Arbres, classes, distances…

Numéro édité sous la direction de Jacques Poitevineau

Cahiers du LCPE
N° 6 - décembre 2002

Méthode des arbres de similarité additifs de Sattath et Tversky : Illustration dans une tâche de catégorisation de situations d'incertitude[1]

Jacques Poitevineau (a,b), Bruno Lecoutre (a,c), Marie Paule Lecoutre(a,d), Katia Rovira (d)

[a]ERIS
[b]LCPE, C.N.R.S
[c]Laboratoire de Mathématiques Raphaël Salem[2]
[d]Laboratoire Psy.co, E.A. 1780, Université de Rouen[3]

Introduction

Dans une tâche de catégorisation d'objets, les données à analyser se présentent sous forme d'un ensemble de partitions des objets. A partir de cet ensemble on peut facilement construire une matrice de dissimilarités entre les objets. La méthode des arbres de similarité additifs de Sattath et Tversky (1977) permet de représenter la structure des objets sous la forme d'un arbre (un ensemble de nœuds reliés par des arêtes). Les objets correspondent alors aux nœuds externes (ou feuilles) de l'arbre, et la dissimilarité entre deux objets est représentée par la longueur du chemin qui les joint. Le programme informatique Addtree (dans la version de Barthélemy et Guénoche, 1991) permet de construire ces arbres.
Nous illustrerons ici l'utilisation de cette méthode pour analyser les données obtenues dans une tâche de catégorisation de situations d'incertitude par des sujets appartenant à trois groupes se différenciant selon leur degré d'expertise en

1 Marie-Paule Lecoutre et Katia Rovira sont les auteurs de l'expérience présentée ici pour illustrer la méthode ; Jacques Poitevineau et Bruno Lecoutre sont plus particulièrement responsables des analyses statistiques.
2 UMR 6085, C.N.R.S. et Université de Rouen Mathématiques, Site Colbert, 76821 Mont-Saint-Aignan Cedex,
bruno.lecoutre@univ-rouen.fr, http://www.univrouen.fr/LMRS
3 UFR Psychologie, Sociologie, Sciences de l'Éducation 76821 Mont-Saint-Aignan Cedex,
marie-paule.lecoutre@univ-rouen.fr, http://www.univ-rouen.fr/psy-socio-sceduc/People
katia.rovira@univ-rouen.fr

probabilités. Les représentations par des arbres sont utilisées ici pour faire ressortir la structure des objets et la comparer à une classification théorique *a priori*. La méthode est également appliquée pour rechercher une classification des sujets à l'intérieur de chacun des groupes.

I. Une expérience de catégorisation de situations d'incertitude

I.1. Problématique de l'expérience

Le hasard est une notion particulièrement complexe et ambiguë qui a donné et continue de donner lieu à de nombreuses interprétations. Au $19^{ème}$ siècle, en accord avec la conception Laplacienne déterministe du monde, le hasard est le nom donné à l'ignorance d'une personne dans un univers entièrement déterminé. Au contraire, quelques théories plus récentes reposent sur l'idée que le hasard est dans l'essence même des choses. Par exemple, en physique quantique le comportement des particules est conçu comme fondamentalement probabiliste. Pour formaliser la notion de hasard le concept de probabilité a été utilisé dès le 17e siècle par les mathématiciens. De nos jours la communauté scientifique s'accorde à reconnaître deux interprétations principales. (1) La plus classique est une interprétation dite *fréquentiste* ou *objectiviste* selon laquelle la probabilité renvoie à des fréquences objectives de réalisation d'événements. (2) Il existe une interprétation plus large dite *subjectiviste* selon laquelle la probabilité traduit un degré de confiance dans la vérité d'une proposition ; c'est à cette dernière interprétation que se réfère la théorie statistique Bayésienne. L'étude expérimentale présentée ici a pour objectif d'apporter des éléments de réponse aux questions suivantes. Que mettent spontanément les gens derrière le mot « hasard » ? Existe-t-il des interprétations dominantes ? Dans quelle mesure sont-elles liées au degré d'expertise en probabilité ? Sont-elles proches de certaines conceptualisations probabilistes ?

