

Statistical issues relating to hierarchical free sorting task

Qannari E. M.

Ph. Courcoux



Y. Taylor

D. Buck

K. Greenhoff



Overview

- **Free sorting task: interest and limitations**
- **Taxonomic free sorting task**
- **Cophenetic distance**
- **Non metric MDS on the average cophenetic distance matrix**
- **Illustration**
- **Conclusion**

Free sorting task: interest and limitations

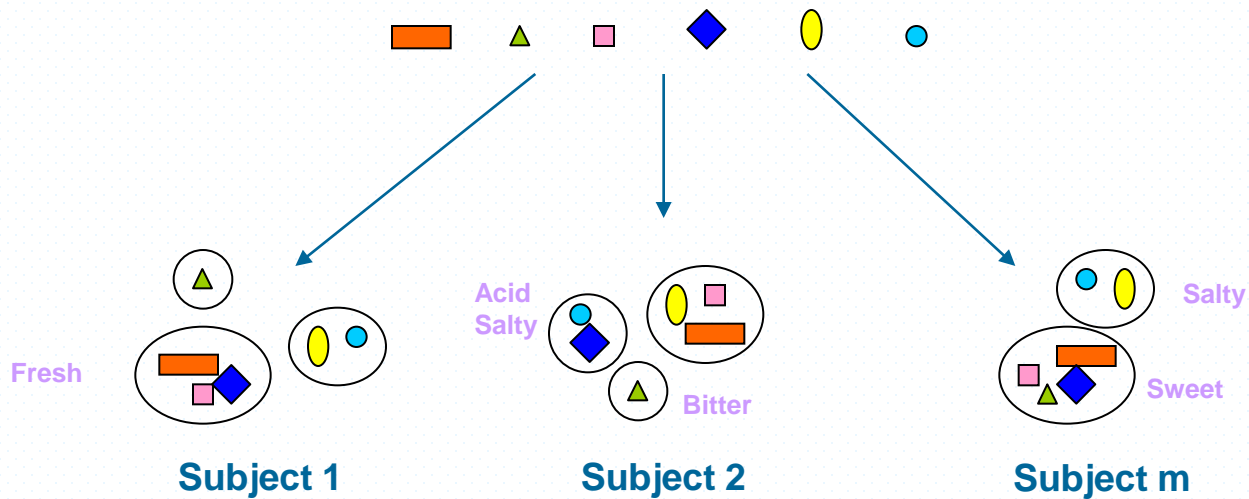


Sorting into categories...

Procedure

n stimuli evaluated by m subjects:

"Please, sort the stimuli in as many groups as you consider necessary with the understanding that stimuli in the same group are perceived as similar"



Interest

- **Categorization is a natural process whereby human beings deal with complexity.**
- **Effective tool to highlight relationships between stimuli.**
- **Several published papers in Food Quality and Preference.**

Statistical analysis

- **Statistical tools to analyze data :**

Factor analytic methods	Cluster analysis
MDS, Multiple Correspondence analysis.	Hierarchical/partitioning cluster analysis; Central partition; Additive trees...

