Un effet Gutman en ACP

A Gutman effect in PCA

Claude Flament and Laurent Milland

Electronic version

URL: <http://journals.openedition.org/msh/2946>

DOI: 10.4000/msh.2946

ISSN: 1950-6821

Publisher

Centre d'analyse et de mathématique sociales de l'EHESS

Printed version

Date of publication: 1 September 2005

ISSN: 0987-6936

Electronic reference

Claude Flament and Laurent Milland, « Un effet Gutman en ACP », *Mathématiques et sciences humaines* [Online], 171 | Automne 2005, Online since 10 April 2006, connection on 23 July 2020. URL : <http://journals.openedition.org/msh/2946> ; DOI : <https://doi.org/10.4000/msh.2946>

UN EFFET GUTTMAN EN ACP

Claude FLAMENT¹, Laurent MILLAND²

RÉSUMÉ – *Le recours à des méthodes multidimensionnelles dans le traitement des données permet d'extraire des structures de données unidimensionnelles, lorsqu'elles existent, et l'on parle alors d'effet Guttman. Si la mise au jour d'un tel effet est bien connu en Analyse des Correspondances, elle peut également se réaliser en Analyse en Composantes Principales, ce que nous exposons ici, en étudiant les propriétés particulières de cet effet (composante de taille, composante diagonale, etc.). L'étude d'une telle structure offre une intelligibilité toute particulière aux résultats des recherches conduites dans le domaine des sciences humaines, comme il en va dans les deux exemples présentés ici à propos des représentations sociales de l'homosexualité d'une part et de l'intelligence au masculin et au féminin d'autre part.*

MOTS CLÉS – Analyse en Composantes Principales, Effet Guttman, Représentations sociales

SUMMARY – A Guttman effect in PCA

The recourse to multidimensional methods in the data treatment allows for extracting structures of unidimensional data, when they exist, and one speaks then about Guttman effect. If the disclosure of such an effect is well-known in Correspondences Analysis, it can be also carried out in Principal Component Analysis, as presented here, while studying special properties of this effect (component of size, diagonal component, etc.). The study of such a structure offers a very particular intelligibility to the results of research undertaken in the field of the social sciences, as shown by the two examples presented in connection with the social representation of homosexuality on the one hand and intelligence in its relationship with gender on the other hand.

KEYWORDS – Guttman effect, Principal Component Analysis, Social representations

1. INTRODUCTION

... construire un espace multidimensionnel ... qui révélerait, lorsqu'elle existe, l'échelle unidimensionnelle sous-jacentes aux données

tel est le conseil que donnent Rouanet et Le Roux [1993]³. Et il est d'autant plus aisé de suivre un tel conseil que le chercheur en sciences sociales dispose couramment de logiciels effectuant des traitements multidimensionnels – alors que l'on n'a pas l'équivalent informatique pour des analyses relevant de modèles combinatoires unidimensionnels⁴.

¹ Laboratoire de Psychologie Sociale, Université de Provence Aix-Marseille I, 29 av. Robert Schuman, 13621 Aix-en-Provence cedex 1, flamentclaud@wanadoo.fr.

² Université de Franche-Comté, 30 rue Mégevand, 25030 Besançon cedex, Laboratoire SAvoir COgnition et pratiques sociales, Université de Poitiers, laurent.milland@worldonline.fr.

³ Référence que, par la suite, nous invoquerons sous le sigle RLR

⁴ Les logiciels d'analyse combinatoire des données, mentionnés par Guénoche et Monjardet [1987, p. 39-43], n'ont malheureusement pas été actualisés.

Donc, un peu au hasard, on effectue des analyses multidimensionnelles sur des données *a priori* quelconques – et si une configuration remarquable émerge, on essaie d'en comprendre la nature. Dans la perspective d'*analyse combinatoire des données* qui est la nôtre [cf. Arabie, Hubert, 1992; Guénoche, Monjardet, 1987], il s'agit de trouver un modèle combinatoire compatible avec ce que fait apparaître l'analyse aveugle. Peut-être est-ce la démarche que recommande Benzécri : «Le modèle doit suivre les données, non l'inverse» (cité par RLR dans leur avant-propos).

C'est par ce genre de hasard (mais Pasteur disait que «Le hasard ne favorise que les esprits préparés»), que l'Analyse en Composantes Principales (ACP) d'un petit questionnaire post-expérimental [Martinie, 2000] a fait apparaître la configuration que reproduit la Figure 1 : un classique «Facteur de taille» – ou un nuage de points disposé selon un arc de cercle ? Nous avons opté pour un modèle référant à l'arc de cercle, car, sur la sphère des corrélations, le chemin le plus court d'un point à un autre est un arc de cercle – ce qui correspond à une structure unidimensionnelle – et nous avons parlé d'un «Effet Guttman en ACP» [Flament, Milland, 2003].

Figure 1. Premier plan factoriel de l'ACP réalisée sur les données de Martinie [2000, p. 191]

2. QUEL MODELE DE DONNEES ?

Par analogie avec l'effet Guttman bien connu en Analyse des Correspondances (AC)⁵, on peut s'interroger sur le résultat que donne l'ACP du tableau d'une Echelle de Guttman classique [Matalon, 1965], comme celle que présente le Tableau 1 (i.e., un tableau en $[0,1]$ tel que l'on peut ordonner les lignes et les colonnes de sorte qu'elles laissent apparaître, entre les «1» et les «0» de ce tableau, une frontière en «escalier») ; notons que nous utilisons le «triangle guttmannien», et non le «scalogramme» (Tableau 1 dédoublé) – modèle que RLR [p. 281] doivent utiliser pour leurs calculs en AC.

⁵ Divers auteurs [Kruskal, Wish, 1991; Podani, Miklos, 2002; Brazill, Grofman, 2002] parlent de «Horseshoe phenomenon» (ou «Horseshoe effect»), de «Arch effect» et «Guttman effect» dans des techniques MDS (Multi Dimensional Scaling) et s'interrogent sur le paradoxe d'une structure linéaire unidimensionnelle qui apparaît sous une forme curvilinéaire bidimensionnelle. Considéré comme artéfactuel, des auteurs cherchent à éviter cet effet que nous analysons ici.

	c_1	c_2	c_3	c_4	c_5
l_1	1	1	1	1	1
l_2	0	1	1	1	1
l_3	0	0	1	1	1
l_4	0	0	0	1	1
l_5	0	0	0	0	1
l_6	0	0	0	0	0

Tableau 1. Une échelle de Guttman classique

Figure 2. Premier plan factoriel de l'ACP sur le Tableau 1

La Figure 2 peut suggérer qu'on est sur la bonne voie, mais, en analysant une échelle plus importante (avec 90 colonnes et 91 lignes, par exemple), on constate qu'on s'éloigne fortement d'un arc de cercle (cf. Figure 3).

Figure 3. Segment de quartique obtenue par ACP d'un triangle de Guttman (tableau de 90 colonnes et 91 lignes)

Rouanet (communication personnelle) résout le problème en rapprochant deux exercices figurant dans RLR : l'analyse du scalogramme en AC [p. 277-281] et [p. 282, question 2], l'étude de l'équivalence entre l'ACP d'un tableau en $[0 \square 1]$ et de l'AC du même tableau, mais «*dédoublé*» (c'est le passage du triangle au scalogramme). On en déduit que les points représentatifs sont, dans le premier plan factoriel de l'ACP, sur une quartique de la forme $y^2 = Ax^2 (B - x^2)$. C'est ce qu'on constate sur la Figure 3.

