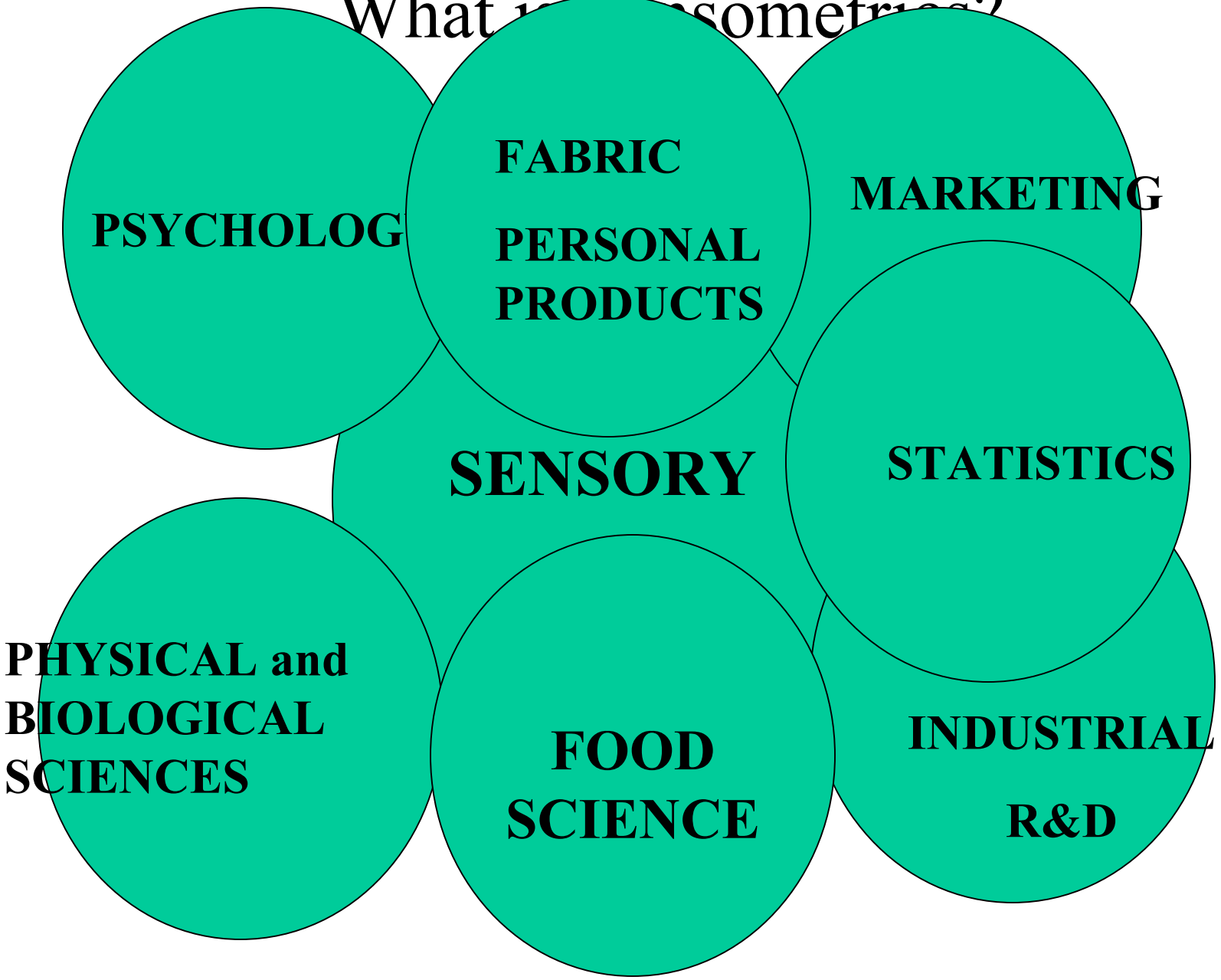
Sensometrics: What have we achieved? Where are we going?

Hal MacFie

Overview

- Achievements
  - Other peoples views
  - Sensometrics Society
  - Panel performance
  - Relating consumer to sensory
  - Shelf life and TI
  - Metrics
  - Role of sensory in consumer decision making and enjoyment

What is Sensometrics?



Current position (Popper)

- Need to get existing methods integrated into practice.
- lack of software, and lack of best practices.
- tendency for the field to offer up new methods where several already exist or to offer twists on existing methods
- little incentive to try to assess the merits of the techniques relative to one another and what their real contribution is to the bottom line for the industrial application -- namely, how much do my recommendations change depending on the approach I use.

What have we achieved?

- Comments from colleagues - Wakeling
  - Increasing sophistication of users particularly in relation to the use of Multivariate analysis,
  - Surprising that sensory people can conduct and interpret a PCA without any difficulties but struggle with a multiple regression.
  - Unnecessary sophistication – our data is not that complex in relation to other fields

What have we achieved?

- Comments from colleagues Ann Noble
  - communication between statisticians and sensory folks
  - accessibility of programs for non-stat person to use validly both resulting in use of increasingly complex multivariate programs for not only analyzing sensory data but relating it to a host of other data sets (chemical analyses, design variables etc etc)

What have we achieved?

- Comments from colleagues Qannari
- Regarding the past :
  - I'd stress the reciprocal stimulation and the cross-fertilisation of sensory analysis and multivariate analysis in general.
  - For instance, GPA contributed a lot in the development of free choice profiling and, on the other hand, sensory analysis has contributed a lot in making GPA very popular.

Conferences

- 7 Conferences now held on a biennial basis
- 100-200 attendees per conference
- Papers that pass the refereeing process published in Food Quality and Preference

Classification of papers in Sensometric Special issues

Topic	No of Papers
Sensory Panel Performance	15
Relating consumer to sensory	13
Difference tests	7
Preference Testing	6
PCA/PLS	6
Design	5
Relating sensory to instrumental	4
Multivariate Testing	3
Free choice	3
Segmentation	3
Neural Nets	3
Quality	2
Choice/ Attitudes	2
Dispersion Tests	1
Multidimensional Scaling	1
Time-Intensity	1

Assessing Sensory Panel Performance

Sensory Panel performance

- Repeatability
  - The ability to score the same product consistently for a given attribute (Rossi 2001 quoting NB Standards)
- Reproducibility
  - The ability to score products the same, on average, as the other panel members (Rossi 2001)
- Discrimination
  - The ability to show a significant signal to noise ratio

Repeatability

- Rossi (2001)
  - Showed how effective simple graphical tools can be
  - Here we see immediately see that there are some problem assessors with regard to repeatability

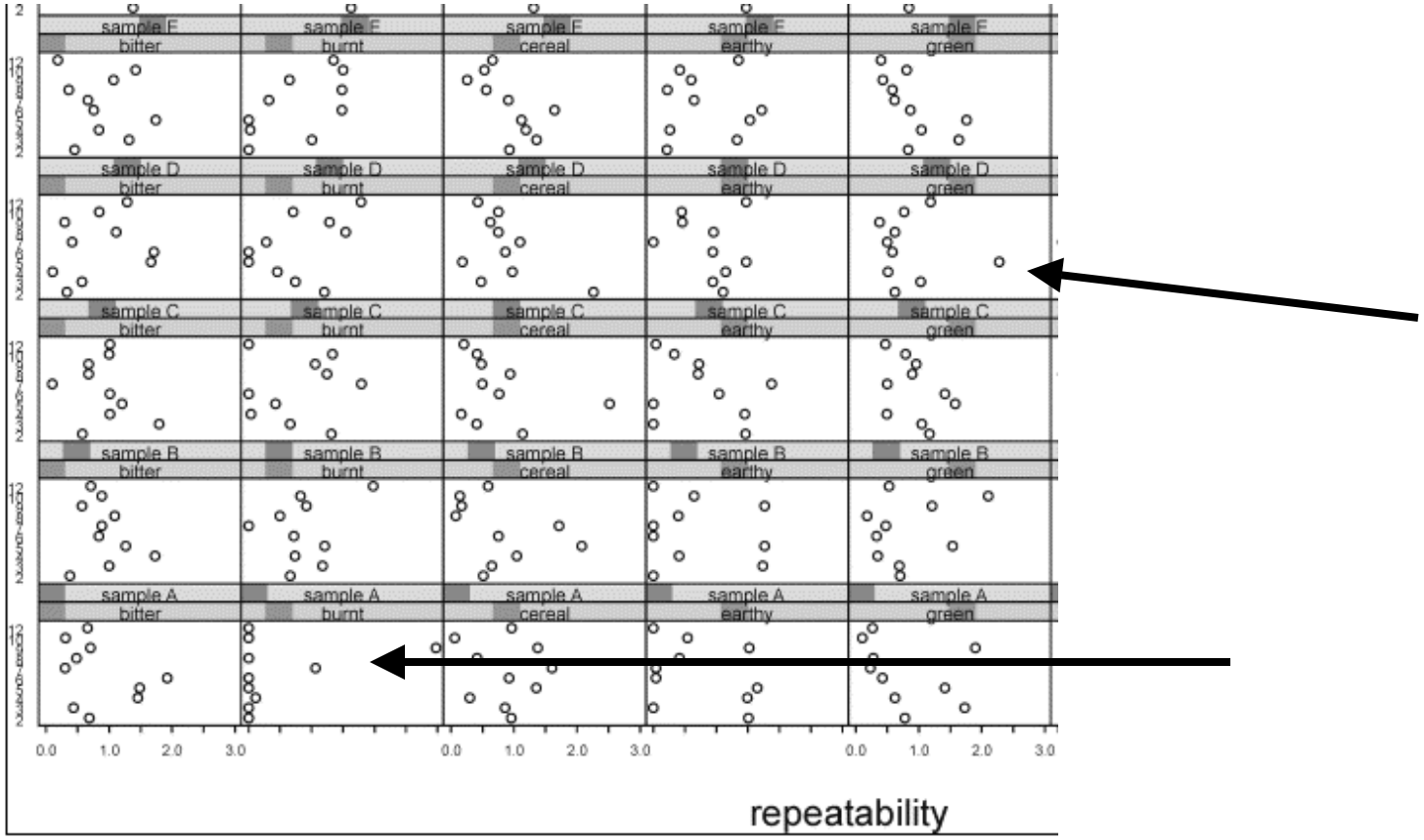
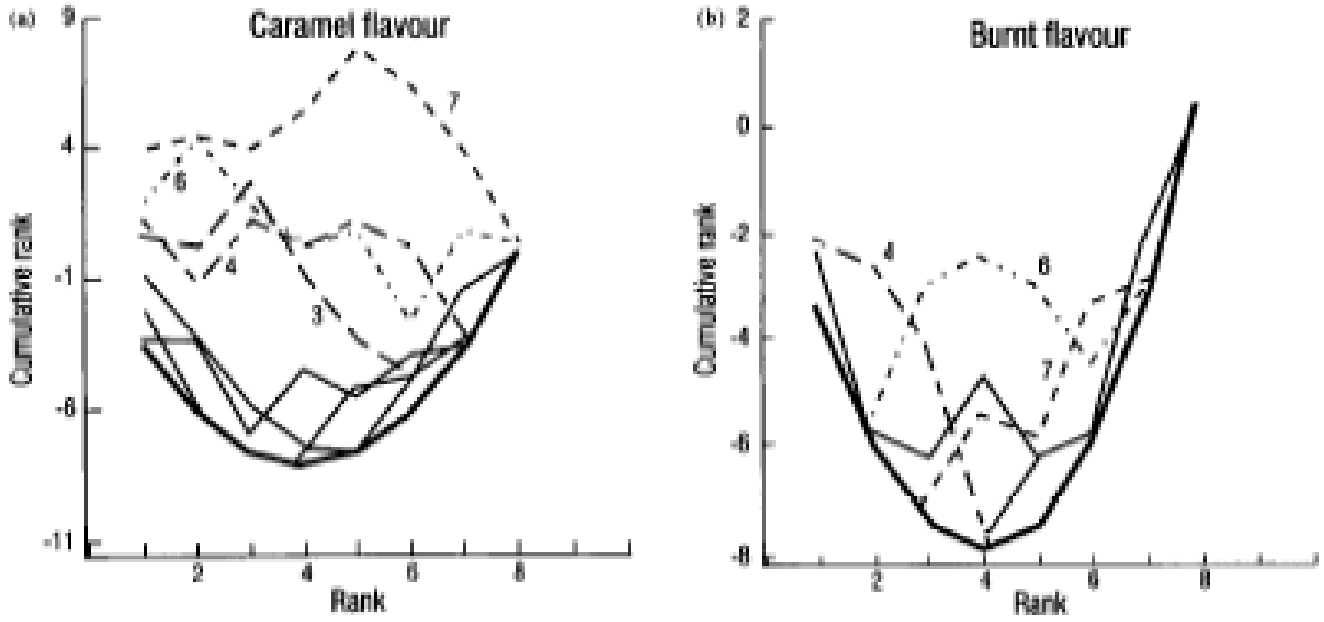


Fig. 5. Individual repeatability measures for panel A

Reproducibility

- Tragon (1960's) - method to partition the assessor by sample interaction
- Naes- eggshell plots (1998)



Reproducibility

- Dijksterhuis (1995)
  - Applied PCA to each assessor on each attribute separately and compared loadings on the assumption that first PCA was signal and the others were noise

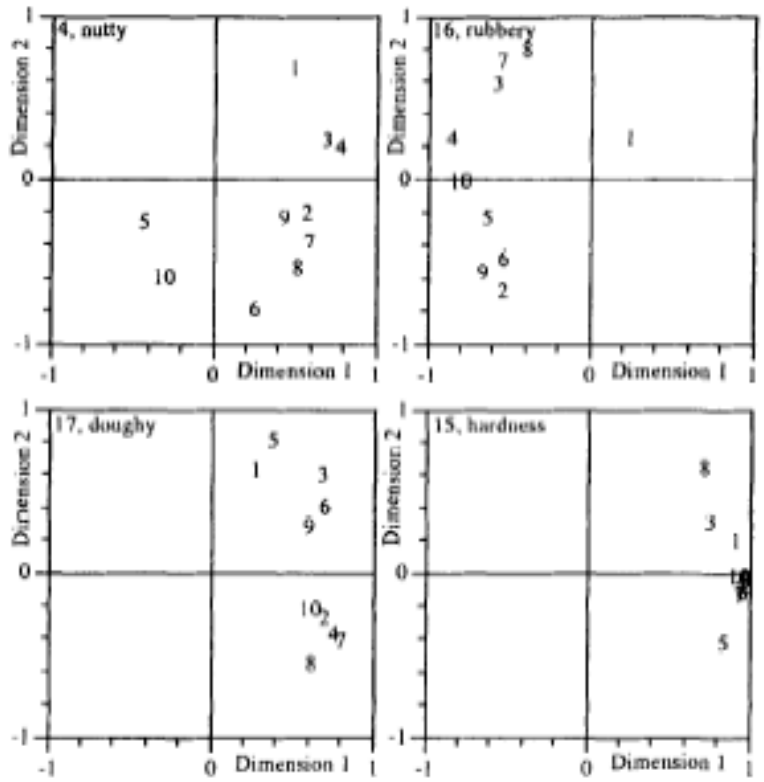


FIG. 4. Plot of the assessors from the PCAs of the most unidimensional attribute (no. 15) and the three least unidimensional attributes (nos 4, 16 17) in the cheese data set.