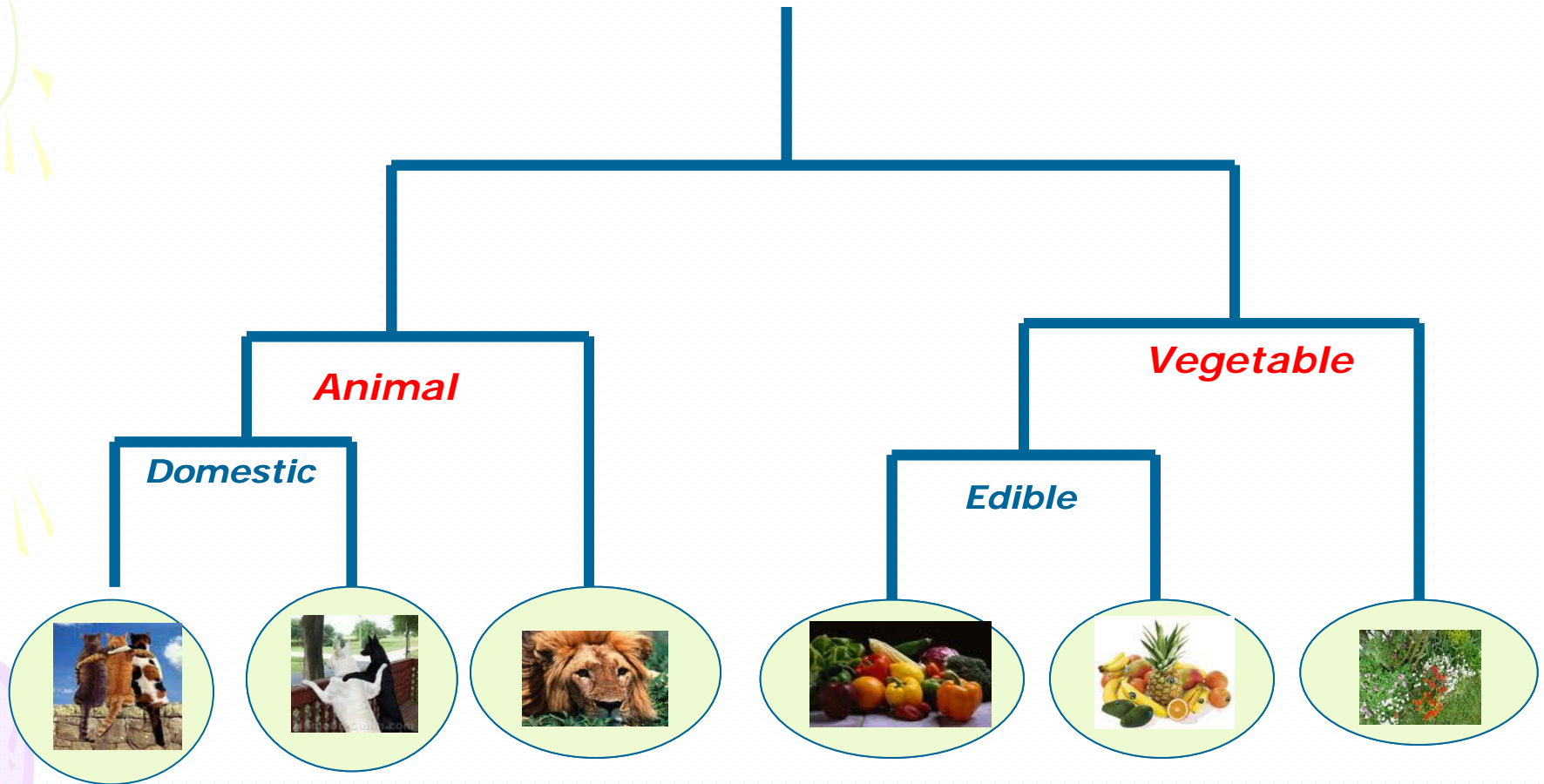
**The underlying principle is that objects within each group are similar (distance=0) and objects in different groups are equidistant.
(inclusive/exclusive strategy)**

Limitations

- *Very often, categories do not have clear-cut boundaries.*
- *The categories are related to each others.*
- *Subjects should be given a chance to assess these facts.*