Il s'avère toutefois que les questionnaires auxquels nous avons recours dans nos études offrent généralement plus de deux modalités de réponse. Nous avons donc envisagé une échelle de Guttman *généralisée*, c'est-à-dire, avec plus de deux modalités de réponse par variable [Matalon, 1965, chap. 4 □ Coombs, 1964, p. 229-236 □ cette généralisation semble être ignorée des auteurs plus récents □ cf. McIver, Carmines, 1981]. Le Tableau 2 donne un exemple (repris de Coombs).

	A	B	C
l_1	4	4	4
l_2	3	4	4
l_3	3	3	4
l_4	3	3	3
l_5	3	3	2
l_6	3	2	2
l_7	2	2	2
l_8	1	2	2
l_9	1	1	2
l_{10}	1	1	1
l_{11}	1	1	0
l_{12}	1	0	0
l_{13}	0	0	0

Tableau 2. Une échelle de Guttman généralisée

On peut considérer cela comme le modèle combinatoire correspondant à un *facteur de taille* □ les lignes sont hiérarchisées en ce sens qu'une ligne est supérieure à une autre si et seulement si elle comporte, dans chaque colonne, une valeur supérieure ou égale à celle de l'autre – dans le vectoriel des n-tuples, les lignes du tableau constituent alors une partie totalement ordonnée. On peut décomposer un tel tableau : si une variable X a K niveaux codés i , de 0 à $(K-1)$, on construit $(K-1)$ variables binaires X_i (pour $i > 0$), en posant :

$$X_i = 0 \quad \text{ssi} \quad X < i, \quad \text{et} \quad : X_i = 1 \quad \text{ssi} \quad X \geq i$$

On recompose par sommation. C'est ainsi que se correspondent les Tableaux 2 et 3. On remarque que le tableau décomposé a nécessairement la structure d'une échelle de Guttman classique. L'ordre des colonnes du Tableau 3 résulte de l'organisation des variables selon l'ordre Guttmanien. Donc, à partir d'une échelle classique (assez vaste – segment de quartique. L'exemple du Tableau 2 est très particulier, en ce sens que, pour chaque niveau i (cf. Tableau 3), la colonne B est toujours entre A et C : on a, soit $A_i > B_i > C_i$, soit l'inverse, selon l'ordre total que présente nécessairement les colonnes d'une échelle parfaite. L'examen de nombreux exemples nous suggère que l'on se rapproche d'autant plus de l'arc de cercle, qu'augmente le nombre de colonnes

respectant un *enchaînement* particulier, pris tantôt dans un sens, tantôt dans l'autre. Et c'est cet enchaînement qu'on retrouve dans la succession des points sur l'arc de cercle. Il y a là, peut-être, une voie de recherche à explorer – mais la présence d'un enchaînement net dans une échelle généralisée reste un cas particulier. Ce cas particulier nous suggère des propriétés qui se révéleront importantes pour la suite de notre étude.

	A_1	B_1	C_1	C_2	B_2	A_2	A_3	B_3	C_3	C_4	B_4	A_4
l_1	1	1	1	1	1	1	1	1	1	1	1	1
l_2	0	1	1	1	1	1	1	1	1	1	1	1
l_3	0	0	1	1	1	1	1	1	1	1	1	1
l_4	0	0	0	1	1	1	1	1	1	1	1	1
l_5	0	0	0	0	1	1	1	1	1	1	1	1
l_6	0	0	0	0	0	1	1	1	1	1	1	1
l_7	0	0	0	0	0	0	1	1	1	1	1	1
l_8	0	0	0	0	0	0	0	1	1	1	1	1
l_9	0	0	0	0	0	0	0	0	1	1	1	1
l_{10}	0	0	0	0	0	0	0	0	0	1	1	1
l_{11}	0	0	0	0	0	0	0	0	0	0	1	1
l_{12}	0	0	0	0	0	0	0	0	0	0	0	1
l_{13}	0	0	0	0	0	0	0	0	0	0	0	0

Tableau 3. Décomposition de l'Echelle généralisée du Tableau 2

Les Tableaux 4 et 5 reprennent l'échelle du Tableau 2, en faisant apparaître les scores factoriels $\square_{F1}$ et $\square_{F2}$ après ACP.

	A	B	C	$\square_{F1}$	somme
l_1	4	4	4	-1.550	12
l_2	3	4	4	-1.270	11
l_3	3	3	4	-1.010	10
l_4	3	3	3	-.780	9
l_5	3	3	2	-.540	8
l_6	3	2	2	-.280	7
l_7	2	2	2	0	6
l_8	1	2	2	.279	5
l_9	1	1	2	.541	4
l_{10}	1	1	1	.775	3
l_{11}	1	1	0	1.009	2
l_{12}	1	0	0	1.270	1
l_{13}	0	0	0	1.549	0
Somme	26	26	26		

Tableau 4. La composante de taille de l'échelle du Tableau 2

	A	B	C	$\square_{F2}$	pente
l_8	1	2	2	1.471	.5
l_9	1	1	2	1.356	.5
l_2	3	4	4	.975	.5
l_3	3	3	4	.860	.5
l_{13}	0	0	0	.496	0
l_{10}	1	1	1	.248	0
l_7	2	2	2	0	0
l_4	3	3	3	-.250	0
l_1	4	4	4	-.500	0
l_{11}	1	1	0	-.860	-.5
l_{12}	1	0	0	-.980	-.5
l_5	3	3	2	-1.360	-.5
l_6	3	2	2	-1.470	-.5
Enchaînement	1	2	3		

Tableau 5. La composante diagonale de l'échelle du Tableau 2

Le Tableau 4 montre que les scores $\square_{F1}$ hiérarchisent les lignes exactement comme leurs sommes : c'est ce que nous appellerons la *composante de taille* (qui diffère du facteur de taille, cf. note 8). Le Tableau 5 organise les lignes de la même échelle selon les scores $\square_{F2}$ ayant codé (1 – 2 – 3) l'enchaînement repéré ci-dessus (dernière ligne du Tableau 5), et ayant calculé la pente de la régression linéaire de chaque ligne sur cet enchaînement (dernière colonne du Tableau 5)⁶, on constate que les lignes contribuant le plus au facteur $\square_{F2}$, sont, d'une part, les lignes croissantes selon l'enchaînement, et, d'autre part, les lignes décroissantes. C'est ce que nous appellerons la *composante diagonale*⁷. À notre connaissance, ce phénomène n'a pas été décrit en ACP.

Mais ce modèle n'a pas la généralité dont nous avons besoin pour rendre compte des divers résultats empiriques que nous avons rencontrés, et qui comportent souvent des *corrélations négatives* (parfois extrêmes) ! Or, les corrélations entre colonnes d'une échelle de Guttman généralisée sont nécessairement positives !

Essayons de préciser ce que nous cherchons : un type de données inscrites dans un tableau de n lignes et p colonnes, telles que, *de façon naturelle*, les points-colonnes se positionnent sur un arc de grand cercle d'une hyper-sphère de rayon quelconque, isomorphe à la « sphère des corrélations » alors, l'ACP standard de ce tableau projetera *ces points-colonnes sur le cercle des corrélations*⁸. Pour que les vecteurs-

⁶ C'est pour des raisons de facilité calculatoire que nous faisons appel à la pente. On remarquera par ailleurs que cette pente peut se calculer en régressant la ligne sur l'enchaînement mais peut également l'être, comme nous le verrons par la suite, en régressant la ligne sur le deuxième vecteur des saturations.

⁷ Il s'agit bien ici de deux composantes de l'effet Guttman (taille et diagonale) qui ne doivent pas être confondues avec les composantes de l'ACP.