Reproducibility

- Couronne (1997)
  - Removed assessor and sample effects and did a PCA on the matrix with assessors as variables
  - Here we see two groups of assessors who are almost zero correlated

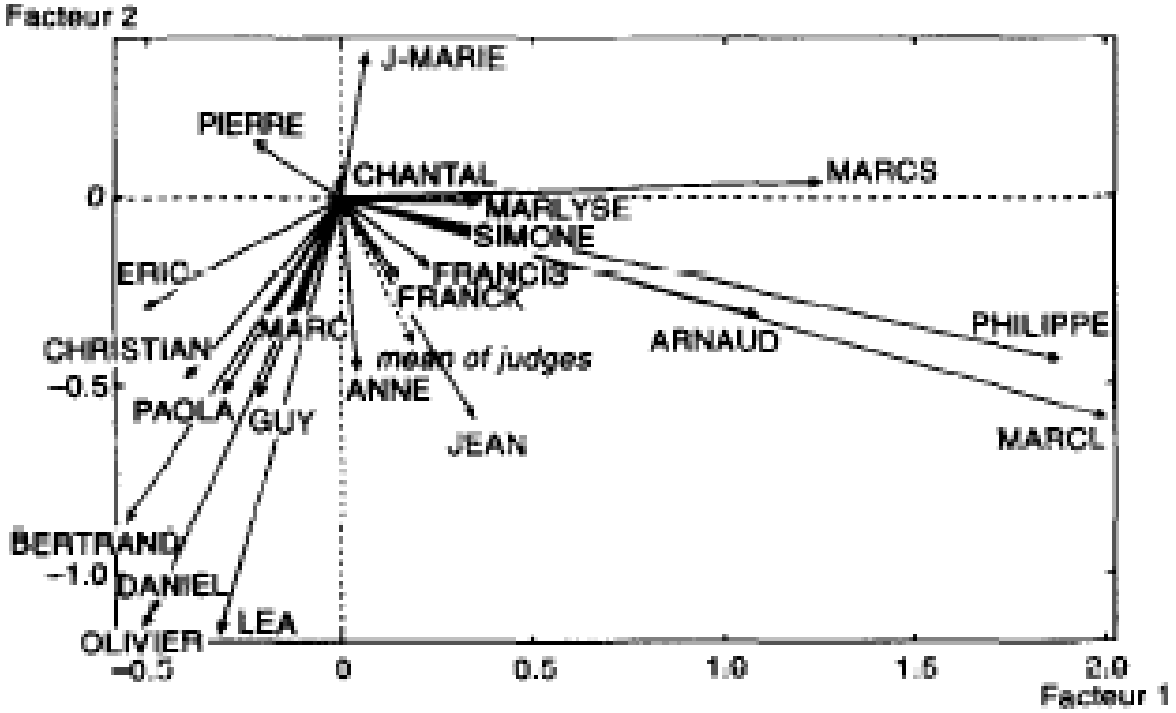


FIG. 5. Factorial plane 1–2 with the assessors as variables.

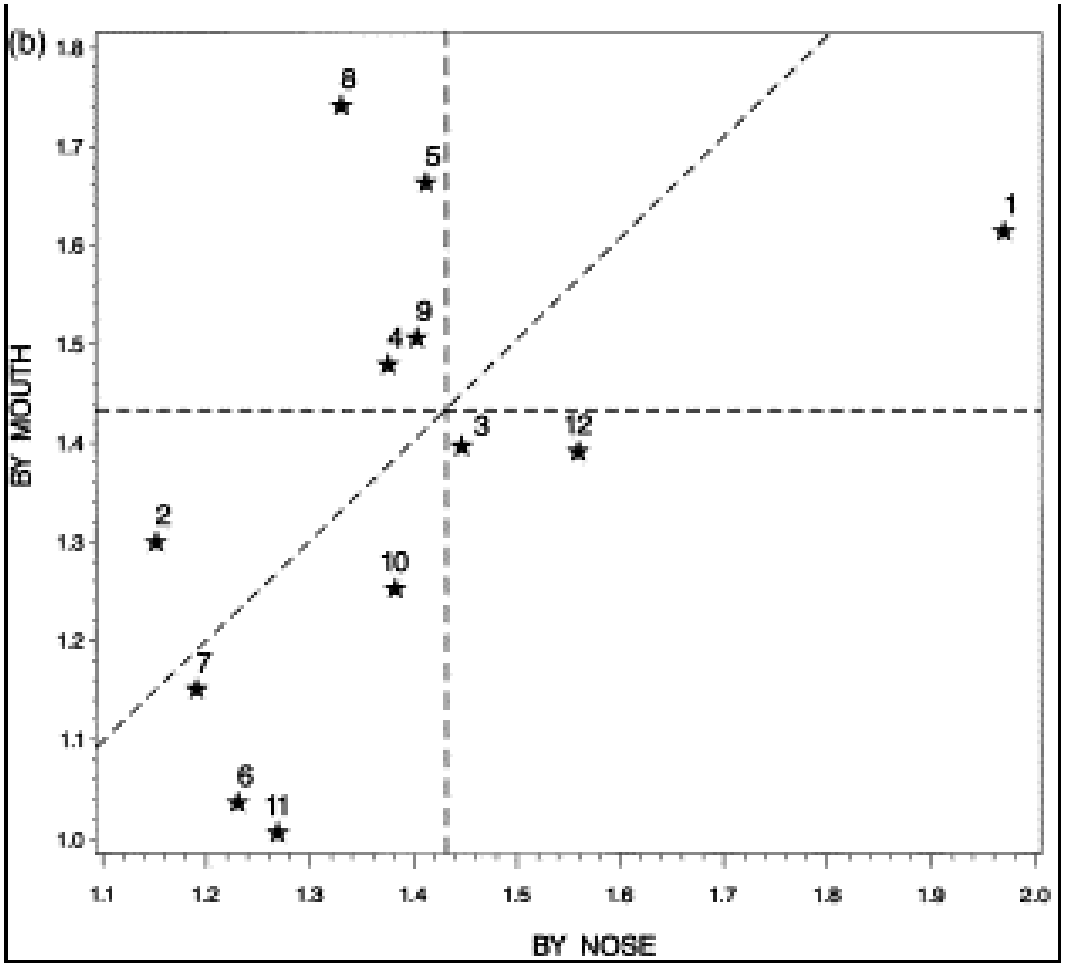
Reproducibility

- Qannari, MacFie and Coucoux (1998)
  - Used isotropic scaling and Procrustes to define performance indices that characterised how similar assessors were performing
  - Here Assessors F and G were found to be performing differently to the rest of the panel.

Table 1. Performance Indices and Isotropic Scaling Factors derived from Gower's GPA Algorithm and from the Alternative Algorithm

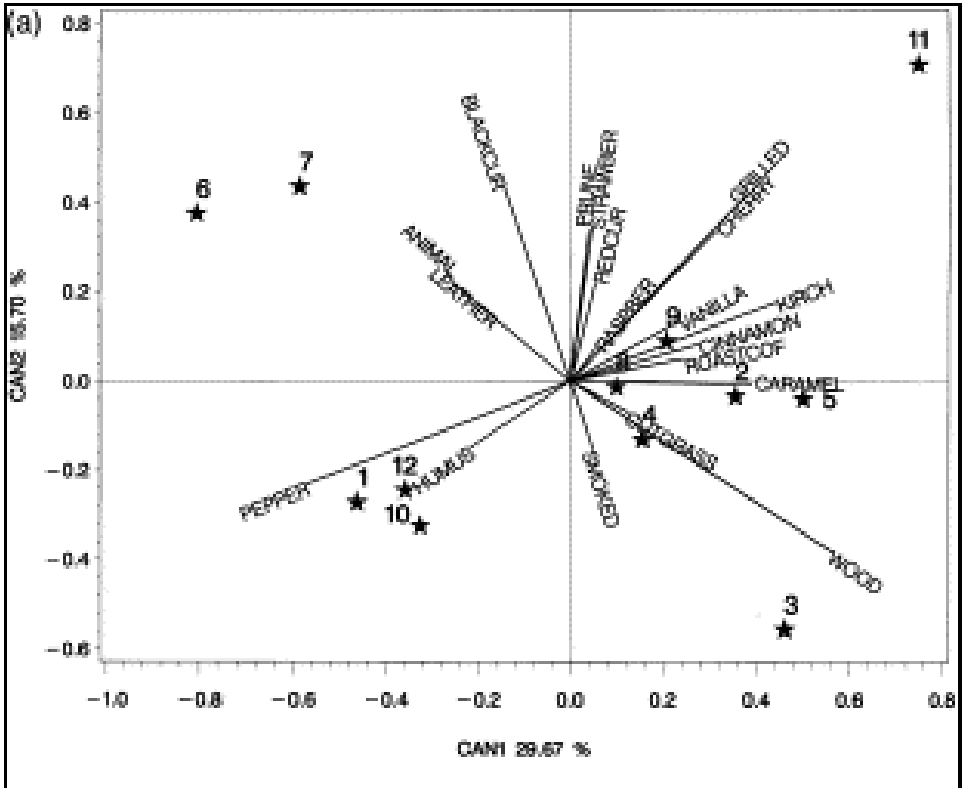
Assessor	Performance indices	Isotropic scaling factors	
		Alternative algorithm	GPA
A	0.383	1.416	1.412
B	0.392	1.032	1.036
C	0.395	0.774	0.776
D	0.382	1.231	1.225
E	0.380	0.992	0.992
F	0.355	0.929	0.926
G	0.357	0.983	0.987

Discrimination



Comparison of 1 way F ratios for assessment of wines by nose and by mouth (Aubrey, Schlich, Issanchou and Etievant 1999)

Reproducibility and Discrimination



Using two way MANOVA and CVA to compare panel performance by nose and by mouth. Thus, panelists 1, 10 and 12 seem responsible for the significant difference in the two modes observed for pepper and humus descriptors; panelist 3, for wood descriptor; panelist 11, for grilled/burned descriptor; panelists 2 and 5, for caramel and kirsch/cherry stone descriptors.

(Aubrey, Schlich, Issanchou and Etievant 1999)

Multivariate panel performances based on MANOVA and CVA

	Univariate	Multivariate
Panelist	**ANOVA model: Prod** ➢ Repeatability: $\sqrt{MS_E}$ (Root mean square of error) ➢ Validity: correlation with panel mean ➢ Discrimination: $F = MS_{Prod} / MS_E$	**MANOVA model: Prod** ➢ Discrimination: MANOVA test of Prod **Canonical Variate Analysis (CVA)** ➢ Dimensionality of product configuration **Stepwise selection of attributes** ➢ Saliency of product differences
Panel	**ANOVA model: Prod + Pan + Prod*Pan** ➢ Repeatability: $\sqrt{MS_E}$ (Root mean square of error) ➢ Validity: $F = MS_{Prod*Pan} / MS_E$ ➢ Discrimination: $F = MS_{Prod} / MS_{Prod*Pan}$	**MANOVA model: Prod + Pan + Prod*Pan** ➢ Homogeneity: MANOVA test of Prod*Pan ➢ Discrimination: MANOVA test of Prod **Canonical Variate Analysis (CVA)** ➢ Dimensionality of product configuration ➢ Product confidence ellipses on CVA map to contrast homogeneity and discrimination

Assessment of Panel Performance Literature

- A useful and comprehensive literature has led to a number of interesting techniques that are possibly unique and may have wider significance.
- Moving forward – how can this information be used to improve panel performance – feedback?