Towards a taxonomic categorization

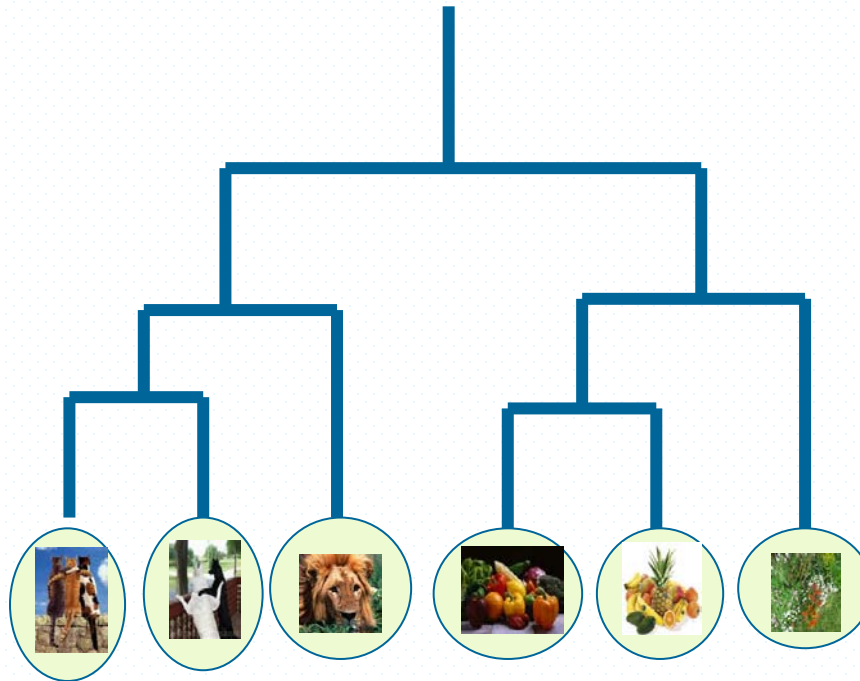


Implications

The taxonomy reflects how the categories are related to each others (class inclusion).

The level of abstraction increases with the level of the hierarchy (from more specific to more abstract categories)

The similarity between stimuli decreases when the level of abstraction increases.



Taxonomic free sorting task

Two stages

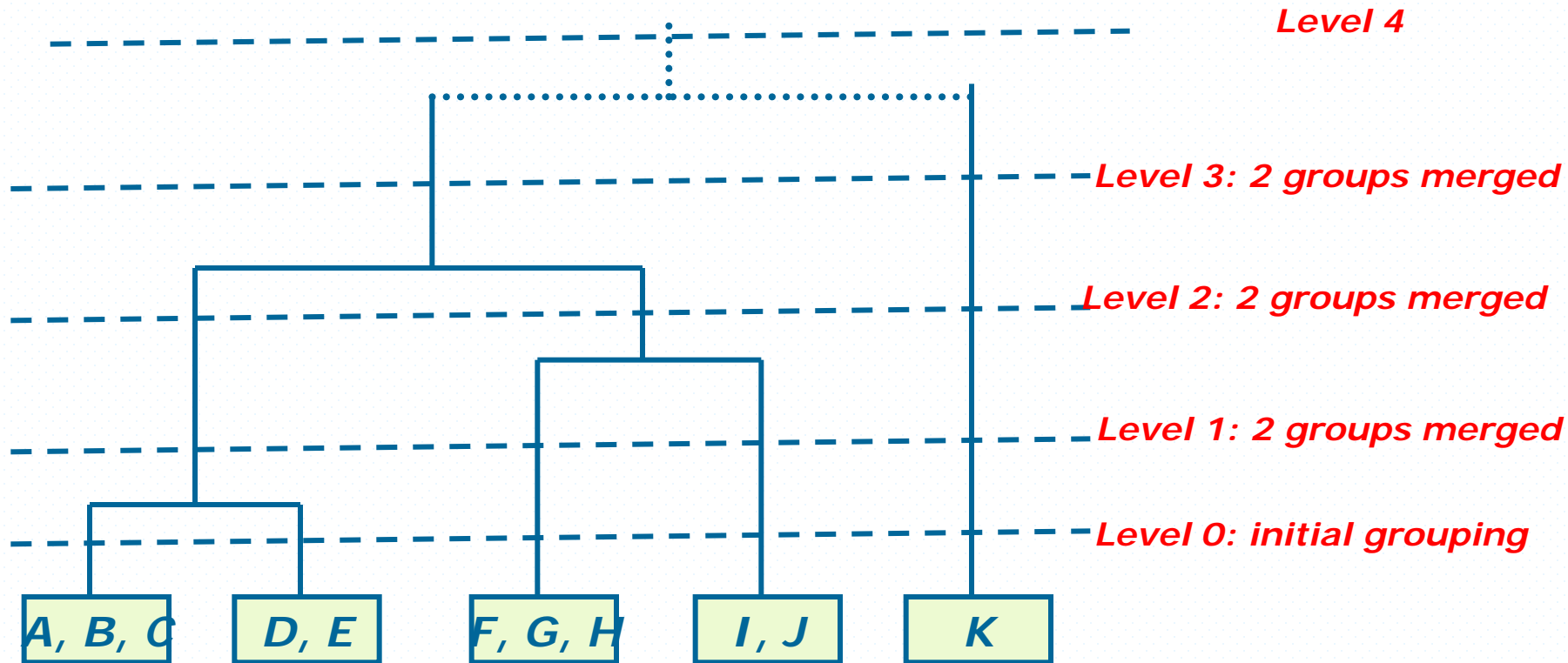
Stage 1

- Subjects form groups of stimuli as in the free sorting task.

