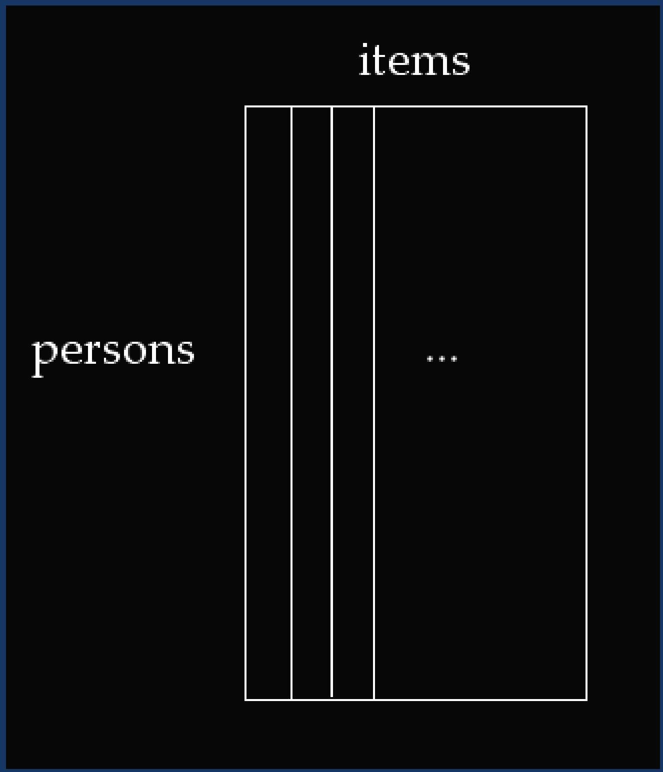
- Analysis tends to be ***multi-dimensional***.

Coombs (1950, 1964), Tucker (1960), Roskam (1968), Kruskal & Carroll (1969), Davison (1977), Heiser (1981, 1989), DeSarbo & Rao (1984), MacKay (1995), Kim, Rangaswamy & DeSarbo (1999), Busing, Groenen & Heiser (2005), Van Deun et al…

Single-Peaked Measurement in Item Response Theory (IRT)

In IRT, an item makes the distinction between individuals or persons.

- Single-peaked ***item curves*** (location parameter = item difficulty);
- ***Person*** gets a scale value or ***score***, usually a ***random effect***;
- Data are ***item conditional*** (items attribute values to persons), and typically have this shape:



- Analysis tends to be ***one-dimensional***.

Van Schuur (1984), Andrich (1988), Hoijtink (1990), Takane (1983, 1998), DeSarbo & Hoffman (1986), DeSarbo & Cho (1989), Roberts (2001, 2008), Roberts & Thompson (2011), Johnson & Junker (2003), Maydeu et al…

Explanatory Unfolding Models

Explanatory models try to predict an ***outcome variable***, in this case judgment or choice data.

Prediction can be based on independent

- Person attributes (both PRT and IRT);
- (designed or measured) stimulus attributes (PRT);
- Item design factors (IRT).

The key concept is to combine a measurement model with a regression model (as pioneered by Karl Jöreskog and Gerhard Fischer), and the key tool is ***reparametrization***.

De Boeck, P. & Wilson, M. (2004), Explanatory Item Response Models. New York: Springer.

Why is Growing Interest in Prediction to be Expected?

- *Measurement is not enough*. It is always driven by a reason: 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 applied fields, e.g. 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.
- …

Crossing the Aspects Measurement/Prediction and PRT/IRT

Some of the proposed software is listed here (not complete…).

	PRT	IRT
Measurement	GENFOLD-1 (DeSarbo) PREFSCAL-1 (Busing et al. in **IBM/SPSS**)	Unidimensional: Roberts' GGUM Multidimensional: MacKay's PROSCAL, Maydeu's NOHARM
Prediction	GENFOLD-2 CANOCO (Ter Braak) PREFSCAL-2 (Busing)	Ordered responses: Böckenholt (2001) Multinomial responses: De Rooij (2009)

Prediction of Judgment in the PRT Paradigm Through Reparametrization of an Unfolding Model

All unfolding models have location parameters for persons (**X**) and location parameters for stimuli (**V**). When $\mathbf{Q}_X$ and $\mathbf{Q}_V$ are the matrices of explanatory attributes for persons and stimuli, respectively, we add the restrictions:

$$\mathbf{X} = \mathbf{Q}_X \mathbf{B}_X,$$
$$\mathbf{V} = \mathbf{Q}_V \mathbf{B}_V.$$

Depending on specific model and optimization method used, usually estimation of regression weights $\mathbf{B}_X$ and $\mathbf{B}_V$ is reasonably standard. In PREFSCAL-2, we use the projected gradient method and Alternating Least Squares for the restrictions (Busing, Heiser & Cleaver, 2010).