⁸ Une échelle de Guttman de classique peut se présenter comme une chaîne sur un hypercube à p

colonnes soient positionnés naturellement sur une sphère, il est nécessaire que les colonnes contiennent des *variables solidarisées* [cf. RLR, p. 62] – et, plus précisément, que chacune des p colonnes contienne n valeurs telles que moyenne et variance soient identiques d'une colonne à l'autre – car l'ACP standard travaille, en fait, sur des variables centrées et réduites. Lebart, Morineau et Piron [1997, référence que nous citerons sous le sigle LMP] considèrent [p. 52] que cela «est particulièrement bien adapté à la distribution normale» mais le Psychologue ne rencontre des courbes normales guère plus souvent qu'il ne rencontre des licornes [Micceri, 1985]. Nous pensons alors à une chaîne de p permutations de n objets.

À chaque permutation est associée un vecteur qui a pour coordonnées les rangs des objets, sans transformation [Degenne, 1972] ou avec [Benzécri, 1971]. C'est un polyèdre à $n!$ sommets qui permet de représenter géométriquement le permutoèdre dans un espace à $(n - 1)$ dimensions. Les sommets de ce polyèdre appartiennent à une hypersphère, qui, dans le cas d'une chaîne minimale de permutations de 4 objets (24 sommets) est une sphère à 3 dimensions.

De la sorte, une chaîne minimale reliant plusieurs permutations est généralement très proche d'un arc de grand cercle de cette sphère⁹.

	<i>abdc</i>	<i>adbc</i>	<i>adcb</i>	<i>dacb</i>	<i>dcab</i>
<i>a</i>	1	1	1	2	3
<i>b</i>	2	3	4	4	4
<i>c</i>	4	4	3	3	2
<i>d</i>	3	2	2	1	1

Tableau 6. Ensemble de 5 permutations sur 4 objets

Le Tableau 6 présente 5 permutations sur 4 objets. Ce qui saute aux yeux, dans ce type de tableau, c'est la répartition des valeurs dans chaque ligne. On a ici $x_{ij} \leq x_{ik}$ pour les lignes A et B, et $x_{ij} \geq x_{ik}$ pour les lignes C et D. Autrement dit, en lisant le tableau de gauche à droite, les valeurs dans les lignes croissent ou décroissent systématiquement.

Ainsi, en organisant le tableau de données de telle sorte que les vecteurs-colonne respectent l'enchaînement, on voit apparaître cette propriété des lignes du tableau et l'on parlera alors de *lignes congruentes*¹⁰.

dimensions, mais cette chaîne (une géodésique) semble s'éloigner de l'arc de grand cercle de l'hypersphère circonscrite (cf. Figure 3) à cet hypercube pour peu que p soit suffisamment grand (il en est de même pour les tresses).

⁹ On rappellera au passage qu'il existe une multiplicité de chaînes minimales permettant de relier deux sommets du polytope contrairement à ce qui se produit en géométrie euclidienne. En s'intéressant au «dépliage» (et plus généralement aux modèles non monotone [Flament, 1963]), on pourrait se limiter aux chaînes blackiennes [Black, 1948–Degenne, 1972] mais ce n'est qu'un cas particulier–ABC-BAC-BCA-CBA est un chaîne blackienne, mais ABC-ACB-CAB-CBA qui ne l'est pas conduit au même effet Guttman.

¹⁰ On affirme par là même que la notion de composante de taille est différente de celle de facteur de taille. Dans ce dernier cas, les lignes sont homogènes et, «[E] pour un individu, une variable prend une valeur forte, toutes les autres variables prennent également des valeurs fortes» [LMP, p. 56]. Dans le cadre de l'effet Guttman étudié ici, l'homogénéité et la congruence étant généralement approximatives, il est possible qu'une même ligne contribue aux deux composantes.

Dans une perspective combinatoire, la distance entre deux vecteurs-colonnes (i.e., permutations) est donnée par la distance de Kendall – correspondant au nombre minimum d'arêtes qu'il faut suivre pour relier ces deux permutations. Cet indice de distance entre deux permutations conduit au calcul du τ de Kendall qui permet d'évaluer la similitude entre ces deux permutations. Si l'on se place dans le cadre de la géométrie euclidienne, et donc au niveau de la sphère à $(n-1)$ dimensions, la distance τ entre deux permutations (distance euclidienne qui, dans le contexte permutatif est généralement appelée distance de Spearman [Degenne, 1972]) conduit à définir le ρ de Spearman (coefficient linéaire de Bravais et Pearson), qui permet d'évaluer autrement la similitude entre ces deux permutations.

On sait, ainsi que nous le rappelons plus haut, qu'un ensemble quelconque de permutations sur n objets peut être très exactement situé sur une hypersphère à $(n-1)$ dimensions. Et il s'avère que, dans certains cas, l'ensemble de permutations considéré – c'est le cas du Tableau 6 par exemple – ne soit pas quelconque.

On observe alors que les permutations de l'ensemble considéré se répartissent sur un «demi grand cercle». On passe ainsi d'un espace à $(n-1)$ dimensions à un espace à 2 dimensions. On obtiendra donc ce que nous recherchions initialement, à savoir que dans le cas (improbable) où on aura une répartition parfaite sur un demi-grand cercle, alors l'ACP standard projettera ces points exactement sur le cercle des corrélations, les deux premiers facteurs expliquant 100 % de la variance.

Figure 4. Permutoèdre et ACP (cf. texte)

Nous pensons que la Figure 4 fait voir ce que nous admettrons le passage (cf. Figure 4-c) entre une chaîne de Kendall telle que celle donnée par les vecteurs-colonnes du Tableau 6 (cf. Figure 4-a) et un arc de cercle de la sphère des corrélations telle qu'il nous est présenté au travers du graphe (cf. Figure 4-b) de l'ACP du Tableau 6.

L'arc de grand cercle repéré par l'ACP (cf. Figures 4-b et 4-c) est calculé à partir des 5 permutations de la chaîne identifiée sur la Figure 4-a), mais, si on remplace *adcb* par *dacb* l'on obtient exactement les mêmes valeurs (cf. note 8, p. 30 ; le dessin de la Figure 4-c nous semble représentatif de cette équivalence).

L'ACP du Tableau 6 montre que les deux premiers facteurs expliquent, ensemble, 97.2 % de la variance – ce qui atteste bien du fait que ces permutations sont presque parfaitement sur le cercle des corrélations, comme l'illustre le graphe de la Figure 4-b. La répartition de ce pourcentage entre les deux facteurs (57.2 % et 40 %) est sans intérêt pour notre problème. En revanche, on peut, à l'instar de Falissard [1995], considérer que l'écart de 2.8 % entre ce pourcentage de variance expliquée par les deux premiers facteurs et celui de 100 %, cas parfait où l'ensemble des vecteurs-colonnes se projetterait exactement sur le cercle des corrélations, est une indication de la bonne approximation de ce que nous recherchions au début : « Un type de données inscrites dans un tableau de n lignes et p colonnes, telles que, *de façon naturelle*, les points-colonnes se positionnent sur un arc de grand cercle d'une hyper-sphère de rayon quelconque, isomorphe à la *sphère des corrélations* ».

On peut également identifier une autre propriété liée à ce type de données. Elle apparaît avec la matrice de corrélation (□ de Spearman ici, mais celle du □ de Kendall conserve cette propriété) comme celle calculée à partir du Tableau 6.