Relating Consumer to Sensory

Consumer to Sensory

- Internal Preference mapping
- External Preference Mapping
- Clustering/Hybrid techniques
- Decision making

Internal Preference Mapping

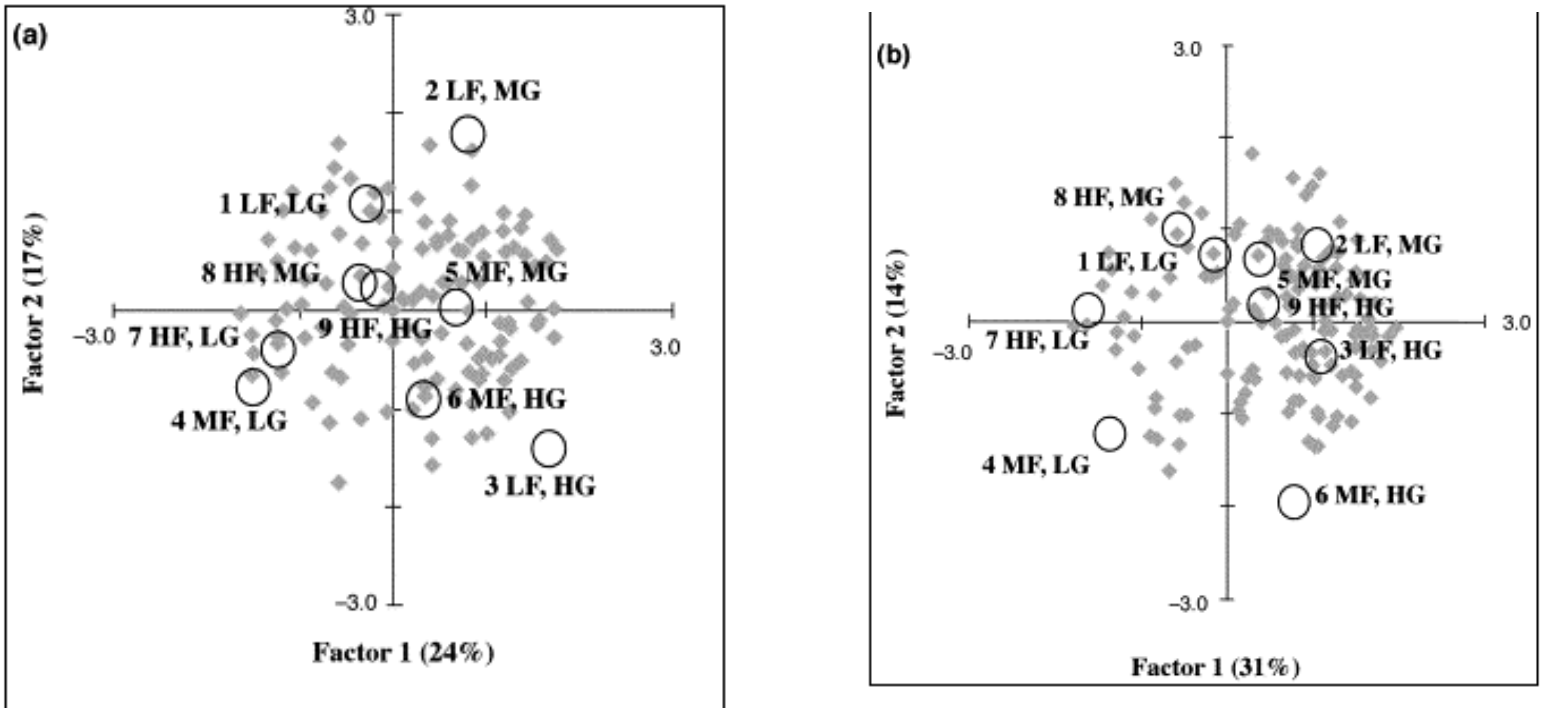


Fig. 4. Internal preference maps—(a) degree of liking of texture/mouthfeel and (b) degree of liking of flavor of Ranch salad dressing (n=144). Squares represent the consumers. Open circles represent the samples. L = low and H = high. F = fat and G = garlic.

Internal preference mapping of hedonic ratings for Ranch salad dressings varying in fat and garlic flavor1 Yackinous*, Wee and Guinard FQAP 1999

Internal Preference mapping

- Based on consumer data
- Easy to correlate in sensory data post-hoc
- Restricted to Vector if using PCA
- Widely used
- Significance testing , confidence intervals around but not in common use

The value of probabilistic unfolding

A"typical" deterministic soln

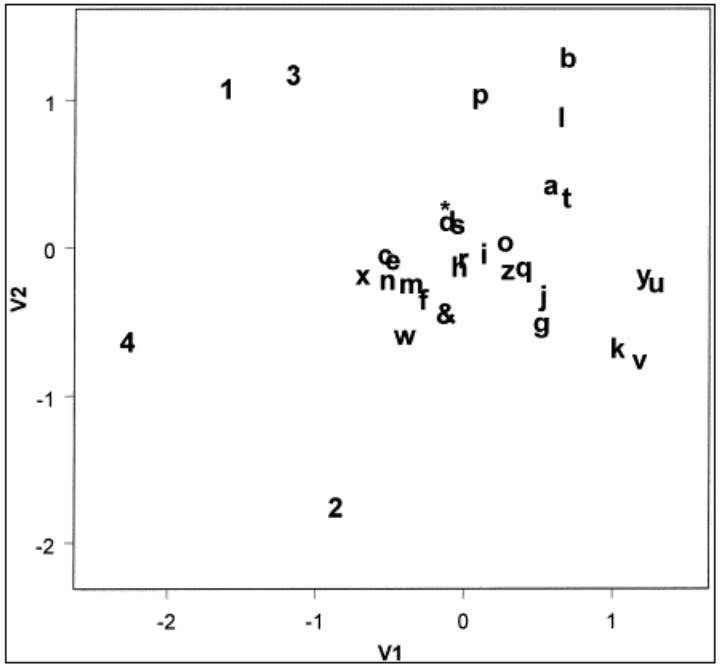


Fig. 5. Deterministic unfolding solution for 28 beverages (*a*, … *z*, &, *) and four consumer segments (1–4).

Probabilistic soln

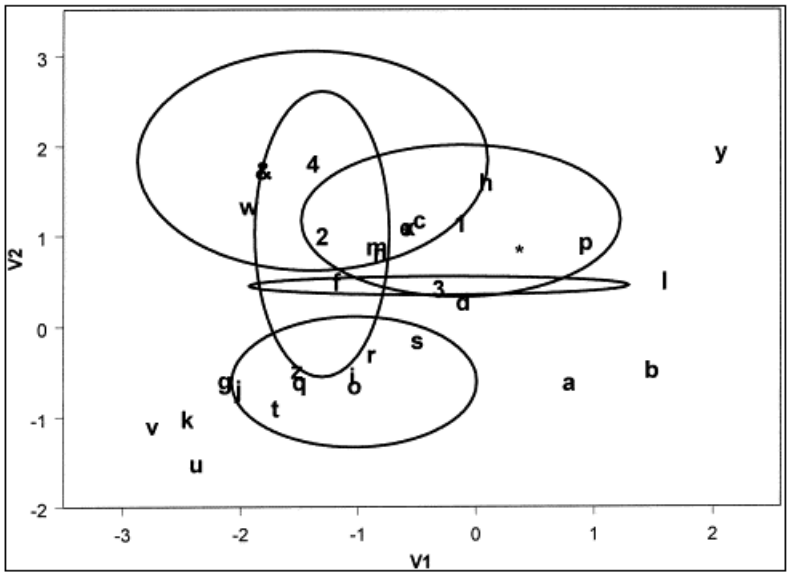


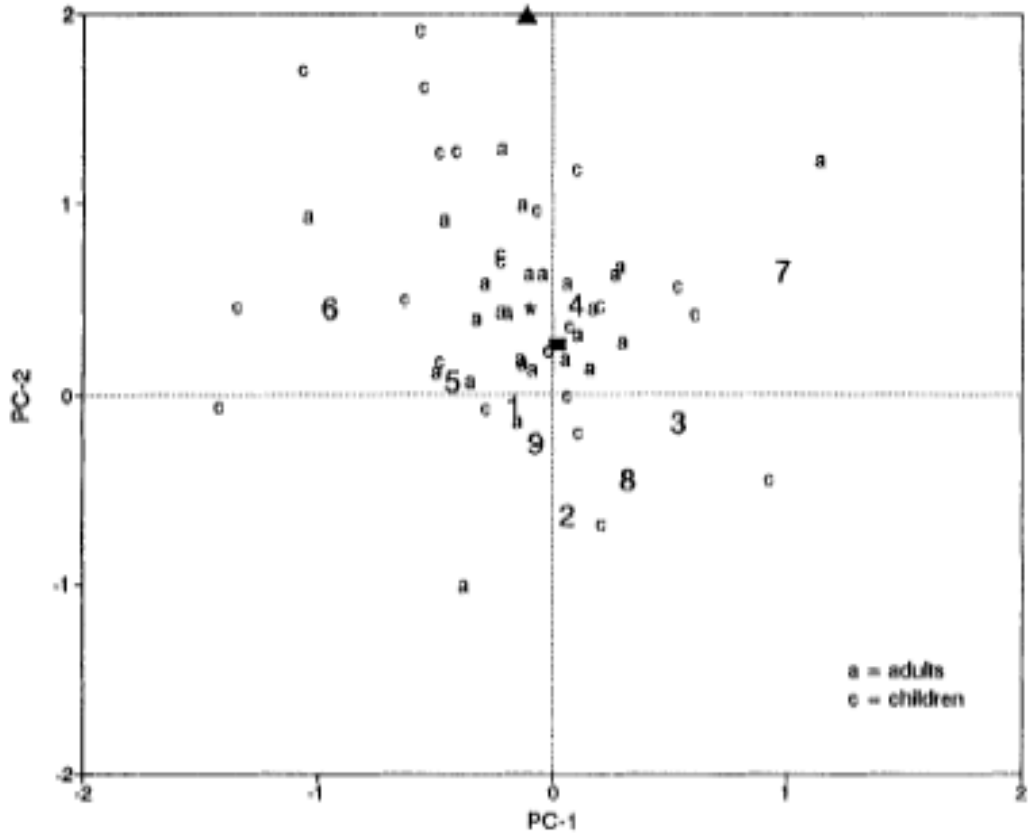
Fig. 6. Probabilistic unfolding solution for 28 beverages (*a*, …, *z*, &, *) and four consumer segments (1,…,4).

These models need to pursued.

Probabilistic unfolding models for sensory data MacKay FQAP 2001

External preference mapping

Descriptive analysis and external preference mapping of powdered chocolate milk Hough and Sánchez FQAP 1998



a-adults

c – children

All elliptical models

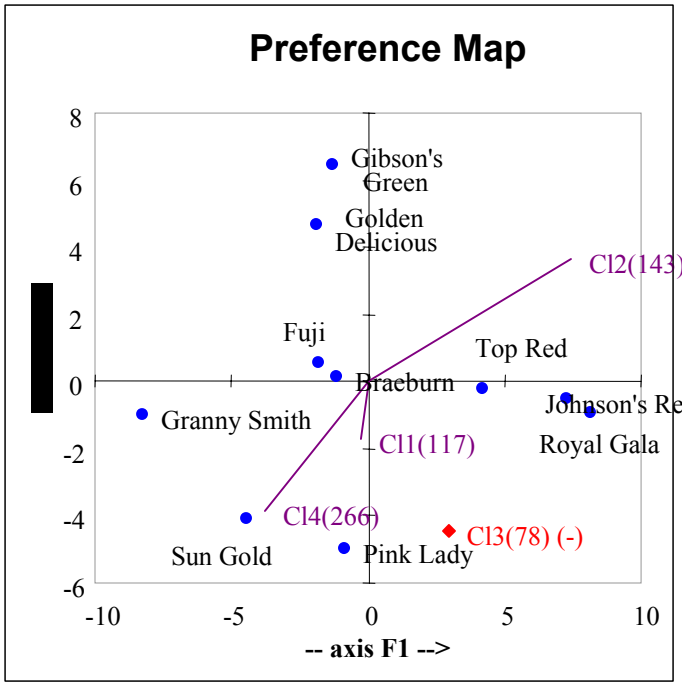
The use of PCA on sensory to derive the external set and then regression of each individual consumer on to the set using a hierarchy of models - vector, circular and elliptical has been widely and successfully applied

External Preference mapping

Solutions and ANOVA significance tests for the best models:

Column	Model	F1	F2	df Model	df Error	SS	MS	R^2	F-ratio	p-value	
Cl1(117)	Vector	-0.011	-0.065	2	7	0.490	0.245	0.076	0.287	0.759	
Cl2(143)	Vector	0.288	0.140	2	7	21.575	10.788	0.669	7.084	**0.021**	
Cl3(78)	Circular ideal point	2.958	-4.442	3	6	33.917	11.306	0.963	51.813	**0.000**	Anti-ideal
Cl4(266)	Vector	-0.146	-0.150	2	7	7.402	3.701	0.418	2.516	0.150	

In bold, p-values equal or lower to the significance level



The position of each product and the best fit model of each group are then plotted.

Easy to get an estimate of a desired product's profile using regression

Problem: Sensory panel weighting of attributes not the same as consumers

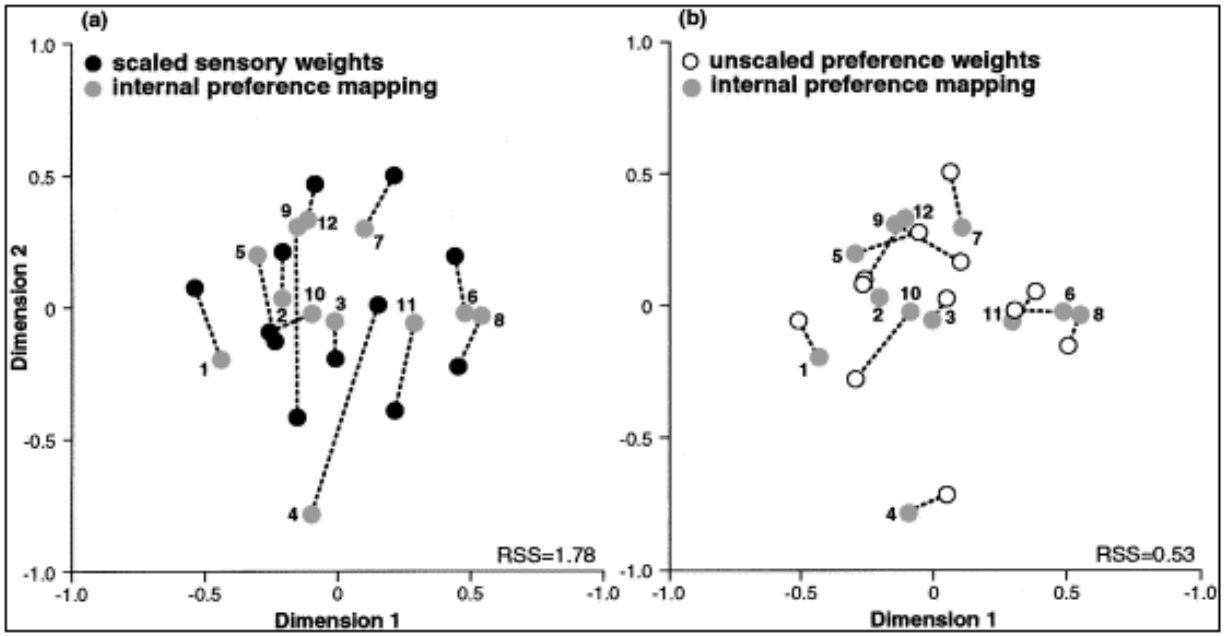
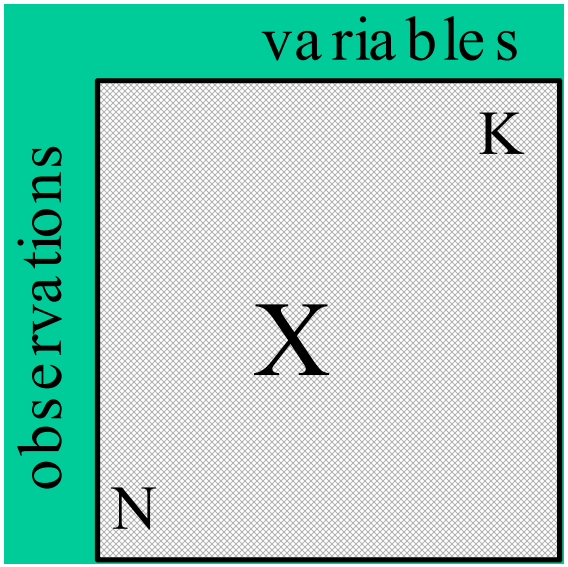