Ter Braak's CANOCO program uses a restricted version of correspondence analysis, is limited to binary and numerical data.

PREFSCAL (in IBM/SPSS) 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

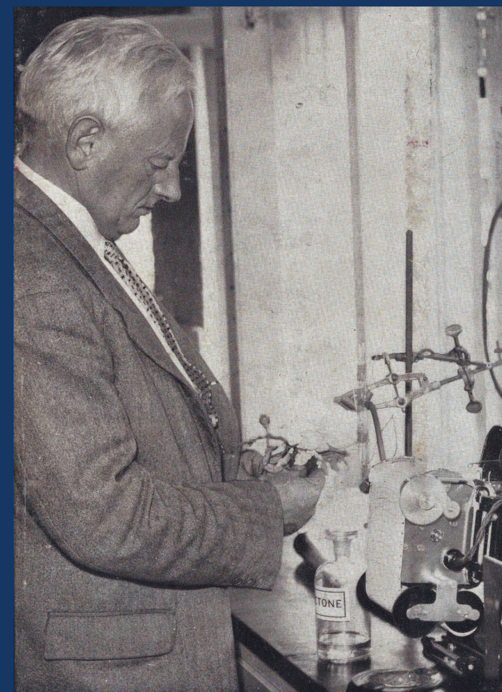
$$\min_{\Gamma,\mathbf{X},\mathbf{V}} \left\{ \sum_i \frac{\|\gamma_i(\mathbf{P}) - d_i(\mathbf{X},\mathbf{V})\|^2}{\|\gamma_i(\mathbf{P})\|^2} \right\}^{\lambda} \left\{ \frac{1}{I} \sum_i \left(1 + \frac{\omega}{\upsilon^2[\gamma_i(\mathbf{P})]} \right) \right\}.$$

Stress term *Penalty term*

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

Example 1: 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:

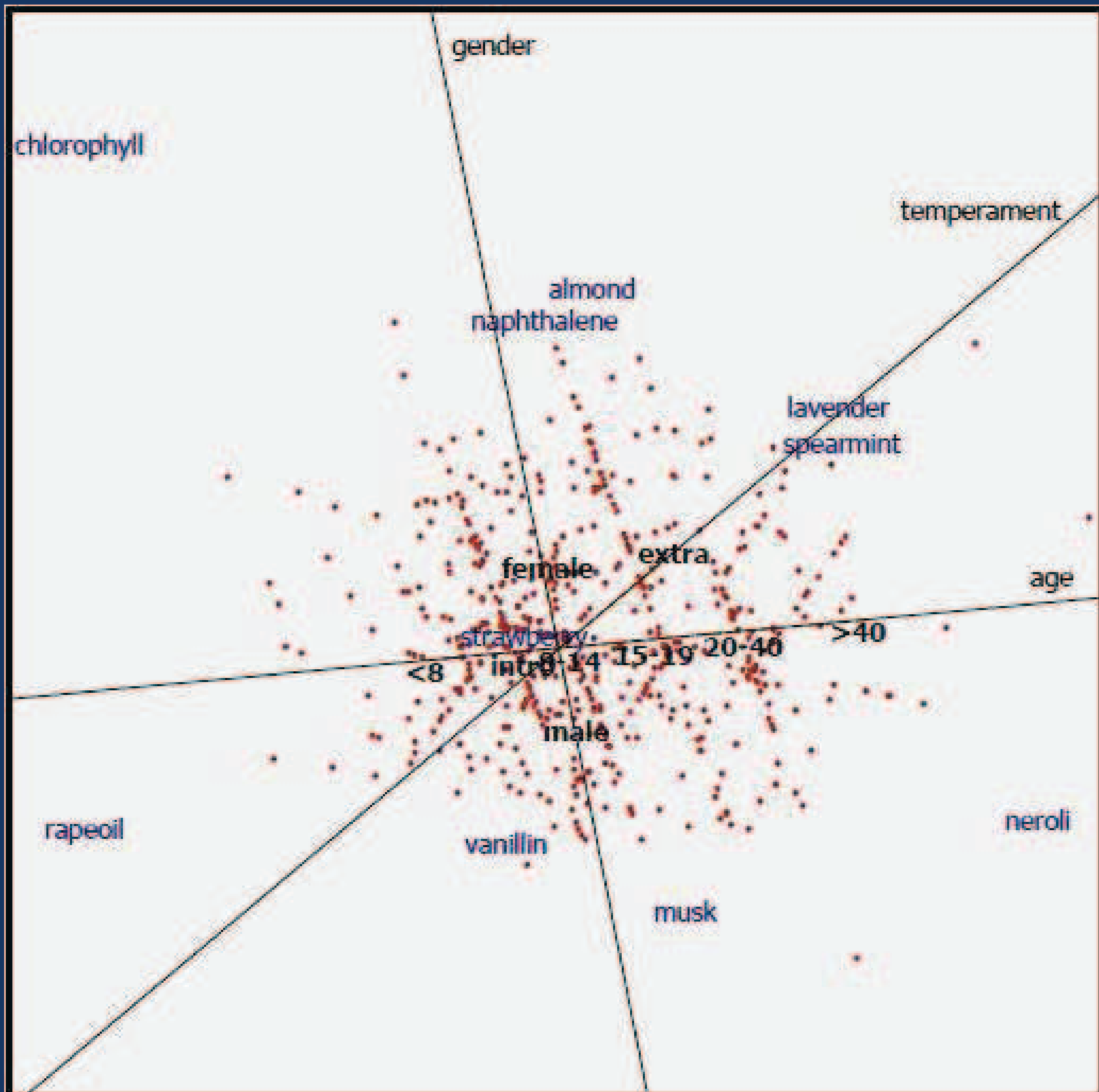


1. **Strawberry** flavoring essence;
2. **Spearmint** oil;
3. French **lavender** oil (with high ester content);
4. **Musk** lactone (100%);
5. **Vanillin** (essential odorant of vanilla pod);
6. **Neroli** oil;
7. **Almond** flavoring essence;
8. **Naphthalene** (smell of moth-balls & fire-lighters);
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.

Triplot of Restricted PREFSCAL on Odour Preferences



VAF of explanatory attributes:

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

Example 2: Soup Rating Data (Busing, Heiser & Cleaver, 2010)