	<i>abdc</i>	<i>adbc</i>	<i>adcb</i>	<i>dacb</i>	<i>dcab</i>
<i>abdc</i>	1	.80	.40	.00	-.60
<i>adbc</i>	.80	1	.80	.60	.00
<i>adcb</i>	.40	.80	1	.80	.40
<i>dacb</i>	.00	.60	.80	1	.80
<i>dcab</i>	-.60	.00	.40	.80	1

Tableau 7. Matrice de corrélations calculée à partir du Tableau 6

La matrice de corrélation présentée dans le Tableau 7 s'organise de telle sorte que, dans chaque ligne et chaque colonne de cette matrice, en s'éloignant de la diagonale, les corrélations sont décroissantes, et l'on parle alors de matrice (ou tableau cf. [Guénoche, Monjardet, 1987]) de Robinson [1951] et par extension de matrice Robinsonnienne lorsqu'il est possible d'identifier cette propriété). Cette organisation ne doit pas être confondue avec la « Structure hiérarchique » de Spearman [RLR, p. 194]¹¹. On remarquera par ailleurs que l'enchaînement des colonnes [Arabie, Hubert, 1992, parlent de « Sérialisation »] peut, non seulement « Se voir » dans le premier plan factoriel, mais aussi, « Se lire » dans cette matrice de corrélation.

¹¹ La décroissance des corrélations en partant de la diagonale s'observe simultanément sur les lignes et colonnes dans une matrice robinsonnienne. Dans une matrice *symétrique* de corrélations (par opposition aux matrices de corrélations dont seule une moitié est présentée) la structure hiérarchique de Spearman s'observe lorsque l'on peut organiser cette matrice, de telle sorte que, ne tenant pas compte de la diagonale de cette matrice, les valeurs décroissent à l'intérieur des colonnes.

Dans un tableau, si dans la ligne i , on a, pour 3 colonnes j, k, l

$$x_{i,j} > x_{i,k} > x_{i,l}$$

Cette ligne tend à ordonner les corrélations $r_{j,l} \geq r_{j,k}, r_{k,l}$ – puisque $r_{j,l}$ est fonction inverse de $(x_{i,j} - x_{i,l})^2$. Donc, si les colonnes présentent un enchaînement et si toutes les lignes sont congruentes, la matrice des corrélations aura une structure Robinsonnienne [Robinson, 1951] selon cet enchaînement. L'essentiel de cette structure peut être préservé, même si quelques lignes ne sont pas congruentes

Donc, on parle d'effet Guttman en ACP, lorsque l'ACP d'un tableau de données nous offre un premier plan factoriel pour lequel la forme prise par le nuage des points-colonnes avoisine le cercle des corrélations, ce qui correspond à un pourcentage de variance expliqué par les deux premiers facteurs proche de 100 %. Cela entraîne une matrice de corrélation robinsonnienne (ou presque parfaitement robinsonnienne). On verra avec l'équation de décomposition des scores que ceux relatifs à l'axe 1 nous donnent des indications sur la composante de taille de cet effet alors que les scores relatifs à l'axe 2 nous donnent des indications sur sa composante diagonale.

3. QUEL TYPE DE CALCULS

Il se peut toutefois que la lecture de ces deux composantes guttmaniennes ne soit pas aussi aisée. C'est le cas des données d'un exemple artificiel et que nous présentons dans le Tableau 8.

	<i>abcdefgh</i>	<i>abhcdefg</i>	<i>habcgdef</i>	<i>hgabcfd</i>	<i>hgfabecd</i>	<i>hgfeadbc</i>	r_{F1}	r_{F2}	moy.	pente
<i>A</i>	1	1	2	3	4	5	.302	-1.392	2.670	.860
<i>B</i>	2	2	3	4	5	7	.656	-.939	3.830	.960
<i>C</i>	3	4	4	5	7	8	.972	-.269	5.170	.960
<i>D</i>	4	5	6	7	8	6	.949	.598	6.000	.570
<i>E</i>	5	6	7	8	6	4	.421	1.106	6.000	-.110
<i>F</i>	6	7	8	6	3	3	-.390	1.218	5.500	-.830
<i>G</i>	7	8	5	2	2	2	-1.262	.544	4.330	-1.310
<i>H</i>	8	3	1	1	1	1	-1.647	-.865	2.500	-1.170

Enchaînement	1	2	3	4	5	6
--------------	---	---	---	---	---	---

Tableau 8. Analyse de 6 permutations sur 8 objets

L'analyse du Tableau 8 montre que les deux premiers facteurs expliquent, ensemble, 91,2 % de la variance – ce qui, à l'appui de la Figure 5, montre bien que ces permutations sont presque parfaitement sur le cercle des corrélations. La matrice des corrélations quant à elle est parfaitement robinsonnienne (Tableau 9).

Figure 5. Premier plan factoriel de l'ACP du Tableau 8

	<i>abcdefgh</i>	<i>abhcddefg</i>	<i>habcgdef</i>	<i>hgabcfde</i>	<i>hgfabecd</i>	<i>hgfeadbc</i>
<i>abcdefgh</i>	1	.64	.19	-.24	-.60	-.83
<i>abhcddefg</i>	.64	1	.79	.31	-.10	-.40
<i>habcgdef</i>	.19	.79	1	.79	.33	.00
<i>hgabcfde</i>	-.24	.31	.79	1	.76	.43
<i>hgfabecd</i>	-.60	-.10	.33	.76	1	.83
<i>hgfeadbc</i>	-.83	-.40	.00	.43	.83	1

Tableau 9. Matrice de corrélations calculées à partir du Tableau 8

On peut donc parler dans le cas présent d'un effet Guttman résultant du Tableau 8. Toutefois, l'arc de grand cercle, tel qu'il est visible dans la Figure 5, ne se présente pas comme ceux, «vertical à droite», des Figures 1, 2 ou 4. Pratiquement, nous aurions intérêt à standardiser la mise en évidence de nos composantes guttmaniennes, en accord avec les formules ci-dessous. Si l'on considère un tableau de données T avec n lignes et p colonnes solidarisées dont l'ACP met en évidence les deux composantes guttmaniennes. Soit l_i la $i^{\text{ème}}$ ligne de ce tableau, $\square_{\square}$ le $\square^{\text{ème}}$ vecteur des saturations et $\square_{\square}$ le $\square^{\text{ème}}$ vecteur des scores¹² produits par l'ACP de T

$$\square(\square_{\square}, l_i) = \square l_i \square \square_{\square, i}$$

Pour avoir un calcul simple nous permettant de connaître la congruence de l_i avec l'enchaînement des colonnes, on calcul la pente de la régression linéaire de l_i sur $\square_{\square}$. On a alors

¹² On s'appuie ici sur une formule des scores donnée par LMP à une transformation linéaire près.

$$a_{(\square_{\square}, l_i)} = \frac{\text{cov}(\square_{\square, i}, l_i)}{\text{var}(\square_{\square, i})} = \frac{\square_{(i \square \square_{\square, i})} \square \square_{\square, i} \square \frac{\square l_i}{p}}{p \square \text{var}(\square_{\square, i})}$$

La coordonnée de l_i sur $\square_{\square}$ peut ainsi s'écrire

$$\square_{(\square_{\square}, l_i)} = p \square \text{var}(\square_{\square, i}) \square a_{(\square_{\square}, l_i)} + \square_{li} \square \square_{\square, i}$$

Où $\square_{li}$ est la moyenne des valeurs qui composent l_i . Lorsque l'ACP fait apparaître un effet Guttman avec un arc de cercle *vertical* – comme nous l'avons obtenu dans les Figures 1, 2 ou 4 – alors, dans le premier plan factoriel $\text{var}(\square_{1, i})$ est faible et $\square_{\square_{1, i}}$ est forte tandis que $(\square_{2, i})$ est forte et $\square_{\square_{2, i}}$ est faible, souvent proche de 0.