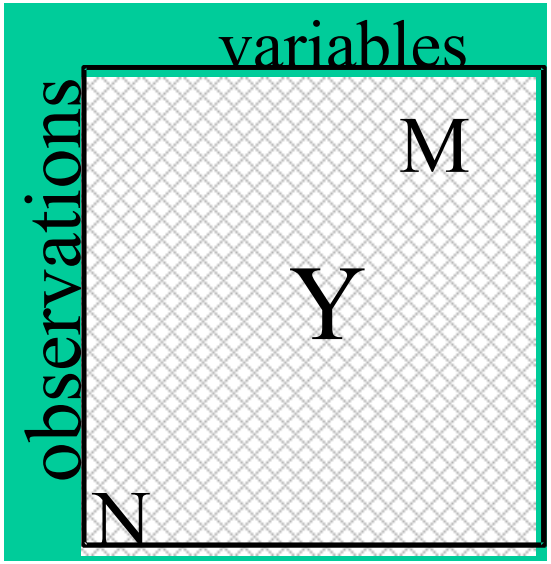
Fig. 1. Graphical comparison of results from Procrustes rotation. Performance of scaled sensory synthesis weights vs. unscaled preference synthesis weights. Samples (Southern Hemisphere apple varieties) are identified by numbers 1–12.

Jaeger and MacFie (FQAP 2000) showed that rescaling weights according to importance for preference gave an external configuration that was closer to that obtained by internal preference mapping.

Think about or go straight to the consumer!

Partial Least Squares

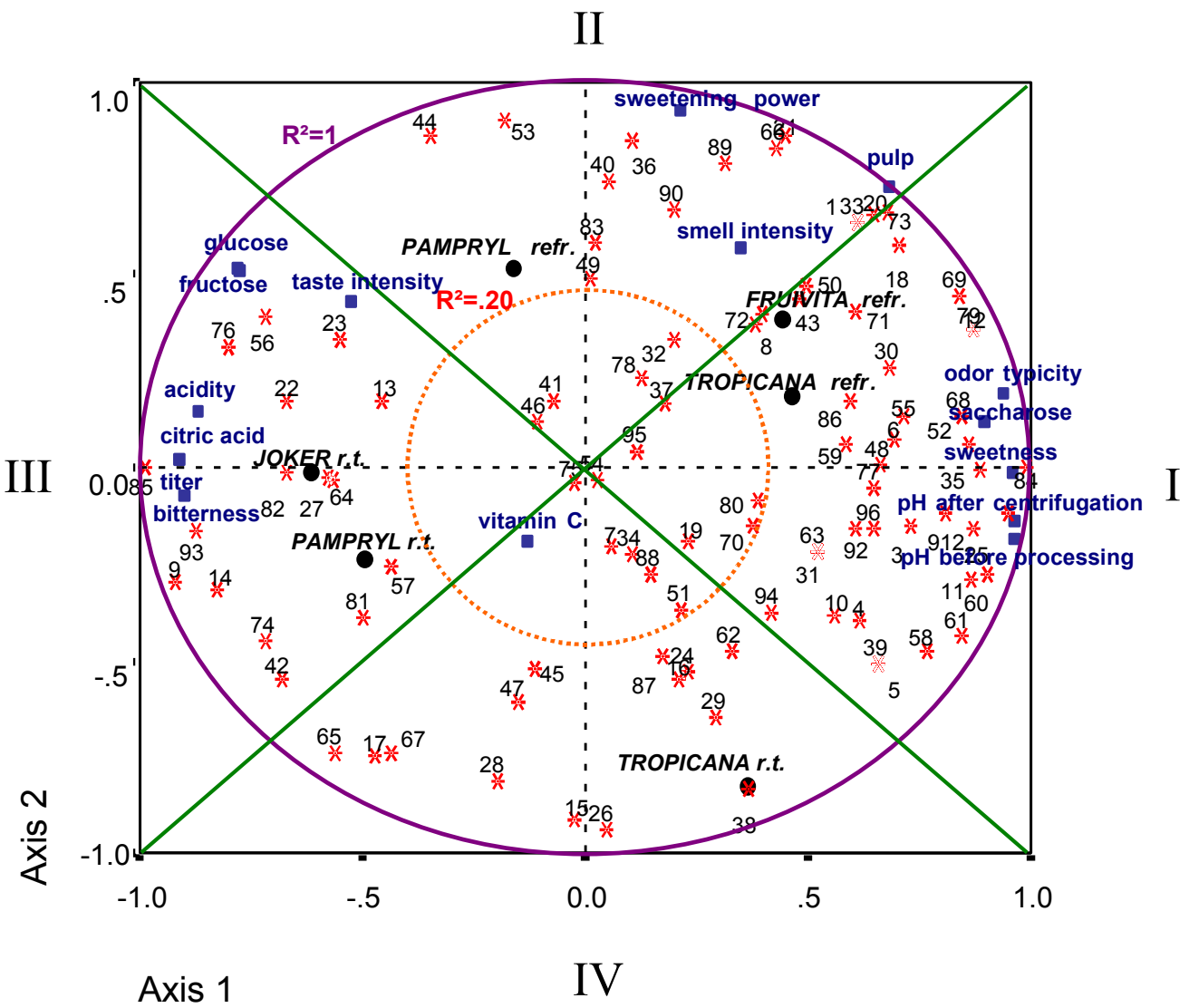




X= sensory data and Y= liking data

X= Principal components of sensory and Y= liking data

X= instrumental and Y= sensory data



Lines bisecting diagram are defining segments numbered in Roman

Tenenhaus, Pages, Ambroisine and Guinot (FQAP in press)

Partial Least Squares Approaches

- Attractive because there is no need to do a data summarisation phase on the sensory data.
- However the method is not invariant to the scaling of X and Y variables
- Cross validation a useful test of predictive value
- Works well when vector model predominates

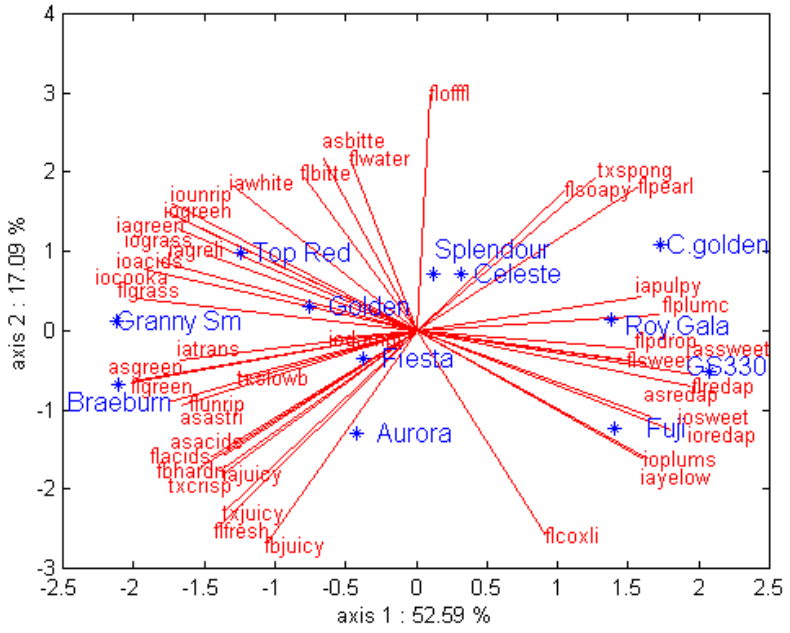
Hybrid methods

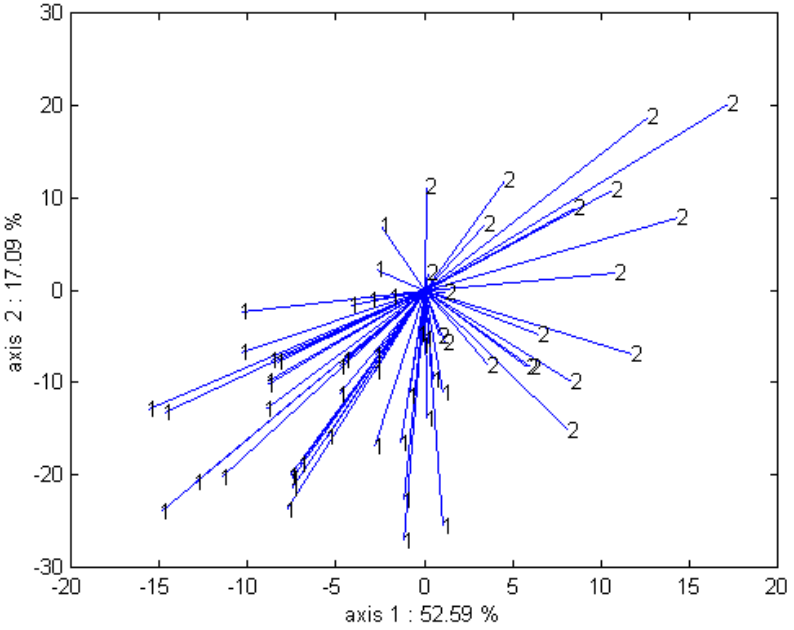
Current position (Popper)

- Need more error-theory based way of selecting the number of factors.
- Clustering is another area that has been lacking in an error-theory. Really, how do I know if these clusters exist and what their stability is? Latent class based models try to address this, and there may be others, so again, perhaps the issue is more focus on an issue rather than new tools.

- Absolutely right!

Segmentation of consumers taking account of external data. A clustering of variables approach E. Vigneau, and E. M. Qannari FQAP 2002