I.2. Méthode

Sujets

Pour étudier l'effet du degré d'expertise en probabilités nous avons observé trois groupes de sujets: (1) 20 enseignants-chercheurs en mathématiques considérés comme experts ; (2) 20 enseignants-chercheurs en psychologie confrontés à la notion de hasard dans leur pratique statistique ; (3) 20 collégiens en classe de 3ème n 'ayant pas reçu d'enseignement en probabilités.

Matériel

La situation expérimentale consiste en une tâche de catégorisation d'évènements incertains. Nous considérons que la façon dont un sujet classe un matériel permet d'avoir accès à ses représentations (*cf.* par exemple Medin, Lynch et Coley, 1997). Le matériel comprend 16 situations d'incertitude (présentées dans le tableau I) écrites sur des cartons blancs que le sujet peut manipuler. Le choix des situations s'appuie notamment sur les travaux de Nisbett *et al.* (1983) et Konold *et al.* (1991). Il y a 8 situations « réelles » et 8 situations « stochastiques ». Les situations réelles sont inspirées d'événements de la vie courante (par exemple le fait d'attraper un rhume dans le courant du mois prochain). Dans 4 situations il y a implication du sujet, ce dernier étant désigné par « vous » dans l'énoncé ; dans les 4 autres il n'y a pas d'implication, le sujet n'étant pas désigné dans l'énoncé. Les situations stochastiques renvoient à des processus répétables (par exemple le lancer d'un dé) ou des situations de tirages au sort. Quatre situations ont des issues symétriques c'est-à-dire équiprobables ; les 4 autres ont des issues non symétriques.

situations « réelles »	
Avec implication explicite du sujet	
A « Ami »	Le fait de rencontrer un ami que vous n'avez pas vu depuis 10 ans
B « Loto »	Le fait que vous gagniez 10000F au loto
C « Tête »	Le fait de dire la première chose qui vous passe par la tête
D « Rhume »	Le fait que vous attrapiez un rhume dans le courant du mois prochain
Sans implication explicite du sujet	
E « Graine »	Le fait qu'une graine mise en terre germe
F « Bourse »	Le fait que le cours d'une action à la bourse de Paris aura progressé de plus de 5% dans trois mois
G « Pluie 1 »	Le fait qu'il a plu à Paris le 15 mars 1936
H « Pluie 2 »	Le fait qu'il pleuve demain à Paris
situations « stochastiques »	
Avec issues symétriques	
S « Dé »	Le fait d'obtenir un chiffre pair à l'issue du lancer d'un dé
T « Pièce »	Le fait d'obtenir Pile à l'issue du lancer d'une pièce de monnaie non truquée
U « Faces »	Le fait d'obtenir Face au 5ème lancer d'un pièce de monnaie non truquée qui est tombée sur Face les quatre fois précédentes
V « Boules 1 »	Le fait de tirer une boule blanche d'une boîte qui contient 10 boules noires et 10 boules blanches
Avec issues non symétriques	
W « Jetons »	Le fait de tirer simultanément 2 jetons rouges d'une boîte qui contient 1 jeton blanc et 2 jetons rouges
X « Chaussettes »	Le fait de constituer une paire de chaussettes « assorties » à partir d'un tirage à l'aveugle de deux chaussettes d'un tiroir qui contient 2 paires de chaussettes différentes
Y « Bonbons »	Le fait de tirer un bonbon au citron d'une boîte qui contient 20 bonbons à l'orange et 10 au citron
Z « Boules 2 »	Le fait de tirer une boule blanche d'une boîte qui contient 10 boules noires et 20 boules blanches

Tableau I - Les 16 situations d'incertitude considérées

Procédure

L'expérience se déroule en deux phases. (1) On demande au sujet de mettre ensemble les situations qui se ressemblent, il peut faire autant de classes qu'il le souhaite ; il s'agit d'une tâche de « catégorisation libre » dans laquelle la notion de hasard n'est pas explicitée. (2) On reprend toutes les situations et le sujet doit dire, pour chacune d'elle, si le hasard intervient ou non ; il s'agit d'une tâche de « catégorisation contrainte » dans laquelle la notion de hasard est alors explicite. Dans les deux cas il est demandé au sujet de justifier ses classements et ses réponses.