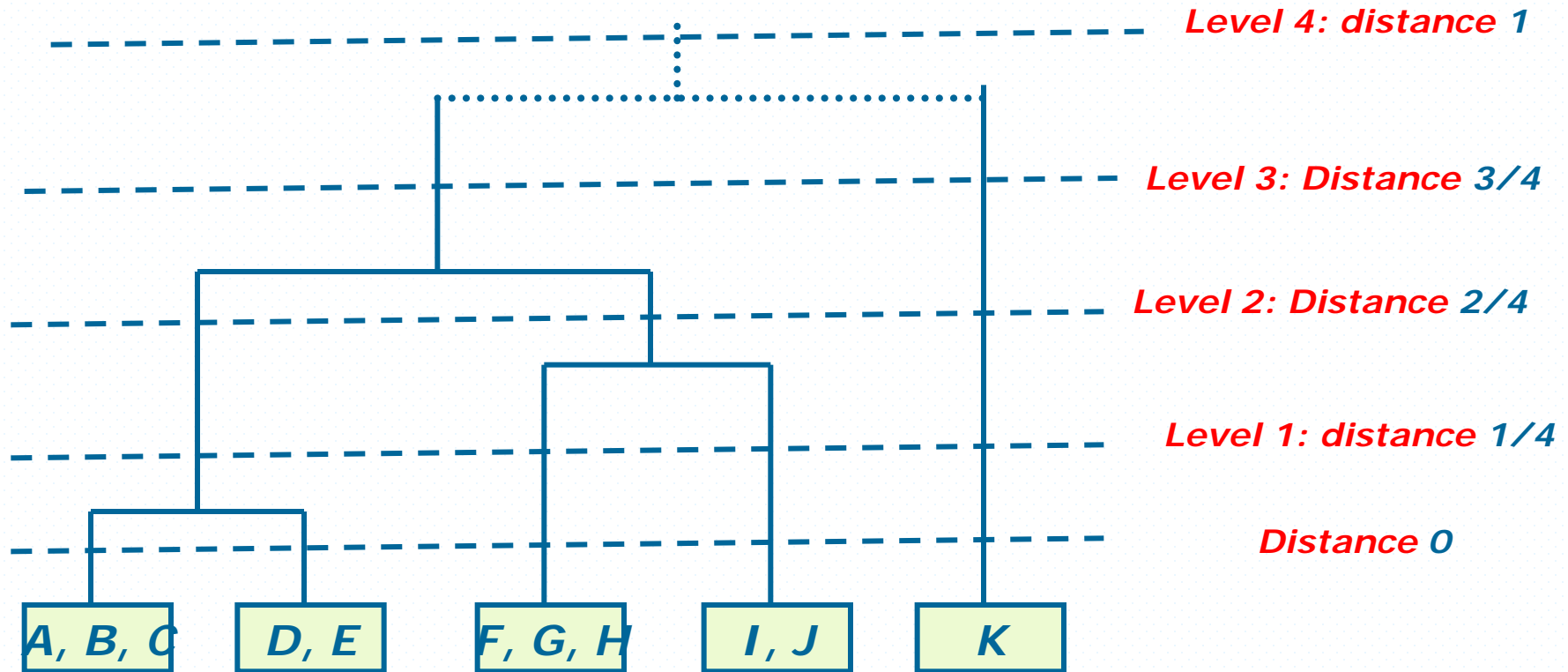
Stage 2

- Subjects are instructed to lump together the two groups which they consider as closest, thus forming a new group.
- This process is reiterated until there are only two groups left.

Cophenetic distances per subject

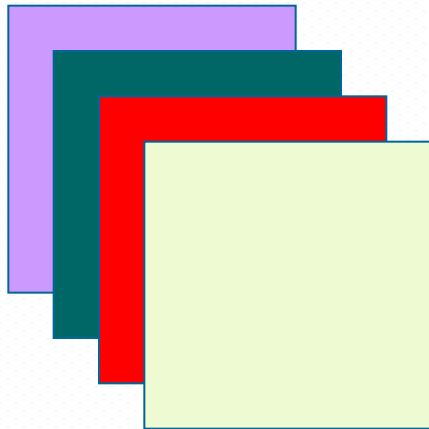


Cophenetic distance per subject



Statistical analysis of data

Cophenetic distances associated with the various subjects



NON METRIC MDS on the average distance matrix

NON METRIC INDSCAL

Cluster analysis

Additive trees

Chocolate data

14 brands of chocolate, conducted among 25 sensory panellists.