De ce qui précède, on en conclut que les scores du facteur 1, comme les moyennes des lignes, permettent de hiérarchiser ces lignes et de mettre en évidence la composante de taille. Par ailleurs, les scores du facteur 2 vont hiérarchiser les lignes de telle sorte que celles qui contribueront le plus à ce facteur sont croissantes (pente positive) ou décroissantes (pente négative) selon le deuxième vecteur des saturations. C'est donc la hiérarchisation des lignes selon les scores du facteur 2 qui permet de dégager la composante diagonale. On comprend dès lors l'intérêt qu'il y a à retrouver systématiquement l'arc de cercle en position verticale.

Ailleurs, nous avons introduit une rotation amenant le centre de gravité du nuage des points-colonnes sur la partie droite du premier axe [Flament, Milland, 2003]. La rotation bissectrice que nous proposons ici s'appuie sur la bissectrice de l'angle au centre des deux points extrêmes du nuage des points-colonnes. Elle a l'avantage de permettre une lecture systématique de la composante diagonale à partir du facteur 2 et de la composante de taille à partir du facteur 1. La représentation graphique s'en trouve également standardisée. En outre, cette rotation pallie les cas sensibles où le centre de gravité du nuage des points-colonnes est situé vers l'une des extrémités de l'arc de cercle.

Soit T un tableau de données dont les points-colonnes, après ACP, se positionnent sur un arc de grand cercle d'une hyper-sphère de rayon quelconque, isomorphe à la «Sphère des corrélations». Soit F la matrice (p lignes $\square$ 2 colonnes) des coordonnées du nuage des points-colonnes dans le premier plan factoriel. Pour faciliter nos calculs ultérieurs et appliquer les propriétés du cercle trigonométrique, on norme¹³ les vecteurs de chaque point-colonne de F , et l'on passe de F à F^* (matrice d'ordre $p \square 2$). Soit X et Y les points extrêmes du nuage des points-colonnes, tels que X soit distinct de Y et X ne soit pas diamétralement opposé à Y $\square 1 < \square_{X, Y} < +1$. Les coordonnées de X et Y , après projection, sont les suivantes

$$F_{X, Y}^* = \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}$$

¹³ En appliquant la formule habituelle $\square_i^* = \frac{\square_i}{\sqrt{\square_1^2 + \square_2^2}}$.

La rotation que nous proposons est telle que le nouvel axe 1 porte la bissectrice de l'angle au centre défini par X et Y .

Soit, dans F^* , les vecteurs

$$X = [x_1, x_2] \quad \text{et} \quad Y = [y_1, y_2]$$

On pose :

$$S_i = x_i + y_i$$

Le vecteur somme de X et Y est

$$(X + Y) = [S_1, S_2]$$

Après projection de X et Y , X et Y sont de longueur égale et le quadrilatère $\{X, O, Y, (X + Y)\}$ est un losange, où le vecteur somme $(X + Y)$, diagonale de ce losange, est porté par la bissectrice de l'angle $\widehat{X, O, Y}$ (cf. Figure 6).

Figure 6. Quadrilatère obtenu par addition vectorielle de X et Y

Le vecteur $(X + Y)$ coupe le cercle des corrélations en un point B . Soit $\square$, l'angle défini par $(X + Y)$ et l'axe F1. Les coordonnées de B sont alors

$$B = [\cos \square, \sin \square]$$

Avec

$$\begin{cases} \cos \square = \frac{S_1}{\sqrt{S_1^2 + S_2^2}} \\ \sin \square = \frac{S_2}{\sqrt{S_1^2 + S_2^2}} \end{cases}$$

Pour amener ce vecteur sur l'axe 1, on procède alors à la rotation définie par la matrice de passage

$$P = \begin{vmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{vmatrix}$$

Il nous faut donc repérer les points X et Y . Soit $\mathbf{v}_k^*$ le $k^{\text{ième}}$ vecteur des saturations obtenu après ACP d'un tableau T et $\mathbf{v}_{k,k}^*$ la coordonnée du $k^{\text{ième}}$ point variable sur ce vecteur des saturations.

Pour repérer X , on construit la matrice (d'ordre $p \times p$) des cosinus des angles au centre des points-colonnes en réalisant le produit de F^* par sa transposée ${}^tF^*$. Soit R cette matrice

$$R = F^* \times {}^tF^*$$

Dans la mesure où l'on parle ici d'effet Guttman si l'on a un arc de cercle, décrit par le nuage des points-colonnes, qui ne dépasse pas 180° (chemin le plus court sur la sphère pour aller d'un point variable à un autre), on repère dans la matrice R la plus faible valeur, correspondant à la corrélation entre X et Y (l'ordre étant, ici, non significatif).

Nous disposons actuellement de nombreux exemples auxquels nous avons appliqué l'analyse proposée ici, exemples dont la mise en évidence des deux composantes de l'effet Guttman apportent une nouvelle lecture des résultats tels qu'ils sont habituellement interprétés à partir d'analyses plus classiques. Certains de ces exemples ont été traités par ailleurs [Flament, Milland, 2003] en appliquant un type de rotation différente, basée sur le centre de gravité du nuage des variables. Les deux nouveaux exemples que nous proposons ici ont pour objectif d'illustrer l'application d'une recherche de l'effet Guttman dans des données réelles, en montrant l'intérêt d'une lecture des résultats sur la base des deux composantes de cet effet après rotation bissectrice.

PREMIER EXEMPLE LA REPRÉSENTATION SOCIALE DE L'HOMOSEXUALITÉ

Le premier exemple que nous retenons à trait à la représentation sociale de l'homosexualité. Le questionnaire, composé d'une liste de 31 traits plus ou moins sexués, a été administré à 70 hommes homosexuels et 70 hommes hétérosexuels [Rallier, Ricou, 2000]. Tous les sujets devaient, dans un premier temps, se décrire à partir de cette liste de traits, en se positionnant à chaque fois sur une échelle allant de 1 (= négatif) à 7 (= positif). Après avoir réalisé cette auto-description, les sujets devaient répondre à ce même questionnaire «Comme le feraient les X en général», la cible «X» pouvant être les hommes, les femmes, ou les homosexuels. Nous disposons ainsi de 8 profils moyens,¹⁴ qui se définissent à partir de la combinaison entre les caractéristiques des répondants et les consignes données pour remplir les questionnaires. Nous travaillons ici sur un extrait des données complètes (15 traits), extrait qui respecte scrupuleusement le type de résultat obtenu sur l'ensemble des 31 traits de l'étude.

¹⁴ Nous avons déjà réalisé ce type d'analyse sur les données de cette même étude [Flament, Milland, 2003], mais en ne disposant à l'époque que de 4 profils, ceux correspondant aux répondants homosexuels.

Pour faciliter le repérage des consignes, nous avons fais le choix de coder les 8 profils en repérant en premier les répondants, puis le type de consigne parmi les 4 possibles

- Ho□Soi = sujets **H**omosexuels répondant à la consigne d'auto-description□Soi□
- Hé□Soi = sujets **H**étérosexuels répondant à la consigne d'autodescription□Soi□
- Ho□H = sujets **H**omosexuels répondant comme le feraient les **H**ommes□
- Hé□H = sujets **H**étérosexuels répondant comme le feraient les **H**ommes□
- Ho□F = sujets **H**omosexuels répondant comme le feraient les **F**emmes□
- Hé□F = sujets **H**étérosexuels répondant comme le feraient les **F**emmes□
- Ho□Ho = sujets **H**omosexuels répondant comme le feraient les **H**omosexuels□
- Hé□Ho = sujets **H**étérosexuels répondant comme le feraient les **H**omosexuels.