II. Résultats

Les résultats ont été analysés avec plusieurs méthodes complémentaires et les interprétations données dans cet article s'appuient sur l'ensemble de ces méthodes. Nous ne présenterons ici que la méthode des arbres de similarité additifs, qui, comme nous l'avons dit dans l'introduction, permet à partir des catégorisations observées de construire : (1) des arbres associés aux situations pour étudier leur structure ; (2) des arbres associés aux sujets pour rechercher une classification de ces derniers.

II.1. Ajustement par un modèle d'arbre

Pour juger de la qualité d'une représentation par un arbre, on dispose d'un ensemble d'indices (*cf.* Guénoche et Garreta, 2001 et l'avant propos de ce numéro pour des définitions), qui se regroupent en deux catégories. (1) Les indices métriques comparent les distances déduites des partitions effectuées par les sujets avec les distances dans la représentation arborée. (2) Les indices topologiques ne font intervenir que la structure de l'arbre.

Il est également intéressant de considérer le rapport des longueurs des arêtes internes à celles des arêtes externes (un arbre est plus facile à reconstituer quand ce rapport est grand). Ce rapport ne fournit pas un critère d'ajustement, mais il joue un rôle important dans les études de simulation qui servent à avoir une idée de la variation des indices quand les distances sont « bruitées » aléatoirement.

Le tableau II donne, pour les arbres associés aux situations, la valeur de ce rapport et de deux indices de qualité de représentation : un indice métrique, le stress de Kruskal, et un indice topologique, le taux de quadruplets bien représentés. La représentation est d'autant meilleure que le stress est plus petit et que le taux de quadruplets est élevé.

	Rapport arêtes internes/externes	*Stress*	*Taux de quadruplets bien représentés*
	Catégorisation libre		
Mathématiciens	0,45	0,06	87%
Psychologues	0,59	0,08	87%
Collégiens	0,43	0,13	85%
	Catégorisation contrainte		
Mathématiciens	0,80	0,08	90%
Psychologues	0,56	0,12	82%
Collégiens	0,73	0,12	71%

Tableau II – Indices de la qualité de représentation des arbres associés aux situations

En se référant aux études de simulation effectuées par Guénoche et Garreta (2001), les valeurs obtenues permettent de considérer qu'un modèle d'arbre apparaît raisonnablement ajusté aux données dans tous les cas. La meilleure représentation est obtenue pour les mathématiciens qui ont le plus petit stress et le plus grand taux de quadruplets bien représentés. En ce qui concerne les arbres associés aux sujets, les indices (non présentés ici) sont moins bons mais restent encore acceptables. Nous commentons ci-après les arbres obtenus dans chacune des deux phases.

II.2. Première phase : Catégorisation libre (hasard non explicite)

Arbres associés aux situations

Les trois structures d'arbres pour les trois groupes sont voisines et reflètent assez bien la classification théorique à savoir : la dissociation entre situations réelles et stochastiques, et à l'intérieur de celles-ci la différence entre implication/sans implication du sujet et symétrie/non-symétrie des issues. Pour les mathématiciens, comme le montre la figure 1, on retrouve même exactement cette classification.

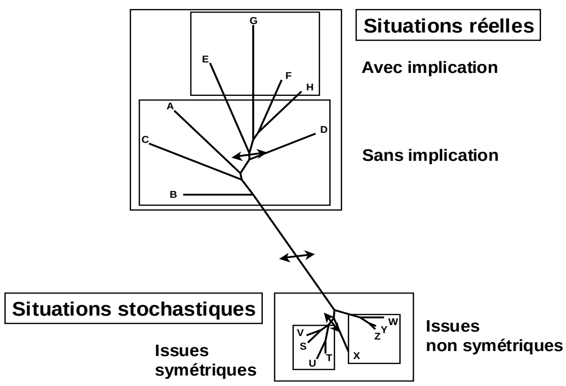

Figure 1 – Mathématiciens : arbre associé aux situations dans la catégorisation libre

Pour les collégiens (voir la figure 2), nous retrouvons les deux classes principales mais la situation B (Loto) est classée avec les situations stochastiques. Pour les psychologues, la structure de l'arbre (non reproduit ici) est intermédiaire.