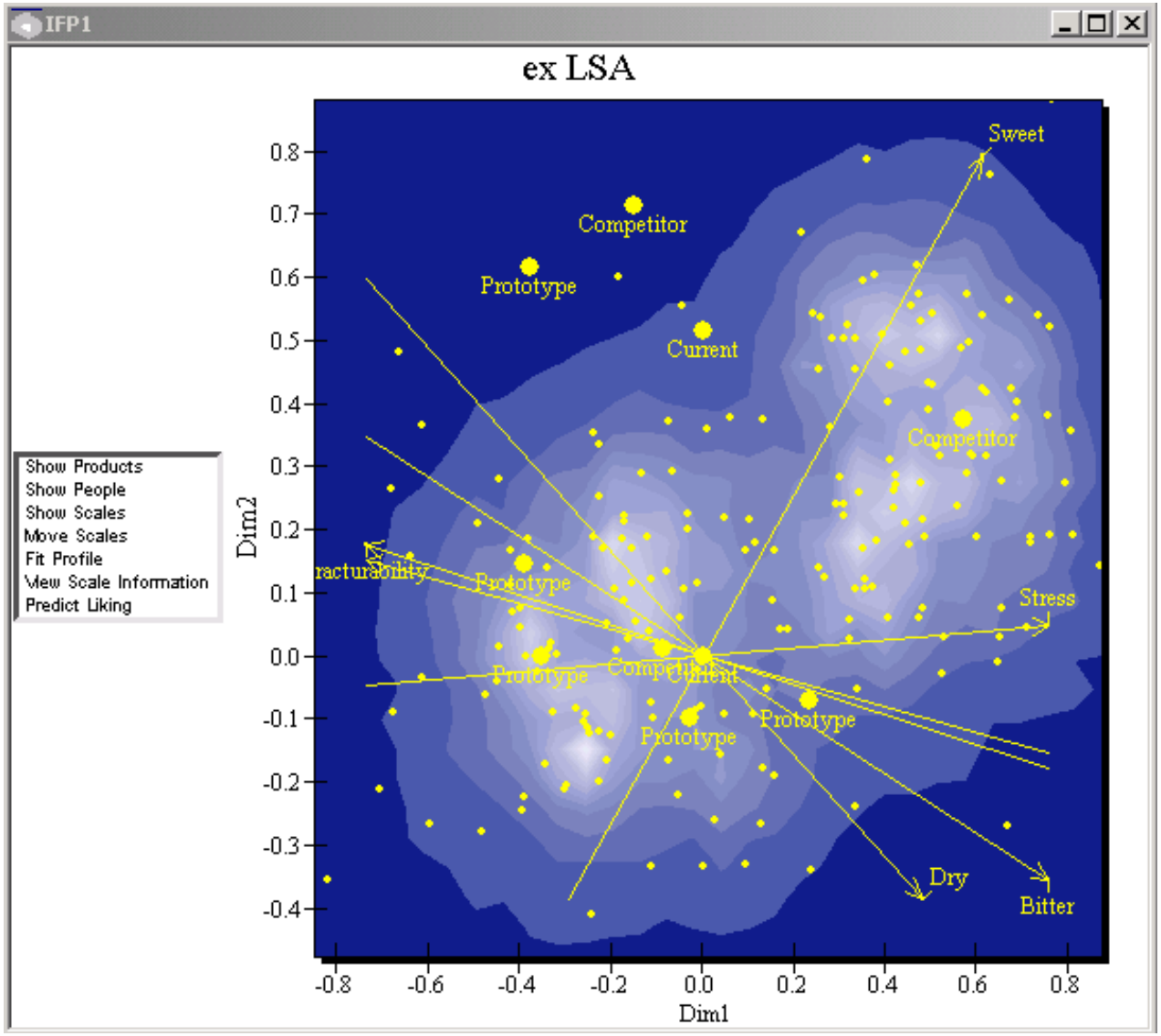




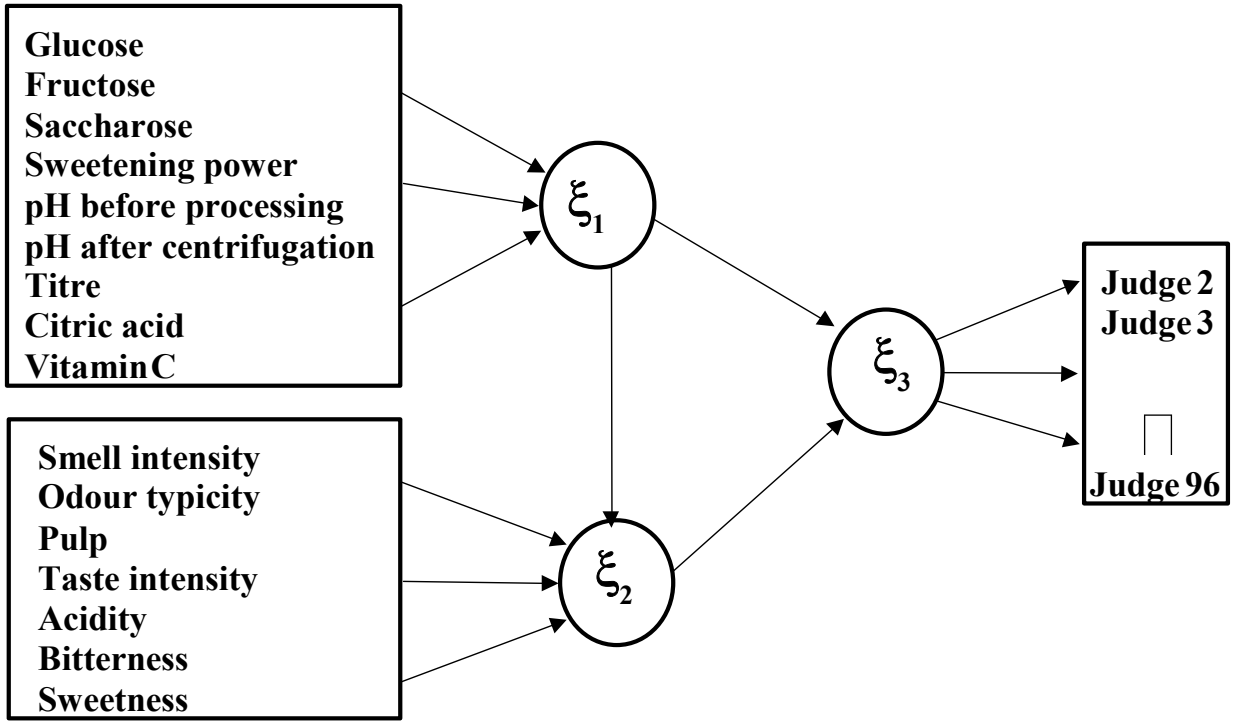
Has shown how to cluster respondents and project them on on to internal and external data. Very clever.

Available in Senstools.

Landscape Segmentation Analysis



Multi-block PLS modelling



Tenenhaus, Pages, Ambroisine and Guinot (FQAP in press)

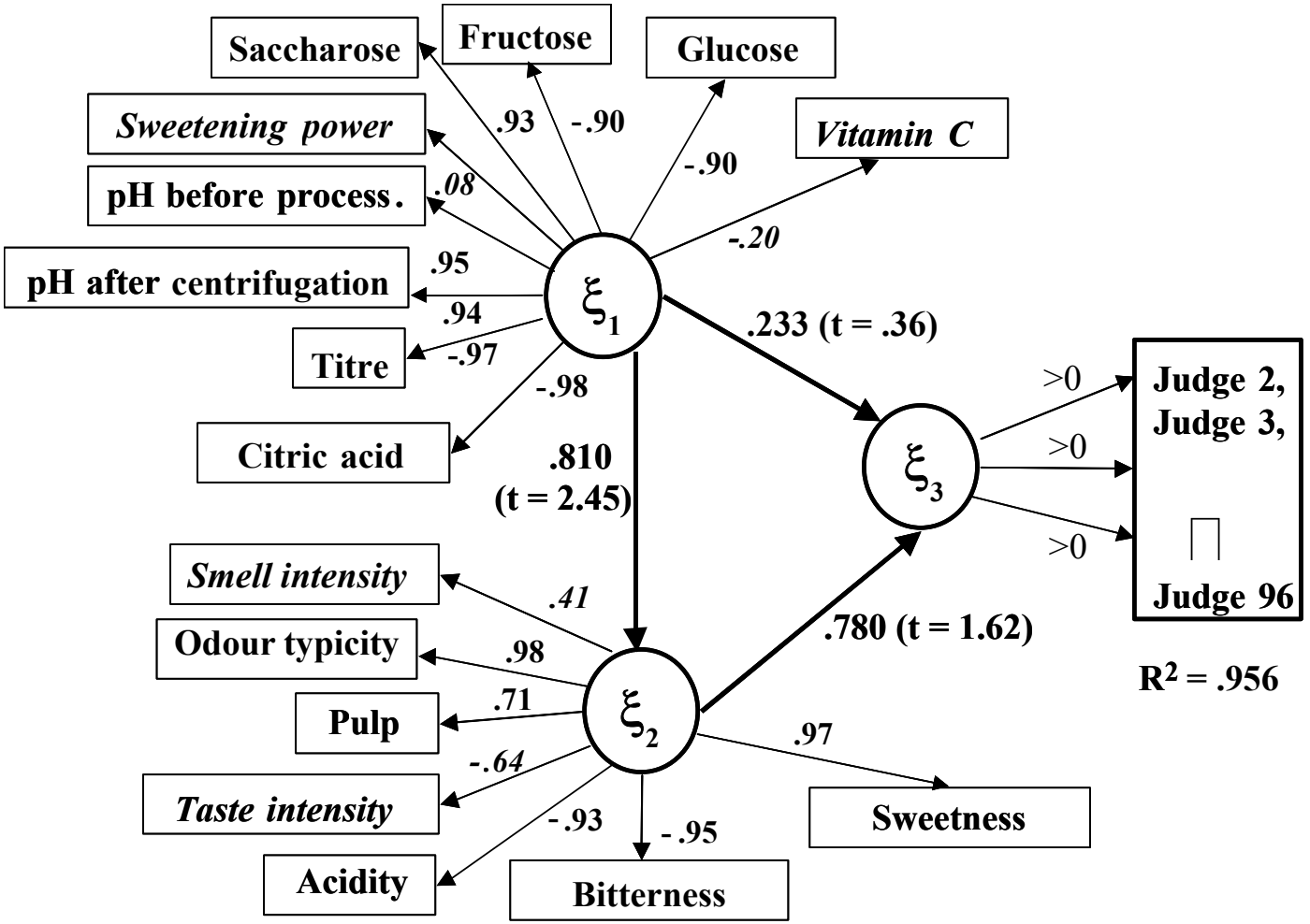


Fig. 8. Estimate of the hierarchical multi-block PLS model (w and c loadings). Non significant loadings are in italic

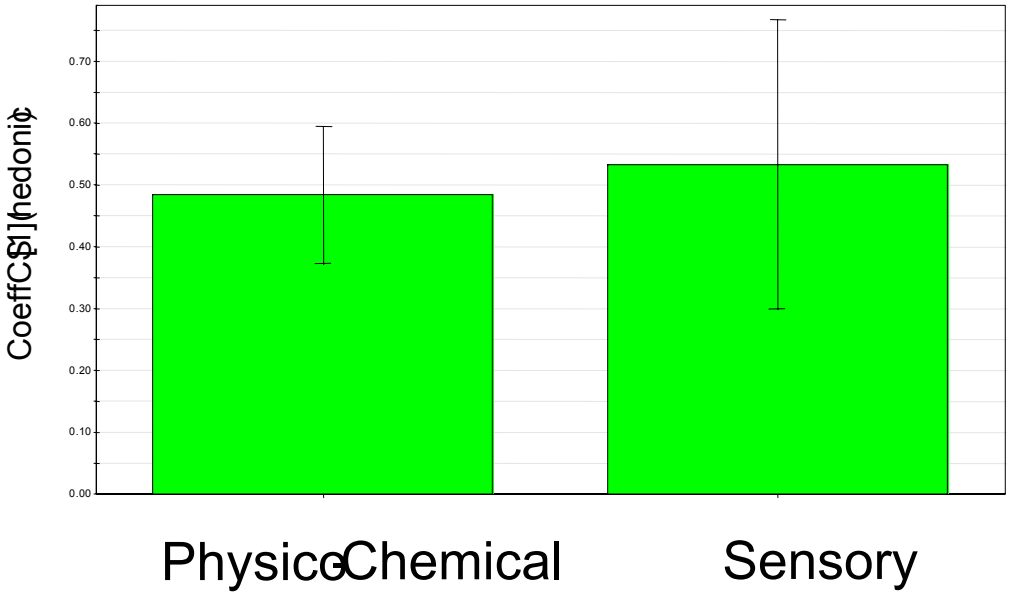


Fig. 7. SIMCA-P® software output, validation of the PLS regression of the hedonic score on the physico-chemical and sensory scores

New PLS methods are being developed outside sensometrics that are finding use – L-PLS

Decision making aids

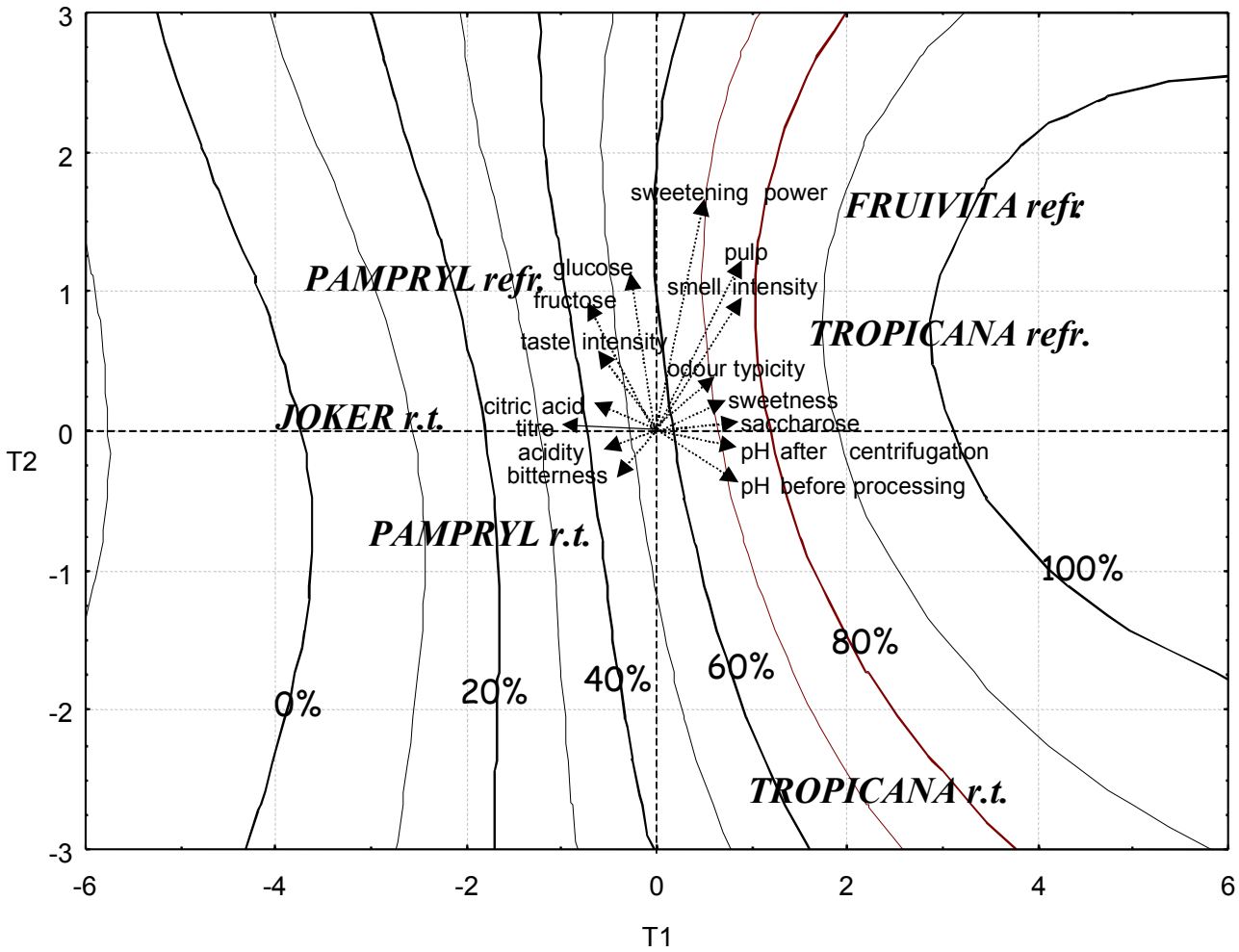
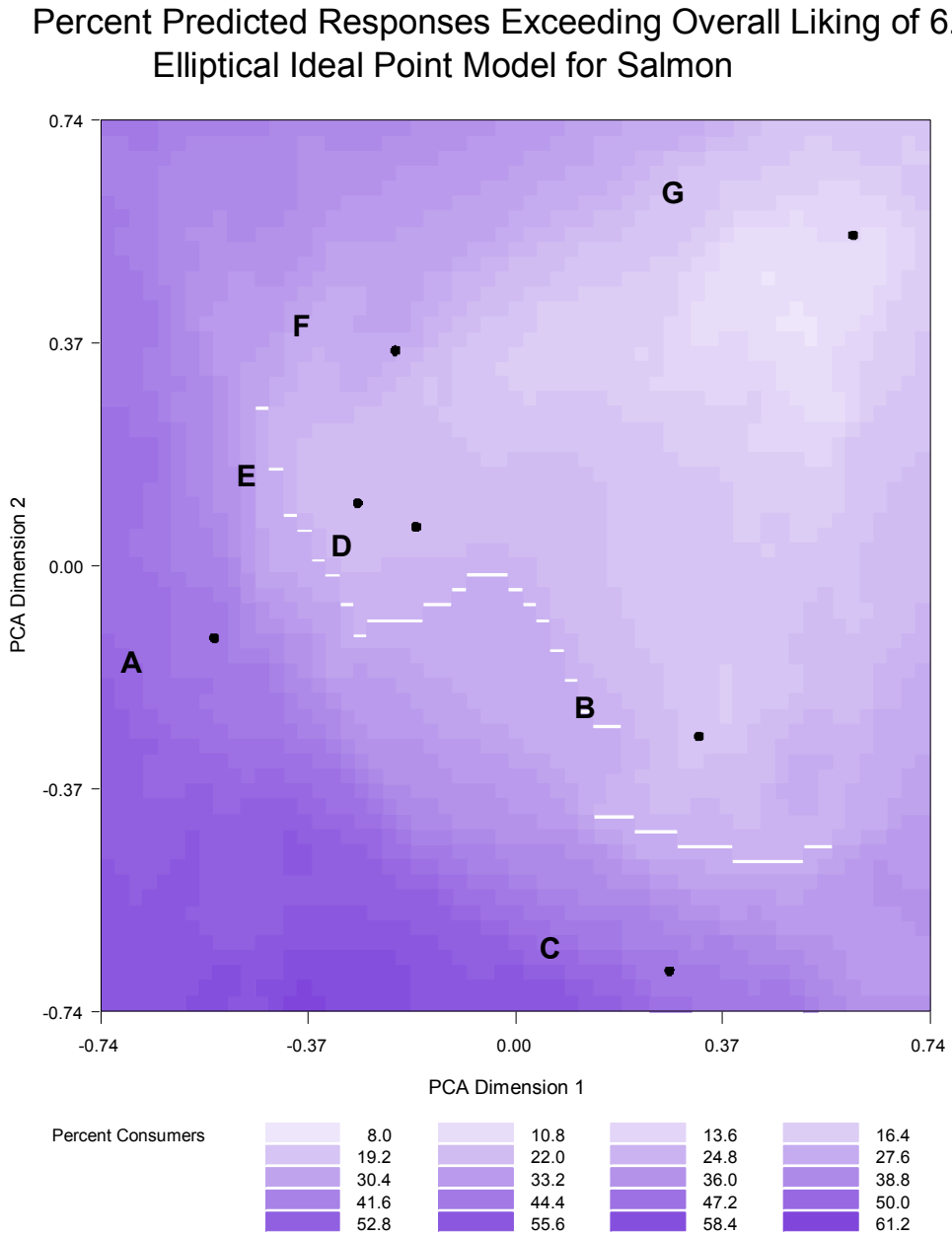