Nous partons ici d'un tableau de données (cf. Tableau 10 ou 11) comprenant, pour chacune des 8 conditions expérimentales, les moyennes de chaque trait calculées sur les 70 réponses obtenues dans chacune des conditions expérimentales. La Figure 7 présente le premier plan factorielle de l'ACP réalisée avant rotation. On remarque ainsi qu'avant projection et rotation bissectrice, l'arc de cercle des points-colonnes se dispose, avec une bonne approximation, sur le cercle des corrélations.

Figure 7. Premier plan factoriel de l'ACP sans rotation du Tableau 10

La Figure 8 présente le premier plan factoriel de l'ACP réalisée sur le même tableau de données après projection des variables sur le cercle et rotation bissectrice.

Figure 8. Premier plan factoriel de l'ACP avec rotation bissectrice du Tableau 10

On retrouve, dans le Tableau 10, le rang (solidarisation des variables) de chacun des 15 traits dans les 8 profils, le premier score factoriel avant ($\square_1$) et après rotation bissectrice ($\square_{B1}$), ainsi que le rang moyen de chacun des traits. Les traits ont été classés de façon décroissante en fonction de leur rang moyen.

On constate ici que les lignes qui contribuent le plus au premier facteur après rotation sont également celles dont la moyenne est la plus élevée. On a donc avec cette hiérarchisation sur les moyennes la mise au jour de la composante de taille. Dans cet exemple toutefois, la composante de taille n'est pas des plus remarquables, les lignes extrêmes étant peu homogènes (du rang 5 au rang 15 pour la première ligne et du rang 2 au rang 9 pour la dernière ligne).

Répondants Cibles	Hé : H	Ho : H	Hé : Soi	Ho : Soi	Ho : Ho	Ho : F	Hé : Ho	Hé : F	$\square_{F_1}$	$\square_{B_1}$	moy
est meneur	5	6	12	13	15	13	14	13	1.134	1.395	11.375
aime la compétition	3	3	13	14	11	14	13	14	1.254	1.019	10.625
FEMININ	15	15	15	15	13	2	4	1	-1.002	1.736	10.000
a confiance en soi	4	8	6	11	14	12	12	12	.956	.674	9.875
dévoué	11	12	10	7	10	11	8	7	-.197	.643	9.500
MASCULIN	1	1	1	12	12	15	15	15	1.759	-.083	9.000
bienveillant	10	10	9	9	7	9	7	6	-.257	.241	8.375
attentif aux besoins des autres	12	13	11	4	9	8	5	5	-.740	.327	8.375
énergique	8	4	5	8	6	10	11	11	.473	-.474	7.875
ambitieux	6	7	3	10	8	7	10	10	.401	-.355	7.625
sensible	14	14	14	2	1	1	1	2	-1.768	-.472	6.125
agréable	9	9	7	5	3	6	6	3	-.668	-.736	6.000
affectueux	13	11	8	1	4	5	2	4	-1.177	-.799	6.000
a du caractère	2	5	4	6	5	4	9	8	.161	-1.138	5.375
défend ses opinions	7	2	2	3	2	3	3	9	-.329	-1.978	3.875

Tableau 10. Synthèse des 8 profils et des principaux résultats de l'ACP utiles au repérage de la composante de taille

La Figure 8 aide à visualiser cette présence un peu moins marquée de la composante de taille dans cet exemple. On note, en effet, que le premier plan factoriel laisse apparaître un «vide» au milieu de l'arc de cercle. Ce vide traduit selon nous l'absence d'une structure fondamentale (dans une approche psychologique et non mathématique cette fois) que nous qualifions de *Thème* et que nous identifions avec la composante de taille [Milland, 2001] Flament, Milland, 2003] en opposition aux *variations* déclinées de cette structure et que la composante diagonale permet de décrire.

Ce qui prévaut ici, c'est la composante diagonale. Le Tableau 11 est une version du Tableau 10 présentant cette fois, le second score factoriel avant ($\square_2$) et le second score factoriel après rotation bissectrice ($\square_{B_2}$) des 15 traits, ainsi que les pentes calculées en régressant chaque ligne sur le deuxième vecteur des saturations après rotation ($\square_{B_2}$). Les traits ont été classés de façon décroissante en fonction de leur pente, mais auraient tout aussi bien pu l'être en fonction de $\square_{B_2}$, la corrélation entre les deux étant de .998 («est meneur» et «a confiance en soi» étant les seuls qui soient inversement hiérarchisés selon ces deux critères de classement).

Les items qui contribuent le plus au second facteur après rotation sont les lignes qui croissent et celles qui décroissent selon l'enchaînement. En d'autres termes, il s'agit des lignes ayant une pente extrême, en positif ou en négatif. La Figure 9 donne une assez bonne visualisation de cette composante diagonale sur seulement 3 items.

<i>Répondants</i> <i>Cibles</i>	Hé : H	Ho : H	Hé : Soi	Ho : Soi	Ho : Ho	Ho : F	Hé : Ho	Hé : F	$\square_{F2}$	$\square_{B2}$	pen
sensible	14	14	14	2	1	1	1	2	-.023	1.704	7.335
FEMININ	15	15	15	15	13	2	4	1	2.058	1.492	5.800
affectueux	13	11	8	1	4	5	2	4	-.517	1.007	4.439
attentif aux besoins des autres	12	13	11	4	9	8	5	5	.533	.851	3.562
agréable	9	9	7	5	3	6	6	3	-.586	.497	2.175
dévoué	11	12	10	7	10	11	8	7	.717	.372	1.511
bienveillant	10	10	9	9	7	9	7	6	.316	.329	1.327
défend ses opinions	7	2	2	3	2	3	3	9	-1.959	-.179	-0.401
a du caractère	2	5	4	6	5	4	9	8	-1.219	-.466	-1.833
ambitieux	6	7	3	10	8	7	10	10	-.472	-.507	-2.061
énergique	8	4	5	8	6	10	11	11	-.615	-.614	-2.387
a confiance en soi	4	8	6	11	14	12	12	12	.446	-.812	-3.559
est meneur	5	6	12	13	15	13	14	13	1.144	-.806	-3.629
aime la compétition	3	3	13	14	11	14	13	14	.725	-1.029	-4.528
MASCULIN	1	1	1	12	12	15	15	15	-.547	-1.840	-7.751
<i>Enchaînement sur $\square_{B2}$</i>	.994	.935	.663	-.548	-.454	-.863	-.922	-.994			

Tableau 11. Synthèse des 8 profils et des principaux résultats de l'ACP utiles au repérage de la composante diagonale

Figure 9. Trajectoires « diagonales » des items « MASCULIN », « FEMININ » et « sensible » à travers l'enchaînement des 8 profils du Tableau 12

On voit ainsi que les sujets hétérosexuels ont tendance à estimer que les homosexuels se décrivent comme féminin plutôt que masculin ce qui ne correspond pas à la perception des sujets homosexuels qui ne se voient jamais comme « FÉMININ » même s'ils font appels à des items identifiés ici comme des caractéristiques féminines plutôt que masculines.