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

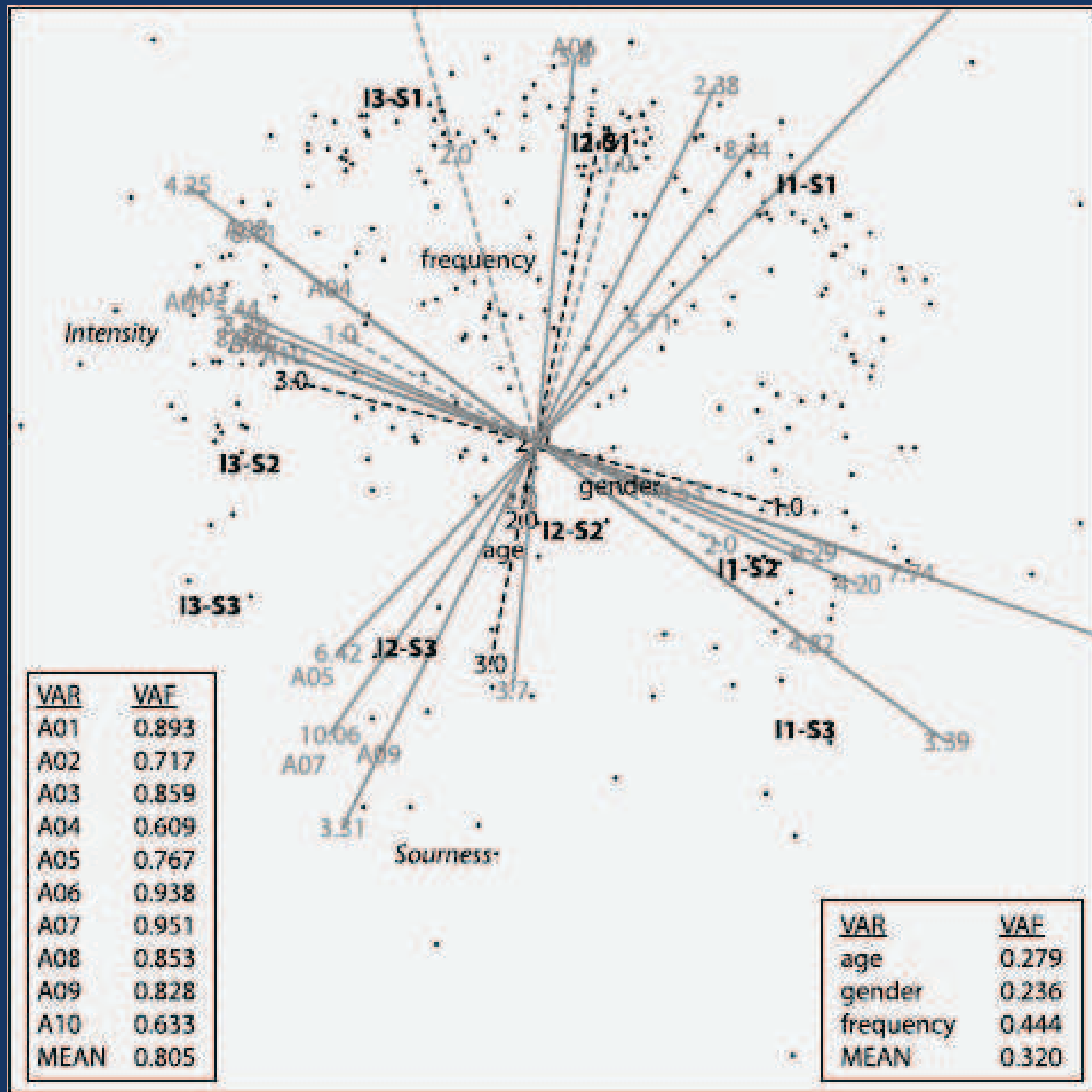
Fixed effects	Numerator DF	Denominator DF	F	Sig.
Flavour Intensity	2	594	5.25	.005
Sourness	2	594	57.40	.000
Flavour Intensity × Sourness	4	1188	2.41	.048
Random effects	**Estimate**	**Standard Error**	**Wald Z**	**Sig.**
Respondents	0.94	0.12	7.50	.000
Respondents × Flavour Intensity	0.50	0.08	5.99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

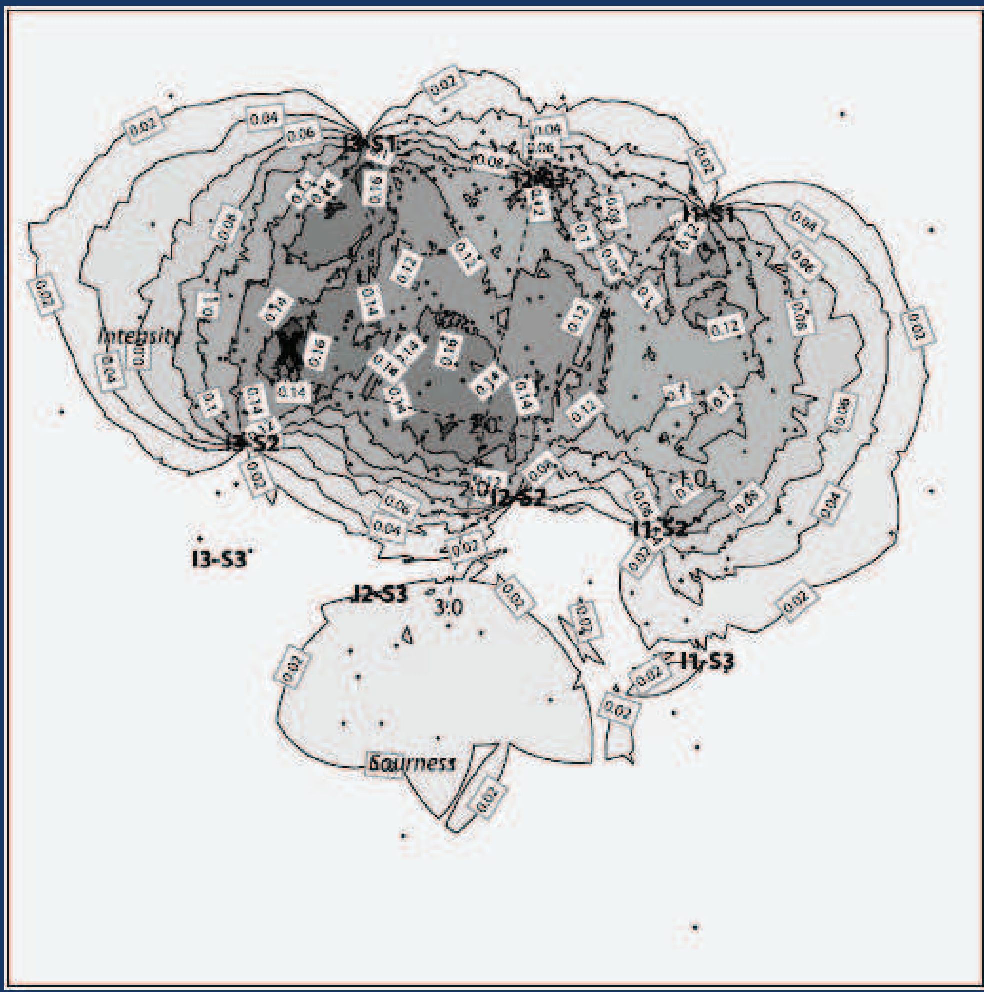
Triplot of Soup Rating Data (*soup locations restricted*)