Products were sorted by packaging/image



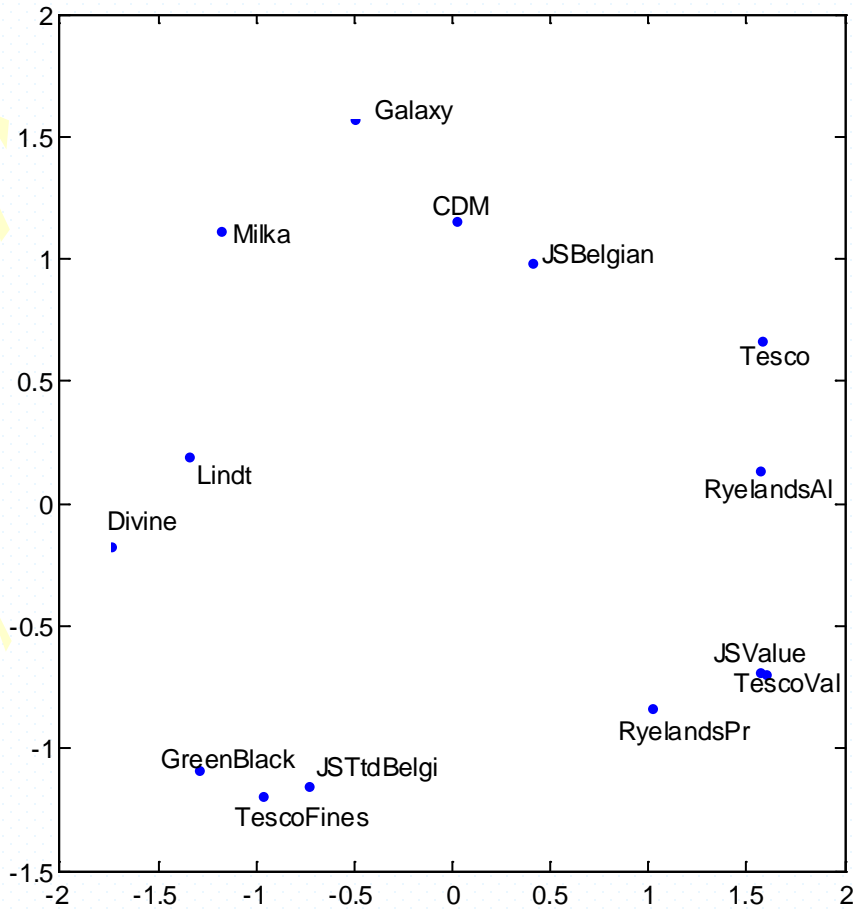
NON METRIC MDS on the average distance matrix

Three dimensional solution

	Free sorting	Taxonomic free sorting
Stress	0.12	0.04
Rank Correlation coefficient	0.83	0.97