DEUXIÈME EXEMPLE □ LA REPRÉSENTATION SOCIALE DE L'INTELLIGENCE MASCULINE VERSUS FÉMININE

Notre premier exemple illustre la nécessité d'une rotation pour bien faire apparaître la composante diagonale de l'Effet Guttman, la composante de taille étant de peu d'intérêt. Notre deuxième exemple aurait pu se passer de la rotation (comparer les Figures 10 et 11) □ mais il illustre de façon limpide le jeu des deux composantes. De plus, il utilise les résultats finaux d'une recherche publiée [Poeschl, 2001], montrant la complémentarité de notre analyse avec des techniques plus traditionnelles.

La recherche porte sur les représentations sociales de l'intelligence, au Masculin et au Féminin, selon des Hommes et des Femmes. Le questionnaire (construit à partir de recherches antérieures) comportait 34 traits pouvant caractériser l'Intelligence (en générale aussi bien que masculine ou féminine). Les réponses possibles allaient de -3 (pas important du tout) à +3 (très important), et étaient codées, dans la publication, de 1 à 7. Les personnes interrogées remplissaient deux fois ce questionnaire, selon les deux cibles (intelligence masculine / féminine), la moitié des sujets dans un ordre, l'autre moitié dans l'ordre inverse¹⁵. D'où un plan [2 (sexes des répondants) * 2 (ordres de passations) * 2 (cibles décrites)], avec mesures répétées pour le dernier facteur (cf. Tableau 12).

Répondants		Femmes		Hommes	
Echantillon		A	B	C	D
Cible	Intelligence féminine	en 1 ^{er}	en 2 ^e	en 1 ^{er}	en 2 ^e
	Intelligence masculine	en 2 ^e	en 1 ^{er}	en 2 ^e	en 1 ^{er}

Tableau 12. Plan expérimental [Poeschl, 2001]

La recherche fut conduite auprès de 56 sujets, chacun donnant deux profils de 34 réponses. Ces 112 profils individuels ont été soumis à une ACP (en plan R), et à l'application des techniques classiques des psychologues [Doise, Clemence, Lorenzi-Cioldi, 1992]. Il en ressort 4 facteurs permettant de définir 4 Echelles. Bien sûr, le questionnaire ayant été construit dans ce but, l'Auteur retrouve des domaines de l'intelligence identifiés antérieurement □

- Aptitudes cognitives (réflexion, capacité de compréhension, esprit critique, ...)
- Compétence sociale (être tolérant, honnête, assumer ses responsabilités, □.)
- Charme et flexibilité (être attractif, émancipé, malin, ...)
- Réussite sociale et professionnelle (avoir l'esprit scientifique, être un dirigeant, intellectuel, ...)

¹⁵ Ce plan expérimental est très naturel pour une recherche de ce type: une étude de la représentation du travail masculin / féminin a été analysée dans Flament et Milland [2003, p. 208-213].

L'hypothèse essentielle de cette recherche est que la description des deux cibles, l'une juste après l'autre, induira une comparaison entre elles – intelligence masculine et féminine se différenciant sur les traits spécifiquement sexués («**C**harme et flexibilité» pour les Dames ; «**R**éussite sociale et professionnelle» pour les Messieurs).

Le Tableau 13 (qui, sous une forme à peine différente, est donné par Poeschl [2001], annexe 2) donne le résumé des calculs – les moyennes des 4 échelles dans chacune des 8 cases du plan expérimental. L'hypothèse de l'Auteur est vérifiée, mais seulement en comparant (test de Student pour chaque échelle) les deux productions successives d'un même groupe de sujets. Les comparaisons entre sujets de sexe différent, notamment, seraient discutables, car le Tableau 13 montre que les Femmes donnent des réponses très supérieures à celles des Hommes – cela dans 15 des 16 comparaisons possibles, l'unique exception concernant la «**C**ompétence sociale» dans la description de l'intelligence masculine en 2^e position, où la moyenne des Femmes est de 5.19, alors que celle des Hommes est de 5.22. Nous ne connaissons pas d'explication à cette supériorité de la positivité des réponses des Femmes par rapport aux réponses des Hommes.

L'ACP du Tableau 13 (tableau annexe final chez Poeschl, tableau initial pour nous) fait apparaître un Effet Guttman très satisfaisant (Figure 10) avec $94 + 4 = 98\%$ de la variance expliquée par les deux premiers facteurs.

Cibles	Intelligence féminine				Intelligence masculine			
	Femmes		Hommes		Femmes		Hommes	
Répondants								
Ordres	en 1 ^{er}	en 2 ^e	en 1 ^{er}	en 2 ^e	en 1 ^{er}	en 2 ^e	en 1 ^{er}	en 2 ^e
Aptitudes cognitives	6.08	5.81	5.58	5.28	5.96	6.23	5.18	5.72
Compétence sociale	5.48	5.18	5.19	4.24	5.10	5.19	4.33	5.22
Charme et flexibilité	5.20	4.57	4.41	3.97	4.64	4.89	3.46	4.19
Succès social et professionnel	4.24	3.69	3.65	3.56	4.59	4.39	3.61	4.13
Moyennes	5.25	4.81	4.71	4.26	5.07	5.18	4.15	4.82

Tableau 13. Moyennes des 4 échelles à travers les 8 profils

Figure 10. Premier plan factoriel avant rotation des données du Tableau 13

Bien que la projection des points sur le cercle des corrélations, et la rotation bissectrice ne changent presque rien, ces calculs ont été faits pour construire la Figure 11, qui permet une explicitation claire de l'enchaînement des 8 conditions expérimentales.

On constate que l'intelligence féminine est située dans la moitié supérieure, s'opposant (faiblement, mais nettement) à l'intelligence masculine. De plus, lorsque la comparaison joue («*en 2^e*»), la différence tend à s'estomper (les 4 flèches sont centripètes¹⁶).

¹⁶ Si on construit un schéma analogue pour le travail masculin *versus* féminin [Flament, Milland, 2003], on constate que les 4 flèches sont centrifuges, révélant un jeu masculin / féminin inverse selon l'objet de la représentation sociale – différences accentuées pour le Travail, diminuées pour l'Intelligence – l'histoire (au cours du dernier siècle) de ces objets expliquant sans doute cette opposition.

Pour mieux décrire ce résultat, on peut visualiser les deux composantes de l'Effet Guttman (cf. Figure 12) – mais pour cela, il faut solidariser les 8 profils (ne serait-ce que pour neutraliser les écarts, inexplicables, des réponses Hommes / Femmes). Ici, notre modèle de données sur le permutoèdre ne saurait convenir sur 4 éléments (les 4 échelles), une chaîne suivant un demi grand cercle ne peut comporter que 6 permutations, et cela seulement lorsqu'elle relie deux permutations inverses (diamétralement opposées sur le permutoèdre) – ce qui est loin d'être le cas dans le présent exemple. On s'est donc contenté d'une solidarisation par simple égalisation des moyennes de chaque profil moyen (qu'on a fixée à 2 pour des raisons de commodité graphique).

Figure 12. Trajectoire des 4 échelles à travers l'enchaînement des 8 profils du Tableau 13.

La Figure 12, d'une grande lisibilité, montre :

- la composante de taille, constituée par les échelles non sexuées (« Aptitudes cognitives » et « Compétence sociale »), dont les légères fluctuations sont sans lien avec l'enchaînement des 8 conditions expérimentales

- la composante diagonale, qui permet une généralisation de l'hypothèse de Poeschl : c'est de façon assez régulière, en suivant l'enchaînement, que « l'Charme et flexibilité » décroît du côté féminin vers le côté masculin – « Réussite sociale et professionnelle » ayant l'évolution inverse.