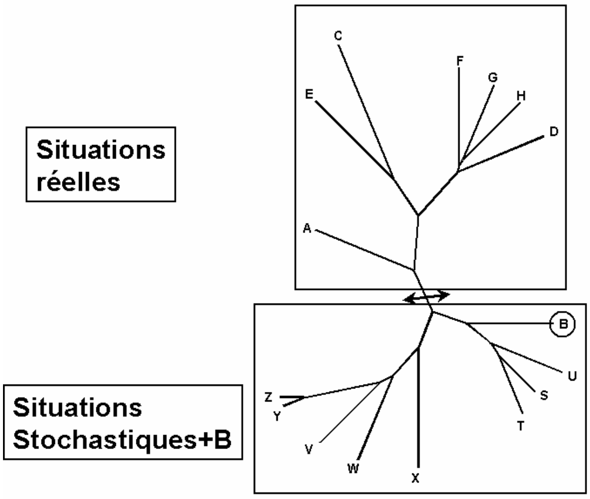

Figure 2 – Collégiens : arbre associé aux situations dans la catégorisation libre

Comparaison des arbres associés aux situations
Pour comparer les arbres, on a calculé la distance de Robinson et Foulds (1981) entre ces arbres (*cf.* le tableau III). Cette distance est purement topologique, c'est-à-dire qu'elle ne tient compte que de la structure des arbres et pas de la longueur des arêtes. Elle est égale au nombre minimum d'opérations élémentaires (fusion ou division de nœuds) nécessaires pour transformer un arbre en un autre et est comprise entre 0 et 26 (c'est-à-dire $2k$-6, k=16 étant le nombre d'objets). Par ailleurs, on a également calculé les distances avec l'arbre théorique correspondant à la structure *a priori* des situations (figure 3).

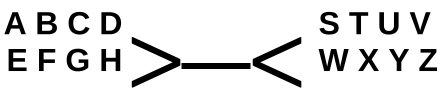

Figure 3 – Arbre théorique associé aux situations

	Théorique	Mathématiciens	Psychologues
Mathématiciens	12		
Psychologues	16	22	
Collégiens	18	20	12

Tableau III - Distance de Robinson et Foulds entre les arbres des situations

On vérifie que l'arbre des mathématiciens est le plus proche de l'arbre théorique, et que les arbres des collégiens et des psychologues, relativement proches entre eux, en sont à peu près équidistants. On pourrait s'étonner du fait que la distance entre l'arbre théorique et l'arbre des mathématiciens n'est pas nulle, alors qu'on a indiqué que les classifications des situations étaient identiques dans les deux cas. Ceci provient de la « structure fine » de l'arbre des mathématiciens : la distance de 12 est entièrement due aux petites arêtes qui distinguent les situations appartenant à une même catégorie *a priori*.

Arbres associés aux sujets

Les arbres associés aux sujets permettent d'affiner les résultats globaux précédents. Chez les mathématiciens (figure 4), seuls 3 sujets réalisent la même classification. On peut cependant identifier 3 sous-groupes de sujets: (1) 7 sujets qui regroupent la majorité des situations dans une seule classe ; (2) 7 sujets qui séparent les situations réelles des situations stochastiques ; (3) 6 sujets qui répartissent les situations dans un nombre élevé de classes (entre 4 et 12).