↰ Passive attributes (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%)

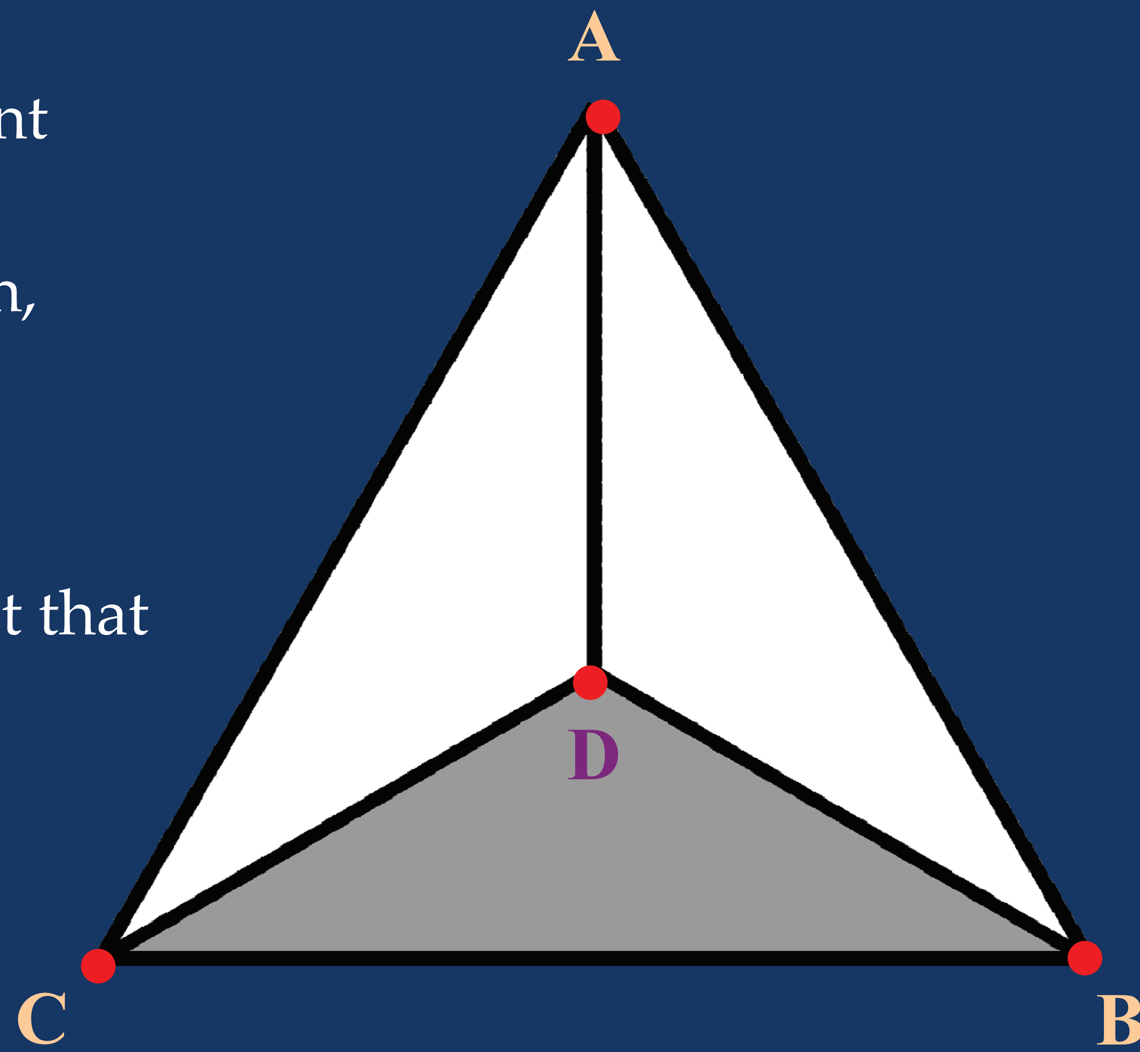
1. Now project this point on the lines for flavor intensity and sourness.
2. Back-transform them into the original scale of the explanatory attributes.
3. **Result** is:
   Flavor intensity = 2.22
   Sourness = 1.66