Fig. 4. Map showing the percentages of the judges in group I classifying a product (t_1, t_2) above average with the products and their characteristics X Tenenhaus, Pages, Ambroisine and Guinot (FQAP in press)



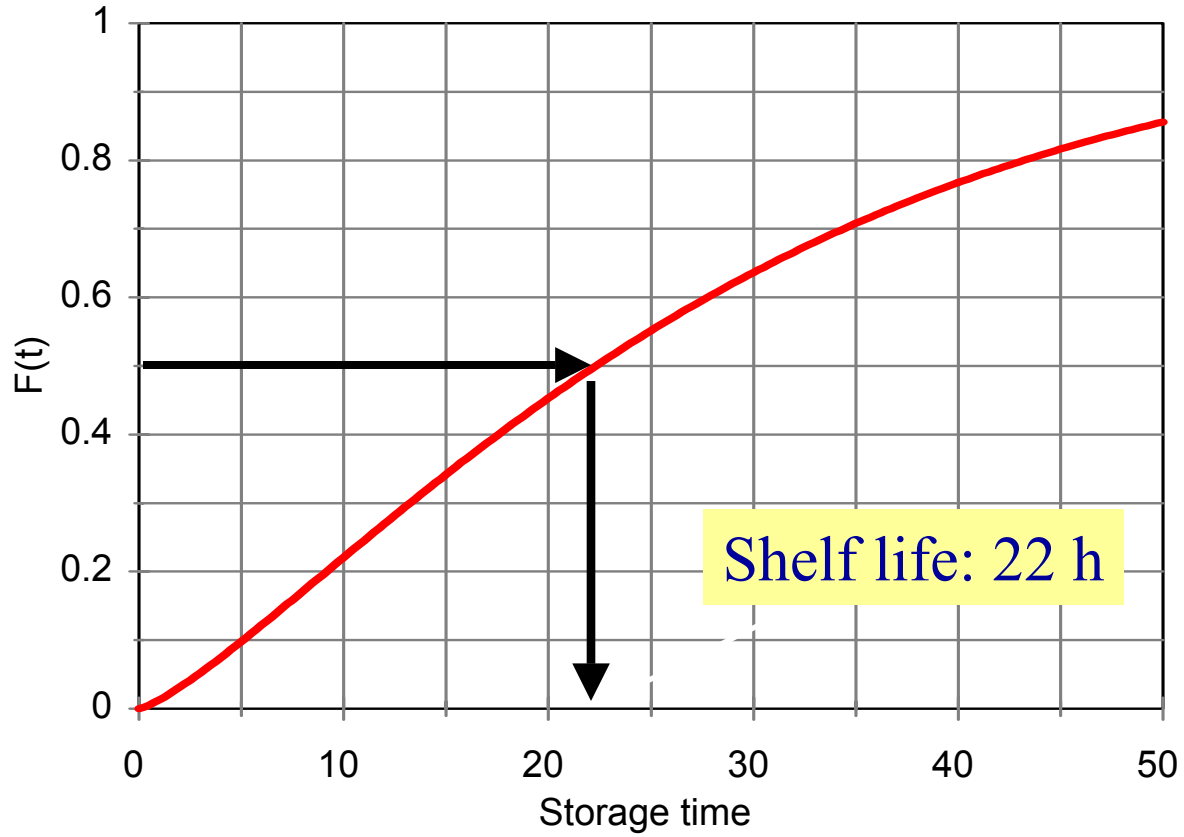
This contour plotting idea, derived by Danzart, is now widely used and is available in FIZZ

Relating sensory to consumer

- Crucial to the relevance of sensory in an NPD commercial environment
- A large number of useful techniques
- More research on significance testing, decision making aids, segmentation strategies, probabilistic methods required
- Structural equation modelling may have promise

Shelf-life and Time Intensity

Probability of rejection



- The yogurt doesn't last 22h.
- With 22 h storage there is a 50% probability that a consumer will reject it.

References

- Hough et al. 2003. Survival analysis applied to sensory shelf-life of foods. J Food Science 68: 359.

- Hough et al. 2004. Determination of consumer acceptance limits to sensory defects using survival analysis. Food Quality and Preference, in press.

- ***Calle et al. 2004. Bayesian survival analysis modeling applied to sensory shelf life of foods. 7th Sensometric Meeting, poster presentation.***

Time -Intensity

- Standard techniques – anova etc
- Overbosch, Liu and MacFie rescaling ideas
- Van beuren – PCA, Dijksterhuis PCA non centred PCA
- Wendin, Janestad – modelling, parametric
- Eilers and Dijksterhuis 2004
- Future – back to basics?

A parametric model for time-intensity curves

Eilers, P.H.C. / Dijksterhuis, G.B., Food Quality and Preference, Apr 2004

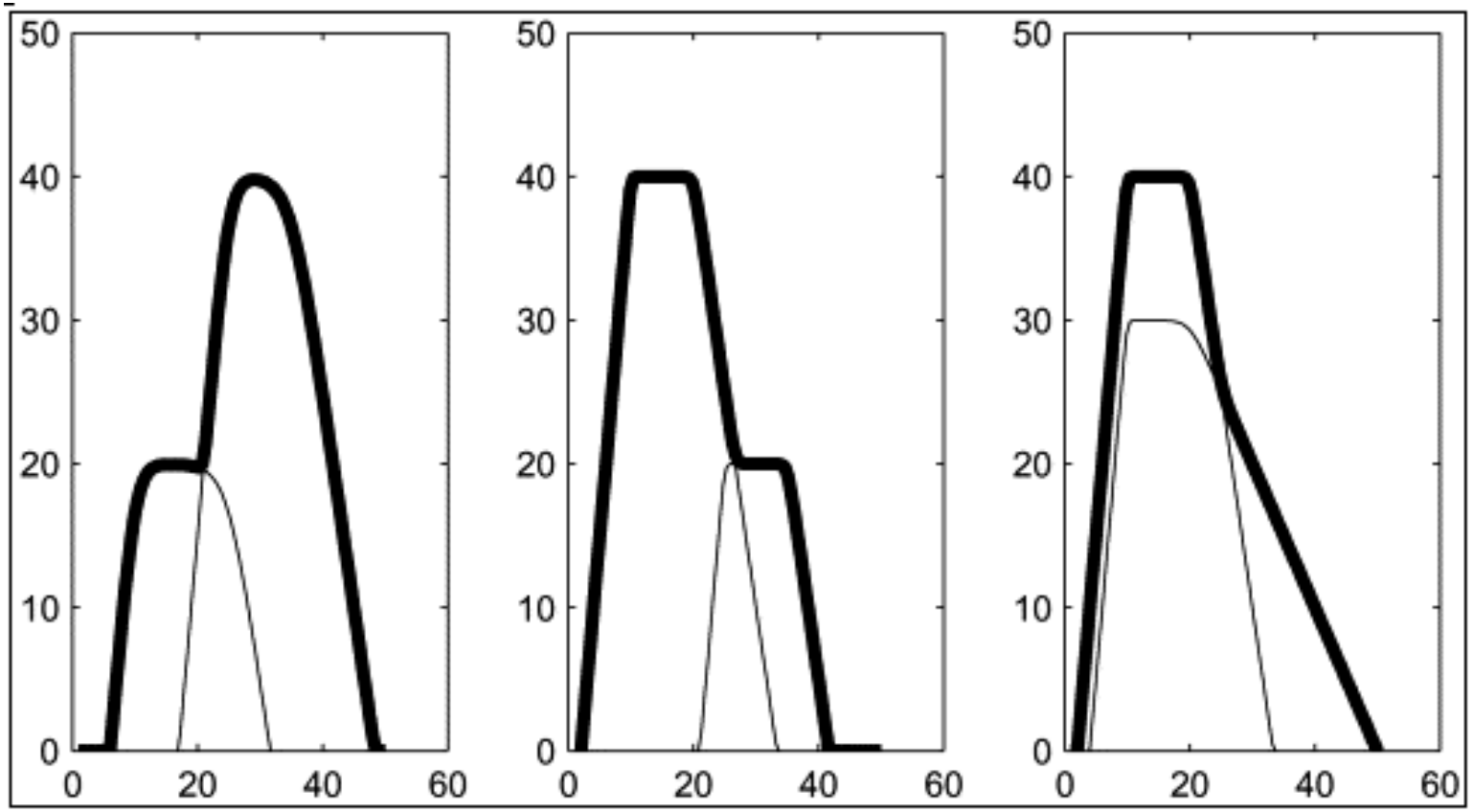
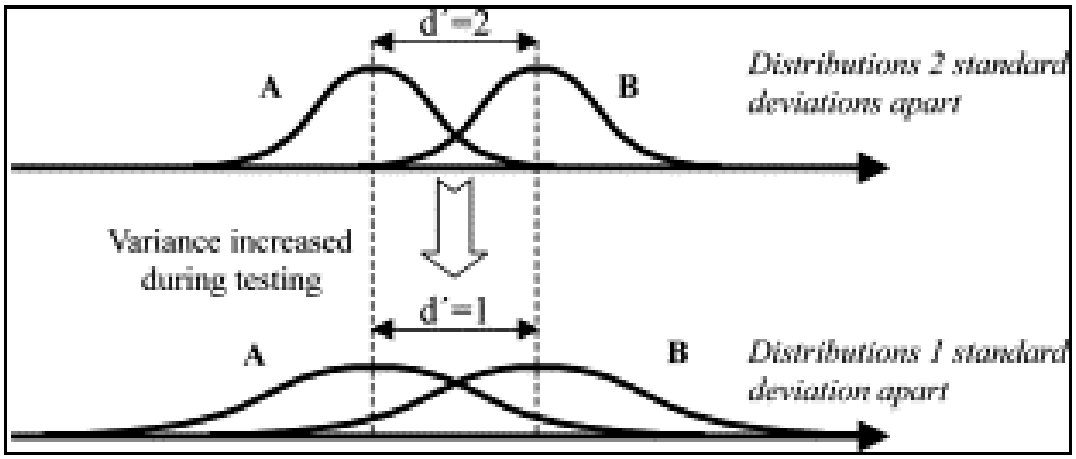


Fig. 6. Some examples of T–I curves according to the extended the model. The response to the individual systems are shown as thin lines. The thick line is the combined response.

Metrics = Measurement

Metrics

- Thurstonian models
- Biases
- New metrics



Using the variance as well as the mean of response distributions has enabled and some clear thinking has enabled O'Mahony and colleagues to develop a valuable programme of testing that is

Comparing the performance of different testing procedures

The effects of warm up samples, memory , retasting etc have been tested.

This programme is producing as series of useful results that resonate with sensory practitioners

Worth pursuing! More workers needed!