4. ANALYSE DE TENDANCES LINEAIRES OU LA RECHERCHE DE PETITS « η »

Certains Psychologues seront peut-être rassurés par des calculs plus traditionnels, susceptibles de mesurer des seuils de significativité en rapport avec les sujets. Une fois l'enchaînement des conditions repéré, et traduit numériquement, d'une façon ou d'une autre – il suffit d'attribuer à chaque profil individuel la valeur repérant la condition où il se trouve (nouvelle variable qu'on peut nommer « l'enchaînement »), et de calculer la régression linéaire entre enchaînement et chacune des variables du tableau initial de la recherche – cela, par une « étude de tendance linéaire » dans le cadre d'une analyse de variance [Abdi, 1987, p.282-287 ; Howell, 1998, p. 436-446], ou, directement, selon le logiciel disponible. Pour peu que les données initiales aient été transformées pour tenir compte de la solidarisation de nos profils moyens – on retrouve nécessairement les mêmes pentes à un coefficient près : les pourcentages de variance expliquée diminuent considérablement, puisque, à la variance inter conditions, on ajoute la variance intra ; et on obtient des seuils p de significativité ! Ainsi :

- dans l'exemple « Homosexuel / hétérosexuel », le trait « Défend ses opinions » n'a une pente significative qu'à $p = .003$, les pentes des 14 autres traits l'étant à $p < .0001$,
- dans l'exemple « Intelligence », seule l'échelle « Succès social et professionnel » a une pente significative à $p < .05$; la pente de « l'Charme et flexibilité » n'a une tendance qu'à $p < .13$.

De tels résultats peuvent plaire à certains – mais ils ne changent rien aux conclusions de notre analyse purement descriptive et nullement inductive.

5. EN GUISE DE CONCLUSION

Certes, s'accordera-t-on à considérer que l'on aurait pu se passer de l'ACP pour traiter de l'effet Guttman. Un des éléments clé de notre analyse repose sur la mesure de la congruence de chaque ligne du tableau avec un enchaînement des vecteurs-colonnes et consiste par là-même à repérer initialement l'enchaînement de ces vecteurs, ce qu'une analyse combinatoire aurait également pu permettre. Le problème revient notamment à trouver une chaîne résumant un ensemble de permutations. Or, la recherche d'une structure de Robinson dans une matrice de distances (pour nous : distance de Kendall, ou de Spearman ; ou, dualement, matrice de corrélations) concourt à révéler cet enchaînement des vecteurs-colonnes de notre tableau de données. On pourrait donc regretter d'en arriver à l'usage de l'ACP dans la poursuite de cet objectif. En effet, si un algorithme analogue à celui de Guénoche [Guénoche, Monjardet, 1987] était disponible, cette approche combinatoire du problème aurait constitué un bon point de départ, mais tel n'est pas le cas.

Finalement, contrairement à ce que prédisait Kolmogorov durant l'International Congress of Mathematicians à Nice en 1970, (cité par [Guénoche, Monjardet, 1987], en exergue de leur article) :

... with the development of the modern computing technique, it will be clear that in very many cases it is reasonable to conduct the study of real phenomena avoiding the intermediary stage of stylizing them in the spirit of ideas of mathematics of the infinite and continuous, and passing directly to discrete models

les logiciels actuels sont toujours dans le continu, nous conduisant à injecter nos chaînes de permutations dans le voisinage continu d'un arc de cercle continu□

Remerciements. Nous remercions des amis plus matheux que nous de l'aide qu'ils nous ont apportée, d'une façon ou d'une autre : M. Barbut, Ph. Cibois, A. Degenne, J.-P. Doignon, A. El Ahmadi, A. Guénoche, M-O Lebeaux, H. Rouanet. Nous sommes reconnaissants à tous ceux qui nous ont aidé à l'amélioration de ce texte, et espérons que d'autres travaux prolongeront notre travail.

BIBLIOGRAPHIE

- ABDI H., *Introduction au traitement statistique des données expérimentales*, Grenoble, Presses Universitaires de Grenoble, 1987.
- ARABIE P., HUBERT L.J., "Combinatorial data analysis", *Annual review of psychology* 43, 1992, p. 169-203.
- BENZÉCRI J.P., «Sur l'analyse des préférences□, *Ordres totaux finis*, M. Barbut (éd.), Paris, Mouton & Gauthier-Villars, 1971, p. 185-205.
- BLACK, D., "On the rational of group decision making", *Journal of Political Economy*, fascicule 56, 1948.
- BRAZILL, T.J., GROFMAN, B., "Factor analysis versus multi-dimensional scaling: binary choice roll-call voting and the US Supreme Court", *Social Networks* 24, 2002, p. 201-229.
- COOMBS, C.H., *A theory of data*, Oxford (England), Wiley, 1964.
- DEGENNE A., *Techniques ordinales en analyse des données statistiques*, Paris, Hachette, 1972.
- DOISE W., CLEMENCE A., LORENZI-CIOLDI F, *Représentations sociales et analyses de données*, Grenoble, Presses Universitaires de Grenoble, 1992.
- FALISSARD B., «Déploiement d'une matrice de corrélation sur la sphère unité de R^3 □, *Revue de Statistiques Appliquées* 43(2), 1995, p.□5-48.
- FLAMENT, C., «Modèles à caractéristiques non monotones dans l'étude d'un questionnaire□, *Revue Française de Sociologie* 4(2), 1963, p. 172-194.
- FLAMENT C., MILLAND L., «Un effet Guttman dans l'analyse des représentations sociales□, *Méthodes d'étude des représentations sociales*, J.C. Abric (éd.), Saint-Agne, Erès, 2003, p.□01-220.
- GUÉNOCHE A., MONJARDET B., «Méthodes ordinales et combinatoires en analyse des données□, *Mathématique et Sciences Humaines* 100, 1987, p. 5-47.
- HOWELL D.C., *Méthodes statistiques en sciences humaines*, Bruxelles, De Boeck Université, 1998.
- KRUSKAL J.B., WISH M., *Multidimensional scaling*, Londres, Sage Publications, 1991 (17^e édition).

- LEBART L., MORINEAU A., PIRON M., *Statistique exploratoire multidimensionnelle*, Paris, Dunod, 2000 (3^e édition).
- MARTINIE M.A., *La réduction de la dissonance cognitive en situation de fausse attribution □ changement d'attitude, trivialisation, rationalisation en acte*, Thèse de Doctorat, Université de Provence Aix-Marseille I, 2000.
- MATALON B., *L'analyse hiérarchique*, Paris, Mouton & Gauthier-Villars, 1965.
- McIVER J.P., CARMINES E.G., *Unidimensional Scaling*, Beverly Hills, Sage Publications, 1981.
- MICCERI T., "The Unicorn, The Normal Curve, and Other Improbable Creatures", *Psychological Bulletin* 105(1), 1985, p. 156-166.
- MILLAND L., *De la dynamique des rapports entre représentations sociales du travail et du chômage*, Thèse de Doctorat, Université de Provence Aix-Marseille I, 2001.
- POESCHL G., "Social comparison and differentiation strategies in social representations of intelligence", *Swiss Journal of Psychology* 60(1), 2001, p. 15-26.
- RALLIER D., RICOU G., *Description de soi et références normatives : étude d'une population masculine homosexuelle et hétérosexuelle*, Mémoire de maîtrise de psychologie sociale, Université de Provence Aix-Marseille I, 2000.
- ROBINSON W.S., "A method for chronologically ordering archaeological deposits", *American antiquity* 16, 1951, p. 293-301.
- ROUANET H., LE ROUX B., *Analyse des données multidimensionnelles*, Paris, Dunod, 1993.