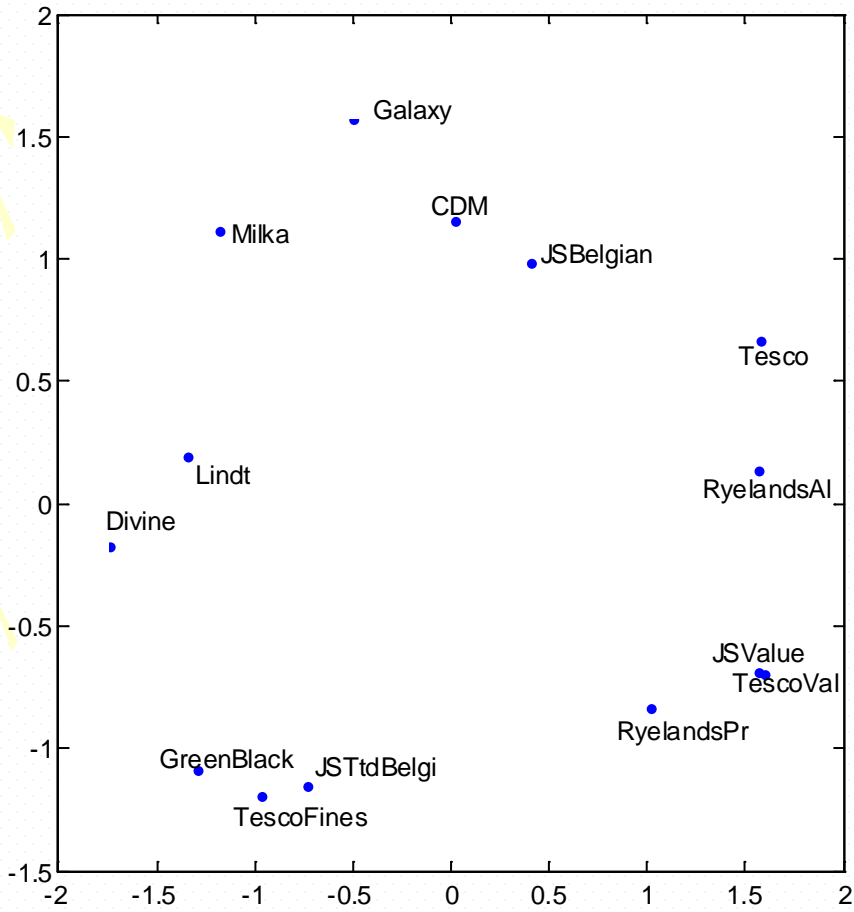
Configuration of the products

Free sorting

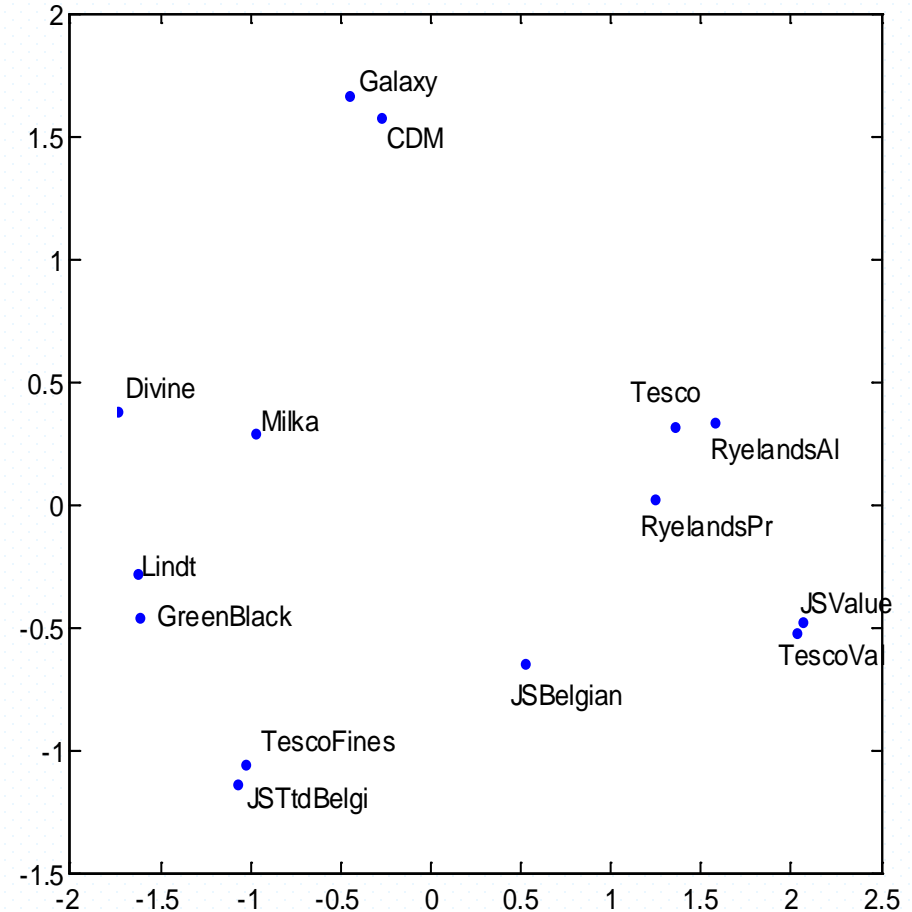


Configuration of the products

Free sorting