Binomial versus Beta Binomial

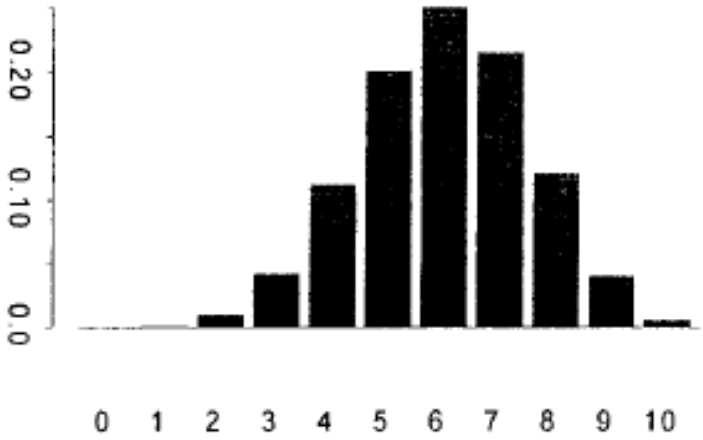


FIG. 23.10. BINOMIAL: n = 10, $P = 0.6$

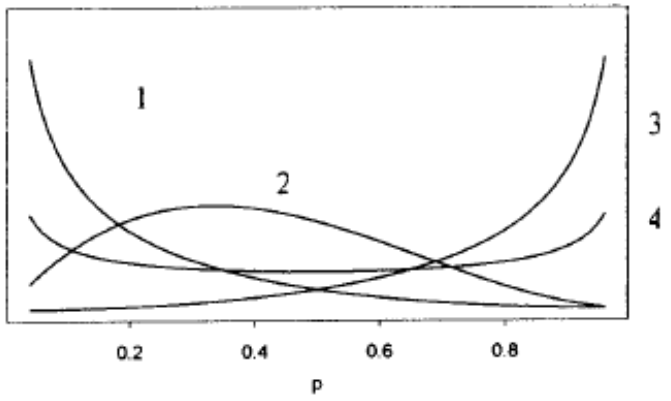


FIG. 23.11. SHAPES OF THE BETA DISTRIBUTION

Ennis and colleagues have developed comprehensive software that takes the test and replicate variability into account.

These tests offer potential improvements in power and sensitivity and, again, resonate with sensory professionals.

My prediction is that this approach will continue to increase in popularity and should be addressed by sensometricians more widely

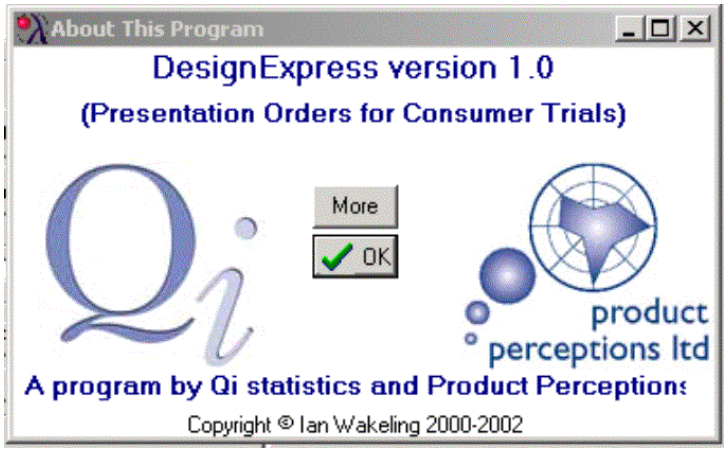
Biases and problems

Balancing designs for positional and carry over effects

Designs widely used.

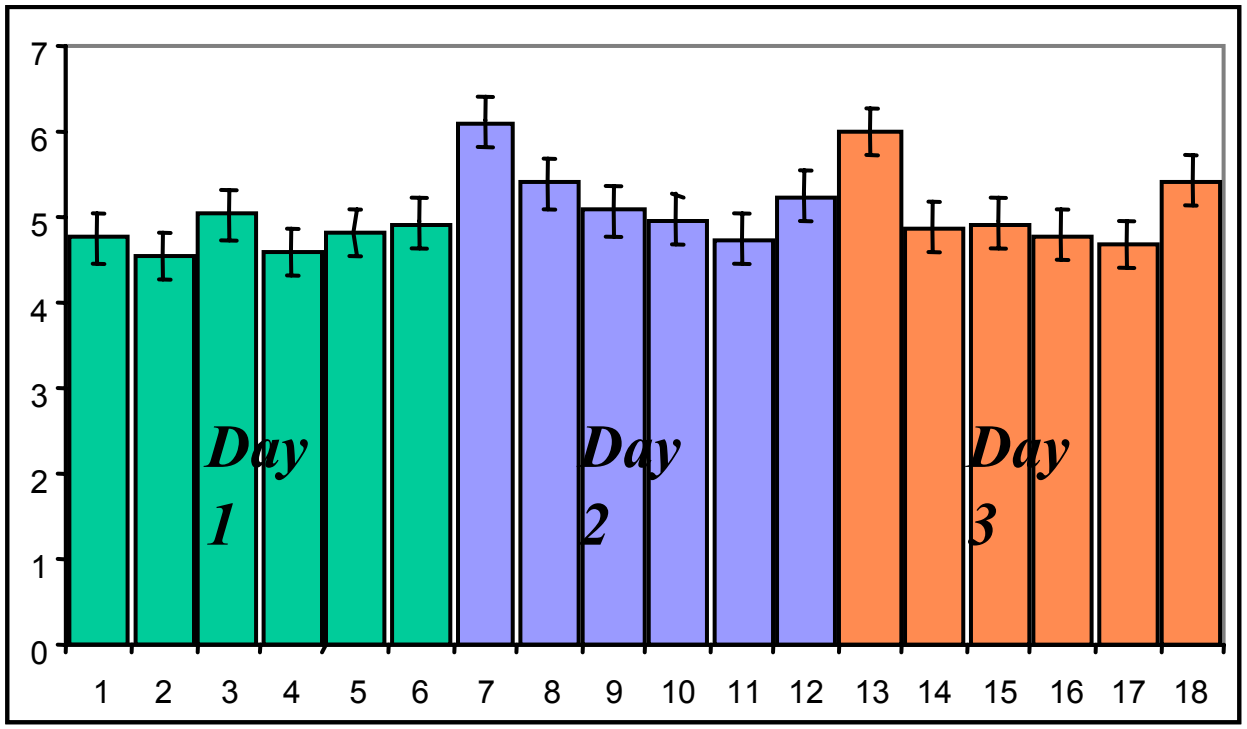
Need for further research?

Kunert criticisms?



Warm up effects

Order effects for Overall Liking (Evaluation)



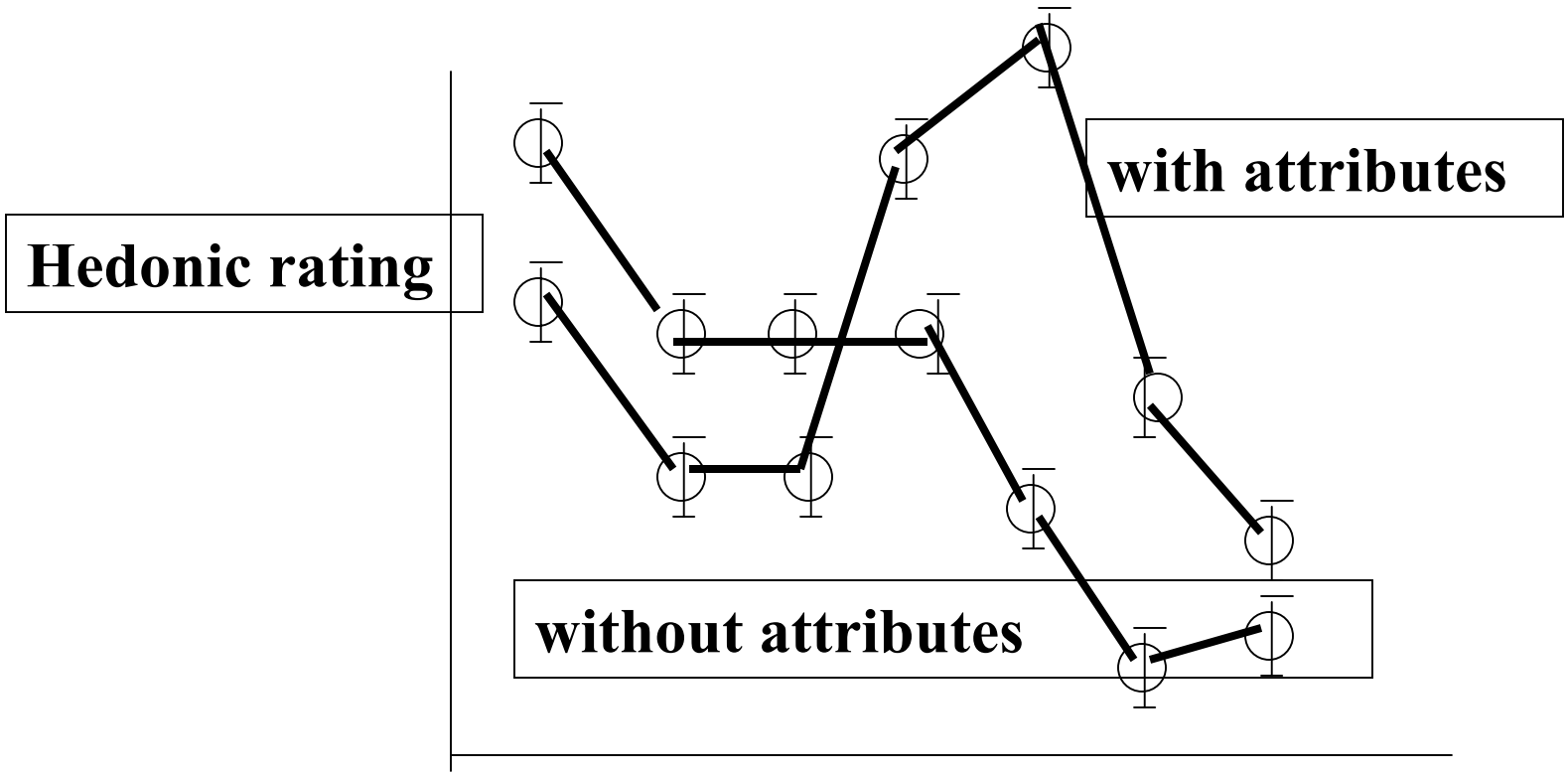
On day 1 we used a dummy sample.

The rise in scores at the end of the trial is also a common effect

ct.

What is the effect of these biases on clustering or preference mapping?

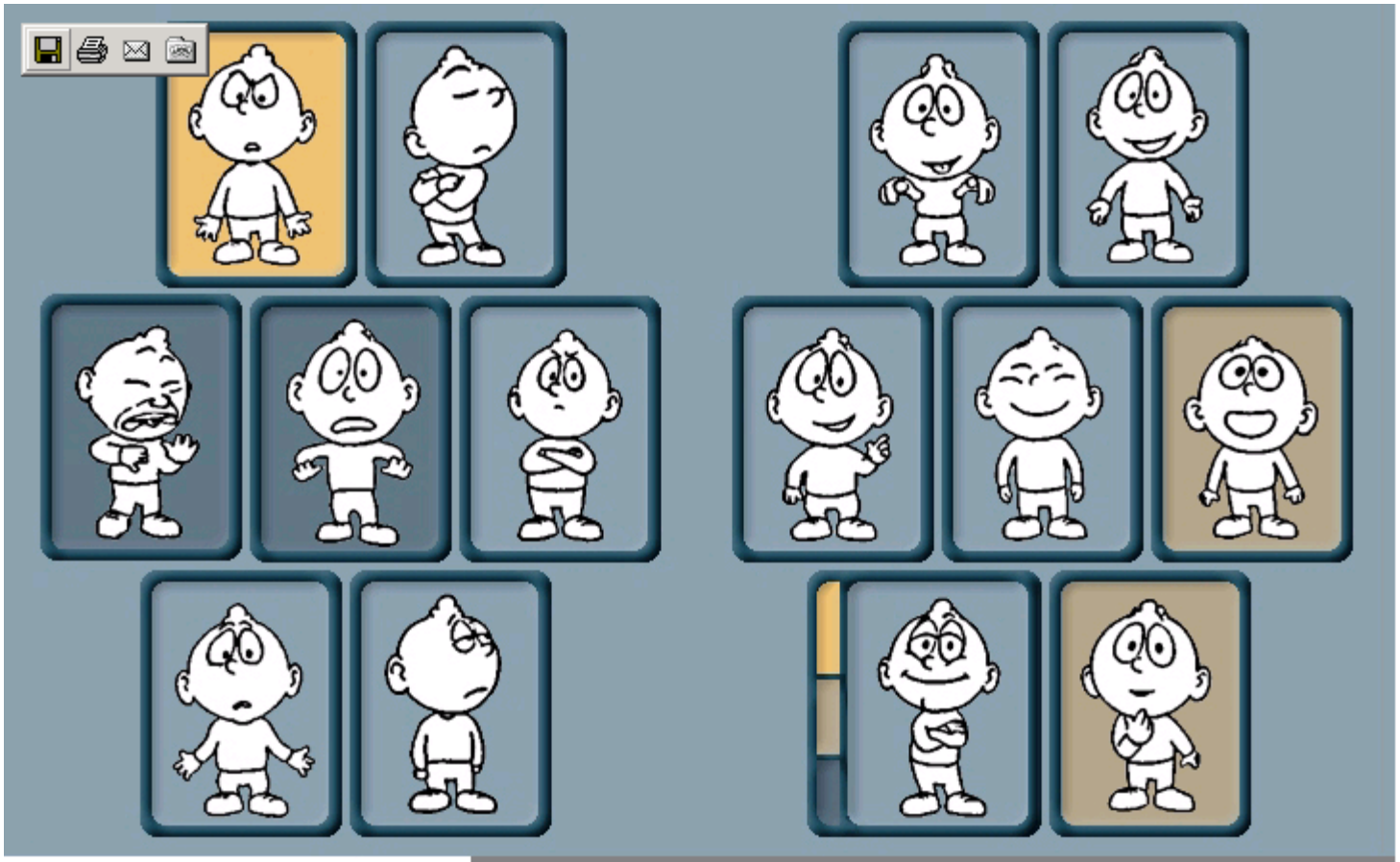
Does the addition of attribute questions alter the hedonic ratings?



milk%	100	75	75	50	25	25	0
dark%	0	25	25	50	75	75	100
sugar gms	9	18	0	9	0	18	9

Is there life after the 9 point hedonic scale?