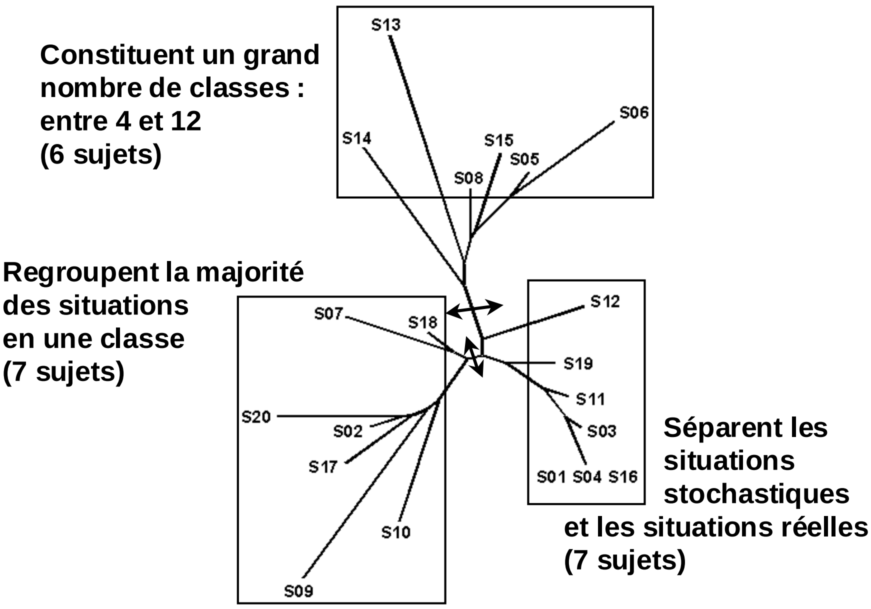

Figure 4 – Mathématiciens : arbre associé aux sujets dans la catégorisation libre

Chez les Psychologues (figure 5) on observe une structure d'arbre nettement différente. (1) 6 sujets regroupent en une classe les 8 situations stochastiques ; (2) 3 sujets regroupent dans une classe les 8 situations stochastiques avec en plus des situations réelles ; (3) 8 sujets constituent un grand nombre de classes (entre 4 et 7). Seuls deux sujets réalisent la même classification, à savoir une classe unique ; un sujet réalise deux classes.