Prediction of Choice in the IRT Paradigm

In multinomial data, every observation is located in one and only one of the corners of a simplex

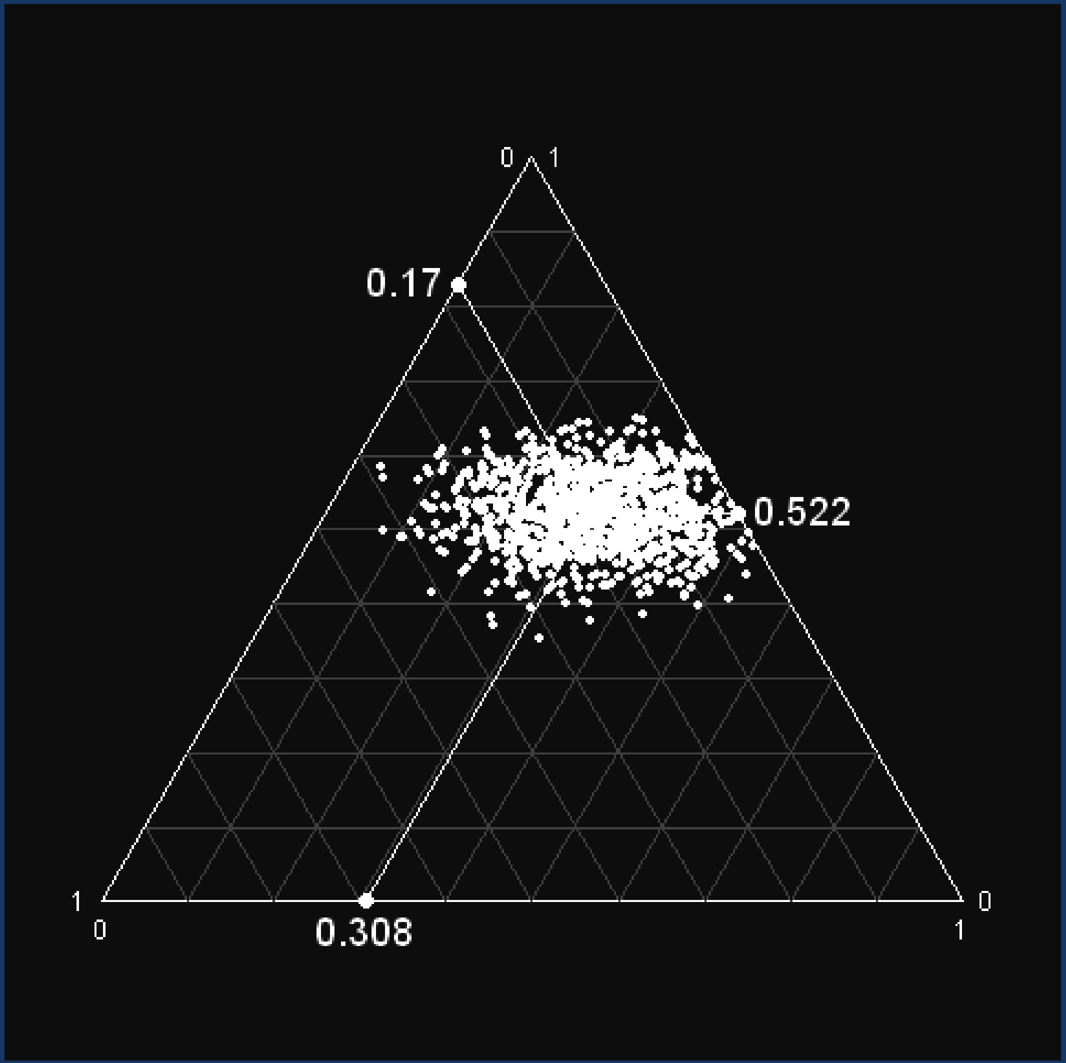
1-dim simplex is a line segment
2-dim simplex is a triangle,
3-dim simplex is a tetrahedron,
and so on.