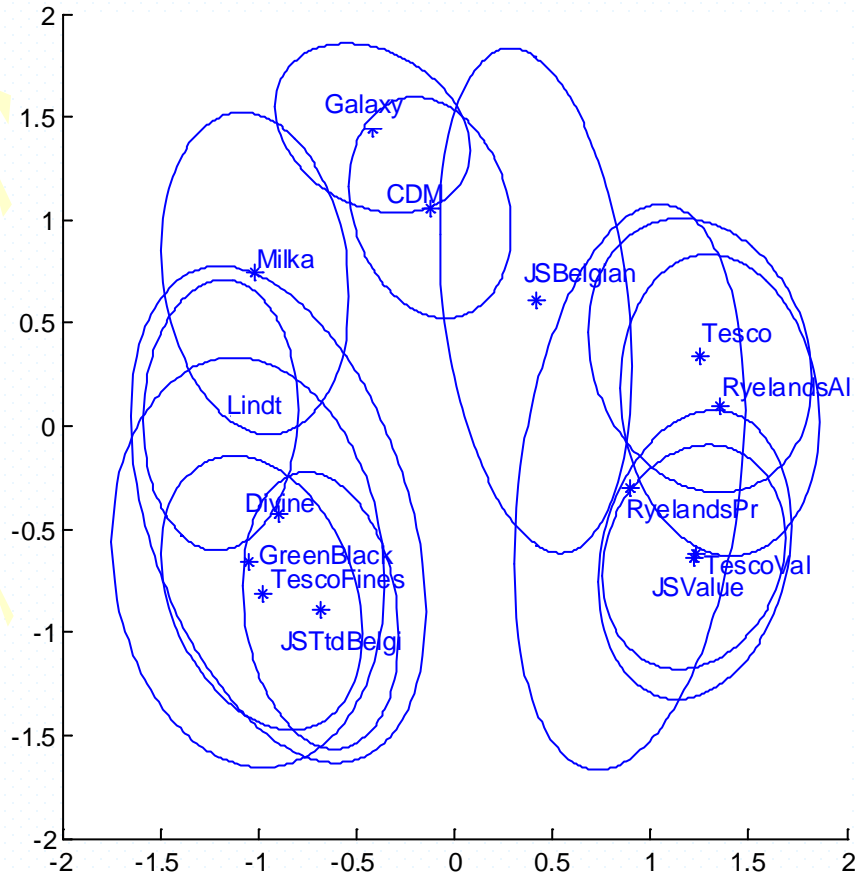


Taxonomic free sorting



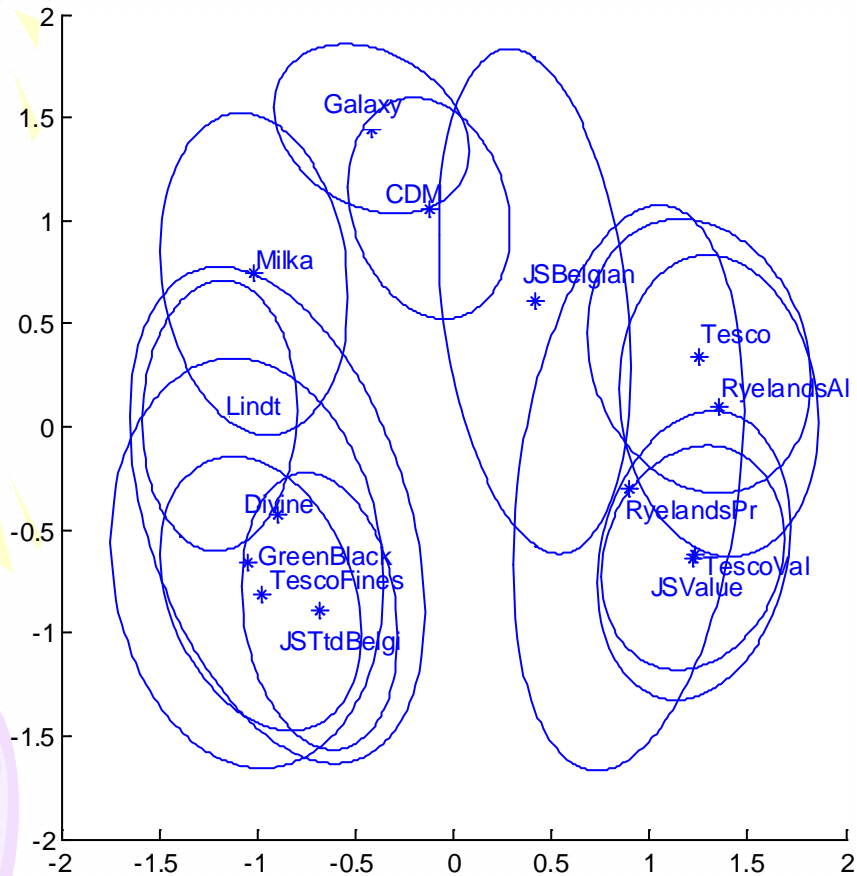
Configuration of the products and confidence ellipses by subjects' resampling (bootstrap)