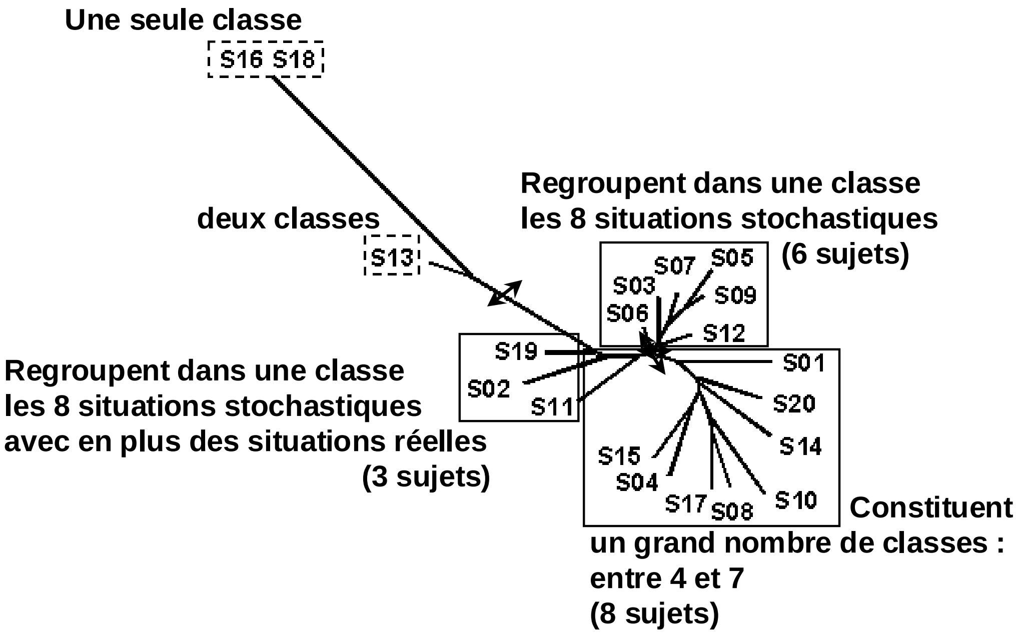

Figure 5 – Psychologues : arbre associé aux sujets dans la catégorisation libre

Chez les collégiens (figure 6) nous notons encore une structure d'arbre différente et autant de classifications qu'il y a de sujets. Il y a également 3 sous-groupes de sujets mais avec des interprétations différentes : (1) 8 sujets regroupent dans une classe au moins 6 situations stochastiques ; (2) 7 sujets regroupent dans une classe 4 situations réelles ; (3) 5 sujets mélangent les deux types de situations.

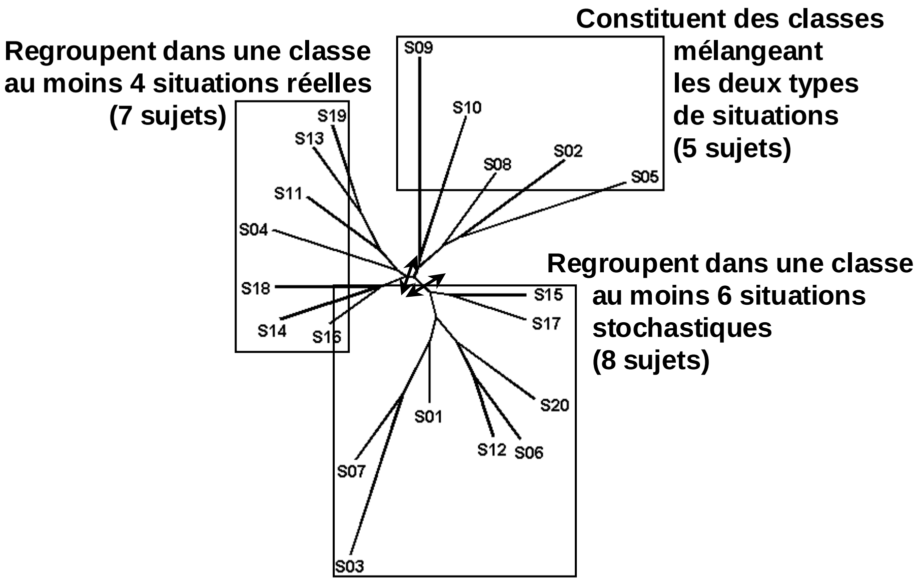

Figure 6 – Collégiens : arbre associé aux sujets dans la catégorisation libre

Si l'on considère les trois arbres, on peut noter une très grande variabilité interindividuelle dans chacun des trois groupes et une variabilité d'un groupe à l'autre, puisque la structure des arbres diffère selon les groupes. Ces résultats montrent que les sujets se différencient dans leur façon de catégoriser les situations et donc selon les critères qu'ils utilisent.

II.3. Deuxième phase : Catégorisation contrainte (hasard explicite)

Dans la situation de catégorisation contrainte, le sujet doit dire si pour chacune des 16 situations le hasard intervient ou non. Les réponses ont conduit à considérer deux classes supplémentaires, l'une *intermédiaire* qui regroupe des réponses nuancées du type « le hasard intervient un peu moins », et l'autre qui correspond à quelques cas où le sujet déclarait qu'il lui était impossible de se prononcer. Ainsi le nombre de classes des catégorisations analysées varie de 1 à 4. Cette situation pose en fait un problème particulier, puisqu'on sait ici qu'une même catégorisation observée correspond dans certains cas à deux réponses opposées. Ainsi par exemple, quelques sujets répondent que le hasard intervient dans toutes les situations, alors que d'autres répondent qu'il n'intervient dans aucune. Dans ces deux cas cela revient à regrouper toutes les situations en une seule classe. Plus généralement deux sujets qui ont donné des réponses diamétralement opposées (pour chaque situation, si un sujet l'a qualifiée de « hasard », l'autre l'a qualifiée de « pas de hasard ») sont considérés comme « équivalents ». Si on applique la méthode directement aux partitions correspondant aux réponses des sujets, on ne tient pas compte de la nature de la réponse.