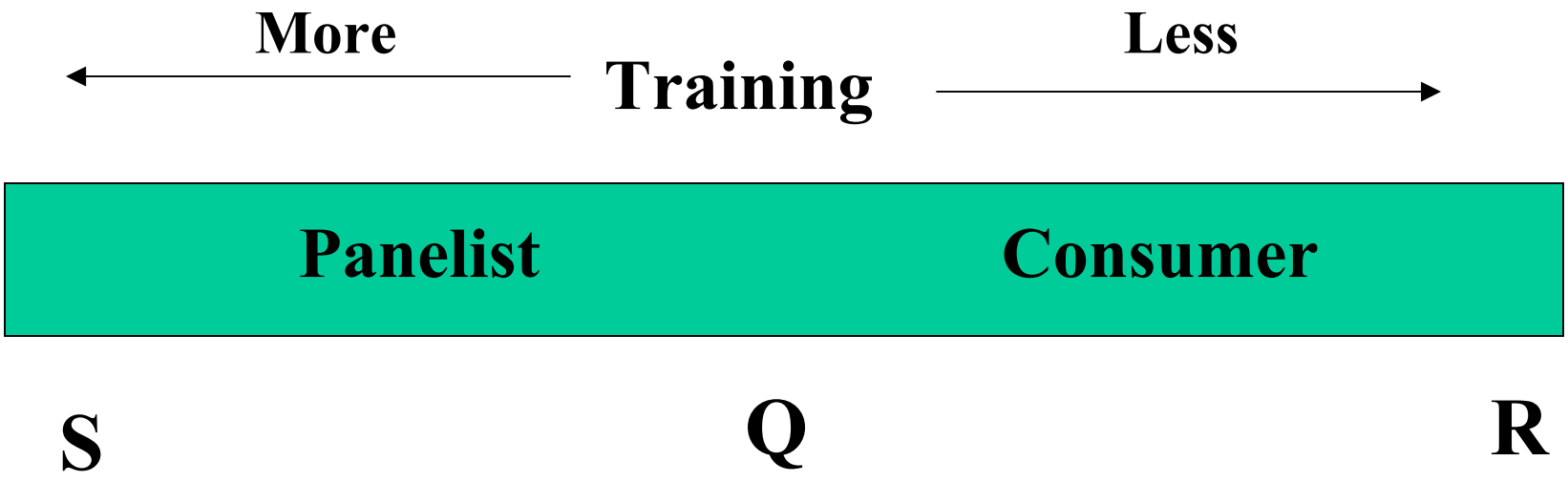
- Non- English speakers generally not happy with it.
- Converting feelings to numbers not natural
  - (Cognition vs emotion controversy)



Please rate the puppets to express what you feel towards this car model.

When you are finished, you can click the grey button

Measuring Intensity



Triadic Elicitation

Which pair of products in this triad are more similar to each other?

In what way are they more similar to each other than the third?

Multi-lingual Consumer Vocabulary for Dessert Apples

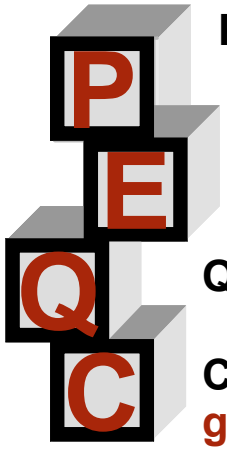
England*		Denmark		Belgium/Flemish		Belgium/French		Spain	
Acidic	20	Sur	19	Zuur	20	Acide	22	Acido	23
Sweet	15	Sod	17	Zoet	13	Sucree	14	Dulce	10
Crisp	11	Sprod	9	Krisp	3	Croquante	4	Crjiente	4
Juicy	7	Saftig	5	Suppig	5	Juteuse	10	Juguso	4
Hard/firm	7	Hard	14	Hard	11	Ferme	13	Dura	8
Soft	6	Blod	9	Zacht	5	Molle	7	Blanda	5
Bitter	7					Amere	3	Amargo	4
Cooking apple	4			Stoofappel	4	Pour la Cuisine	4		
Tasteless	5					Fade	13	Insipido	11
Mealy/Floury	[A]	Melet	9	Melig	12	Farineuse	8	Harinoso	6
Coarse	5	Grov	5			Granuleuse	4	Aspera	3
Dry	4	Tor	3	Droog	3				
Spongy	4			Loose Stru	5				
Crumbly				Korrelig	4				

* Based on RGM interviews with 25 consumers in each country. Descriptors correlate significantly (r^2>0.05) with GPA consensus

[A] Only descriptors used by 3 or more (>12%) consumers are shown

Repertory Grid is a method that elicits consumers' response by gathering their knowledge, feelings and judgments on a particular subject for which they have experience.

- Rep Grid consists of 4 primary components:



Purpose—what you want to achieve

Element Class/Elements—samples representative of the subject content area

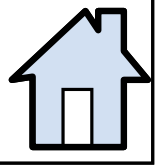
Qualifiers—statements used to direct the respondent's thinking

Constructs—bi-polar scales generated by the respondent; once generated, the respondents rate each element on each of the constructs

- And is executed with the following key steps:

1 Provide Experience

- *Ensure relevant use experience*



2 Create Grid

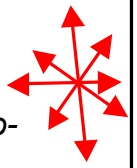
- *Generate constructs*
- *Ladder up & down*

3 Rate Elements

- *Rate elements on constructs*
- *Eliminate redundancies*
- *Refine elements*

4 Analyze & Interpret

- *PCA biplots*
- *Other techniques-- Resistance-to-Change*



Generating and handling descriptors

- A rather specialist subject of sensory
- Repertory grid does it but other methods exist
- **The application of a text clustering statistical analysis to aid the interpretation of focus group interviews** Eric Dransfield FQAP 2004
- More papers appearing – new methods needed and topic where we can make a contribution
- (maybe free choice profiling is not dead – Mostaffa)

Metrics

- An important area of Sensometrics
- Application of Thurstonian principles producing important results
- More questioning of current practice needed
- New metrics needed to replace 9 point scale
- Methods to handle consumer generated descriptors needed
- More researchers needed

Future (Popper)

- Then there is the atheoretical stance of sensometrics. Where is the perceptual/cognitive theory that be should be striving to build?
- I wish there was more theoretical work and focus on estimating parameters of a perceptual/cognitive model rather than parameters is some all-purpose statistical model.
- This is what makes the Thurstonian framework and what Danny is doing attractive to me.

Fundamental modelling of sensory processes related to flavour structure activities

- Not much published in FQAP or Journal of SS
- Chemical Senses?
- Ennis has published in this area
- Expect this area to grow

Don’t forget the package

"A DISAPPOINTING WINE, CONSIDERING THE GREAT LABEL GRAPHICS."

"Winner" Element
Scored +17

A truly great beer from the company that cares

The crystal, golden quencher

The taste everybody's gotta love

Now's the time to enjoy

"Loser" Element
Scored -6

How interested are you in this concept?
(1 = Not at all interested . . . 9 = Very interested)

The impact of olfactory product expectations on the olfactory product experience Scharf and Volkmer FQAP 2000

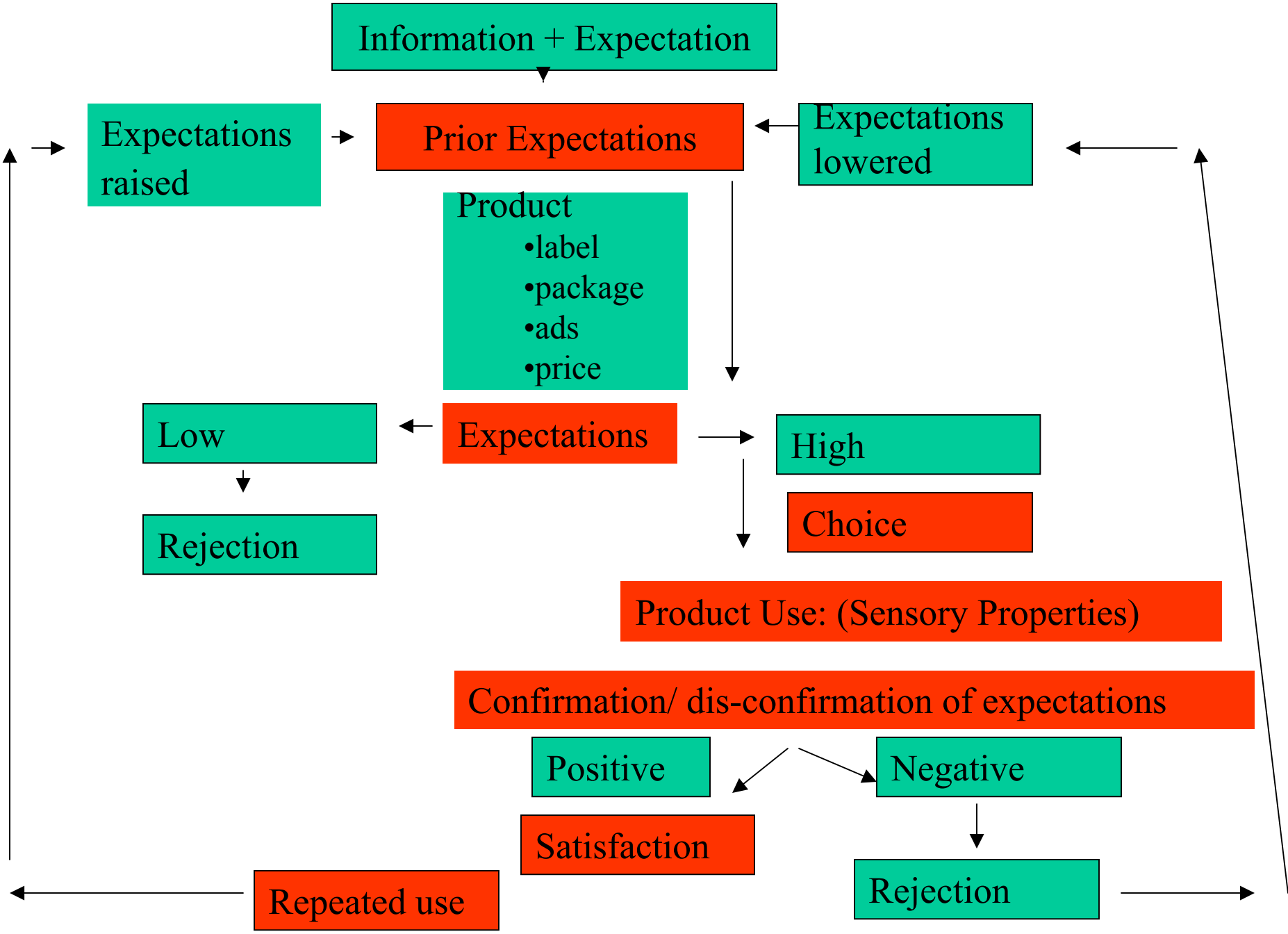
Table 3. Results of the assessment of the fragrance/packaging combinations (averages from 0 to 100 and significance of the difference between the averages; n.s.=no significance; P> 0.1; all interactions were not significant)

		Dependent variable				
Independent variable		Acceptance	Intensity	Sweetness	Freshness	Floweriness
Fragrance	"Anaïs"	43,7	59,6	52,9	53,7	58,1
	"Diva"	36,4	67,2	61,8	44,8	51,7
Intensity	"Low"	39,9	63,5	57,8	48,2	54,8
	"High"	40,2	63,3	56,9	50,3	55,0
Colour	"Pastel green"	40,9	60,8	53,7	54,4	55,2
	"Dark red"	39,1	66,0	61,0	44,1	54,6
Brand name	"TIP woman"	35,4	64,1	57,9	48,2	55,4
	"C. Dior"	44,6	62,7	56,8	50,3	54,3
Significance of effects	Fragrance	0,0	< 0,01	< 0,01	< 0,01	0,02
	Intensity	n.s.	n.s.	n.s.	n.s.	n.s.
	Colour	n.s.	0,02	0,01	< 0,01	n.s.
	Brand name	< 0,01	n.s.	n.s.	n.s.	n.s.

A case study in relating sensory descriptive data to product concept fit and consumer vocabulary

Carr, B.T. / Craig-Petsinger, D. / Hadlich, S., Food Quality and Preference, Jul 2001

Fig. 3. PLS map of fit-to-concept ratings and sensory flavor data illustrates the primary and secondary flavor drivers of fit-to-concept. The map shows that the East and West segments have different flavor drivers.


Information + Expectation
Expectations raised
Prior Expectations
Expectations lowered
Product
•label
•package
•ads
•price
Low
Expectations
High
Rejection
Choice
Product Use: (Sensory Properties)
Confirmation/ dis-confirmation of expectations
Positive
Negative
Satisfaction
Repeated use
Rejection

Success= Choice + Satisfaction + repurchase

- Methods to establish the role of product performance (assessed by the user) in the total marketing mix, cheaply, easily and convincingly are required.
- Choice models, expectation models, satisfaction models, repurchase models

Future (Punter OP&P)

- Sensory-on-Demand.
- users do not need any software except a browser and internet connection.
- The logical next step is the analysis and reporting on demand, so that the the client only has to provide place, products and respondents.
- Ie a 100% internet application.

Future (Van Dongen)

- --Better define what it is and what it is for - I still don't know what the organization does other than the meetings every other year....
- --Help statistical folks communicate results better - write standards for showing different types of data like ASTM standards
- --Educate sensory folks in statistical tools - should there be minimum statistical knowledge required for sensory folks?

Future (Mostaffa)

- how to determine completely new concepts (new products with new tastes, new appearances... never heard of).
- This poses a real challenge in terms of statistical designs, optimisation, statistical treatment of data because we need to go beyond the limits of the experimental universe.

Summary

- Impressive literature and practical methods
- Have helped sensory move towards consumer insights
- Sensory measures and methods – more fundamental approaches
- Make Metrics an important part of our conferences and research area
- Methods for helping decision making
- Keep focus on product performance
- Role of sensory in satisfaction/repurchase