Free sorting

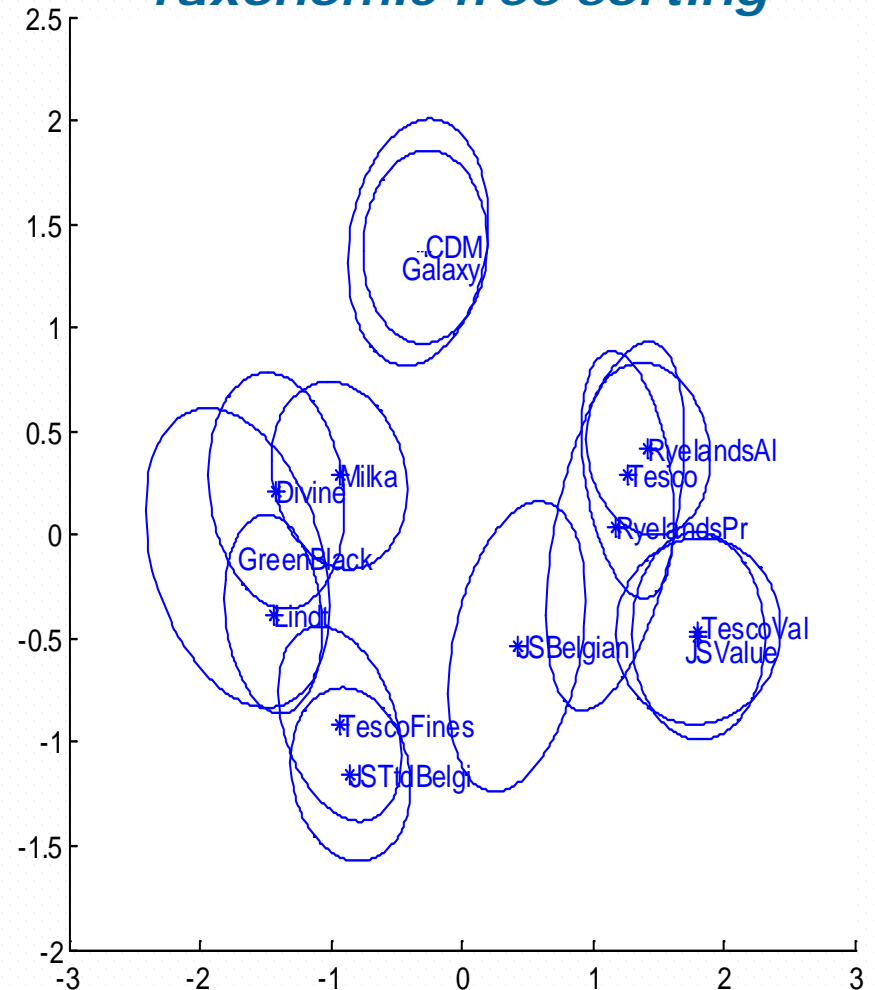


Configuration of the products and confidence ellipses by subjects' resampling (bootstrap)

Free sorting



Taxonomic free sorting



Conclusion

- *Taxonomic free sorting gives a better insight into the relationships among the products than the usual categorization task;*
- **The portrayal of the products on the basis of *MDS* dimensions is more stable;**
- **Further investigations are needed in order to :**
 - Take account of the subjects' differences (e.g. non metric INDSCAL)
- **More generally, this new kind of data poses a real challenge.**