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



Creating a Latent Space: Prediction of Expected Choice

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.

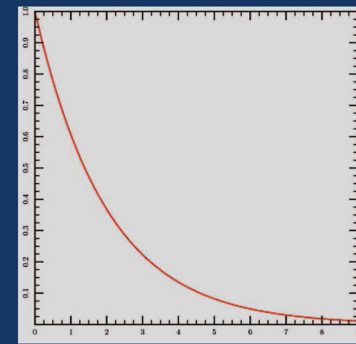
Every point within the triangle is a different set of probabilities (p_A, p_B, p_C), positive numbers summing to one.

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

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; model goes back to Shepard (1957, 1987). Next we use the reparametrization

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

Here γ_{cm} is location of choice option point, and η_{itm} is reparametrized as

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

So the ideal point is built up from fixed effects and random effects. De Rooij assumes MV normal distribution for the random effects $\mathbf{u}_{im}$.

Why Random Effects?

- 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 tested and generalized;

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) and and for ranking data by Böckenholt (2001).

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

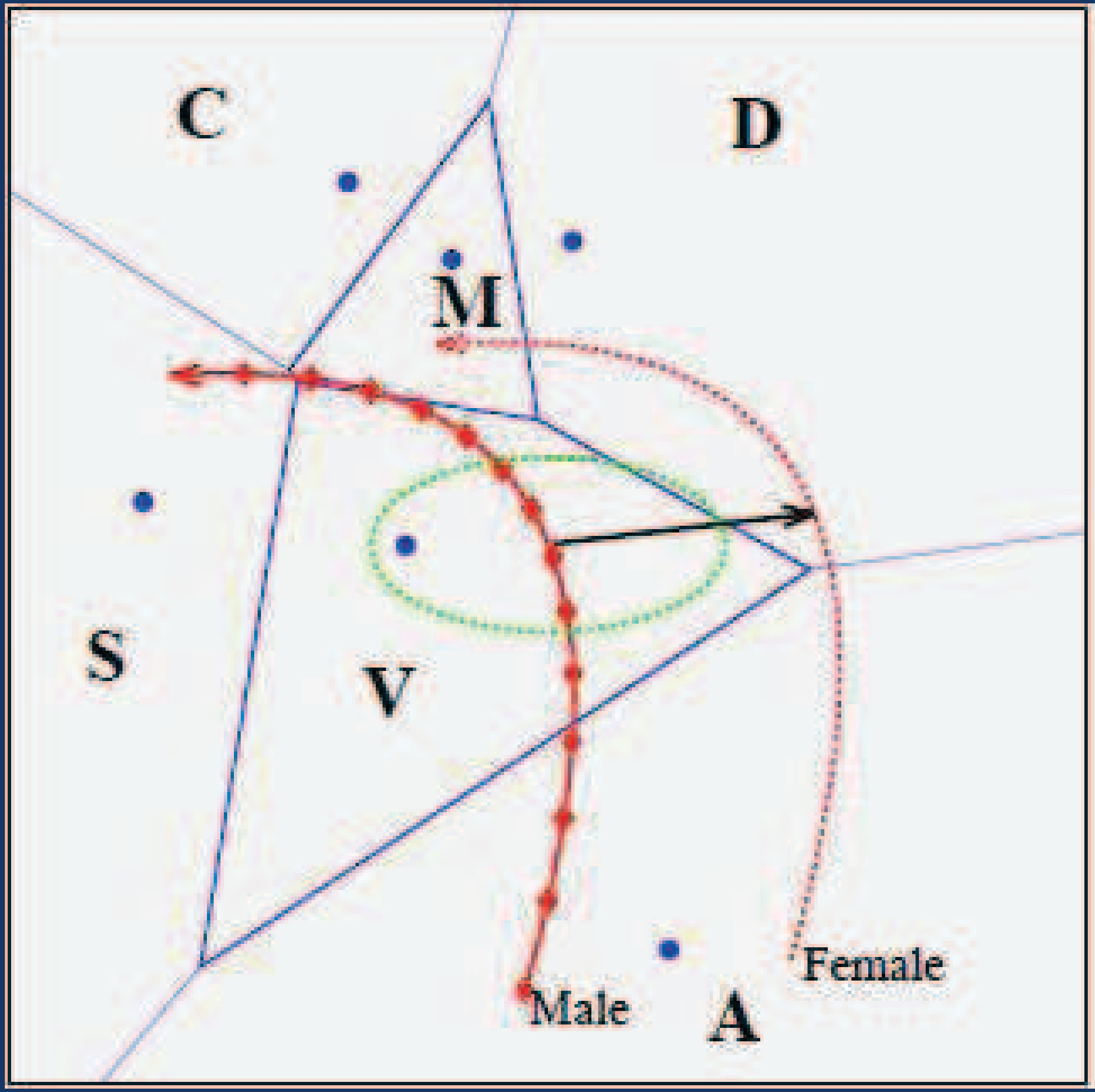
Fit statistics for several models are:

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;
- $T \mid G$ 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.

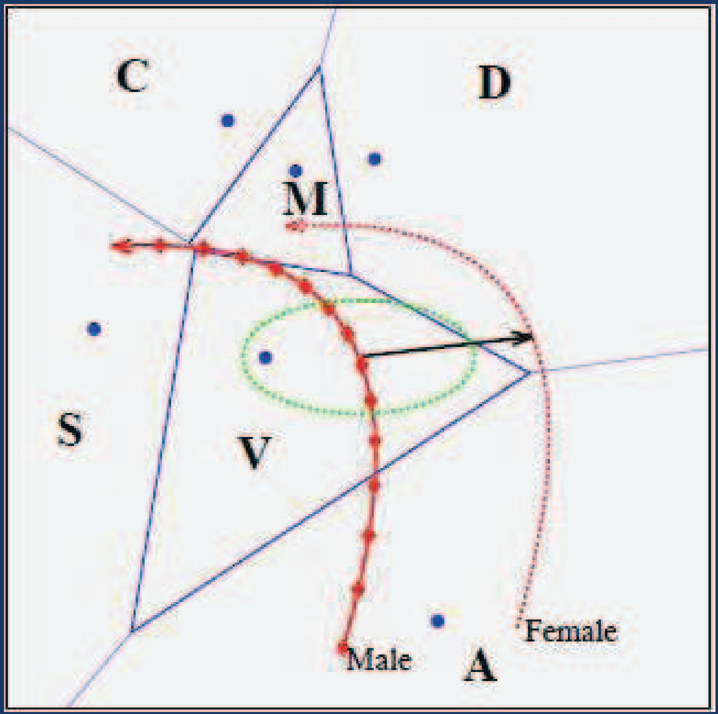
Females tend more to *Drama*, males to *Sport*.

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.

Conclusion 1

The mixed model unfolding plot combines ideas from at least:



- Coombs: ideal points;
- Shepard: probability = inverse of distance
- Kruskal: slide vector for asymmetry
- Carroll, De Soete, DeSarbo: wandering ideal points
- Takane: ideal point discriminant anal.
- Heiser, Zielman, Adachi, De Rooij: slide vectors for change
- Böckenholt: mixed ideal point model

Some more Conclusions

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

1. **PREFSCAL** is first program that avoids ***degeneracies*** in ordinal (nonmetric) unfolding, using an effective penalty function (available via IBM/SPSS). It can fit ***three-way models***, too.
2. Move from pure measurement to (context dependent) prediction will determine the research agenda for next decade.
3. Single-peaked IRT did not really begin with this move & could learn from PRT.
4. PRT (the multidimensional scaling world) did not really take very seriously the need to work with random effects and mixed models & could learn a lot from IRT (and biostatistics!).

If you are interested in using **PREFSCAL**, you are most welcome to contact us for advice (mailto: **Heiser@Fsw.Leidenuniv.nl**, or **Busing@Fsw.Leidenuniv.nl**).