Nous avons néanmoins directement appliqué la méthode dans un premier temps. Nous avons ensuite envisagé la modification suivante. Quatre objets *fictifs*, « hasard », « pas de hasard », « intermédiaire » et « ? » ont été ajoutés. Dans chacune des classes d'objets constituées par un sujet, on ajoute l'objet fictif correspondant à la qualification de la classe. En outre, si le sujet n'a pas lui-même créé toutes les 4 classes, on ajoute les classes manquantes qui ne contiennent alors que l'objet fictif correspondant. Ainsi, deux sujets réalisant la même partition mais avec des qualifications différentes sont distingués. Il est important de noter que cet ajout d'objets fictifs ne change pas les distances entre les « vrais » objets (non fictifs). Par construction, les 4 objets fictifs seront toujours disjoints et la distance entre eux sera maximale. En outre, ces objets facilitent l'interprétation des classes dans les arbres associés aux situations.

Arbres associés aux situations

Application directe de la méthode

Les trois structures d'arbres associés aux situations sont encore voisines pour les trois groupes de sujets, et on retrouve bien la distinction entre situations stochastiques et réelles. Cependant les deux situations réelles A (Ami) et B (Loto) apparaissent intermédiaires pour les mathématiciens (comme on peut le voir sur la figure 7), et même nettement plus proches des situations stochastiques que des réelles pour les deux autres groupes (ce qui pour les collégiens est conforme au résultat obtenu dans la catégorisation libre). Par ailleurs, on ne retrouve plus, même chez les mathématiciens, la distinction entre les sous-catégories définies a priori par l'implication du sujet et par le type d'issues.

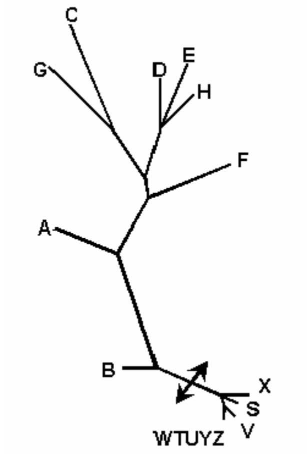

Figure 7 – Mathématiciens : arbre associé aux situations dans la catégorisation contrainte (application directe de la méthode)

Ajout d'objets fictifs

La classification des situations est peu modifiée (voir la figure 8 pour les mathématiciens) de même que les indices de qualité de représentation. Dans les 3 groupes les situations stochastiques sont proches du hasard et les situations réelles plus proches de l'absence de hasard, à l'exception de A et B. Chez les mathématiciens, c'est la situation C (dire ce qui vous passe par la tête) qui est la plus proche de l'absence de hasard ; chez les psychologues et les collégiens, c'est d'abord la situation E (Graine) puis la situation C.

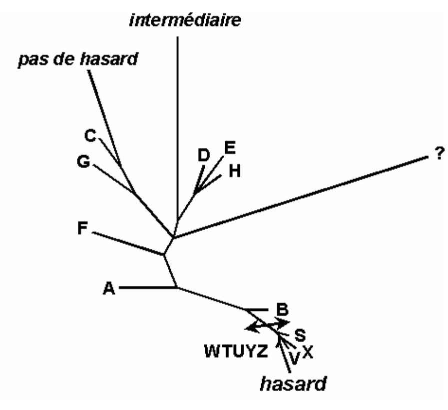

Figure 8 – Mathématiciens : arbre associé aux situations dans la catégorisation contrainte (avec ajout d'objets fictifs)

Arbres associés aux sujets

Les arbres associés aux sujets reflètent une variabilité inter-individuelle (et inter-groupe) encore plus importante pour la catégorisation avec contrainte que pour la catégorisation libre. On peut le voir pour les mathématiciens sur la figure 9 pour l'application de la méthode avec ajout d'objets fictifs.

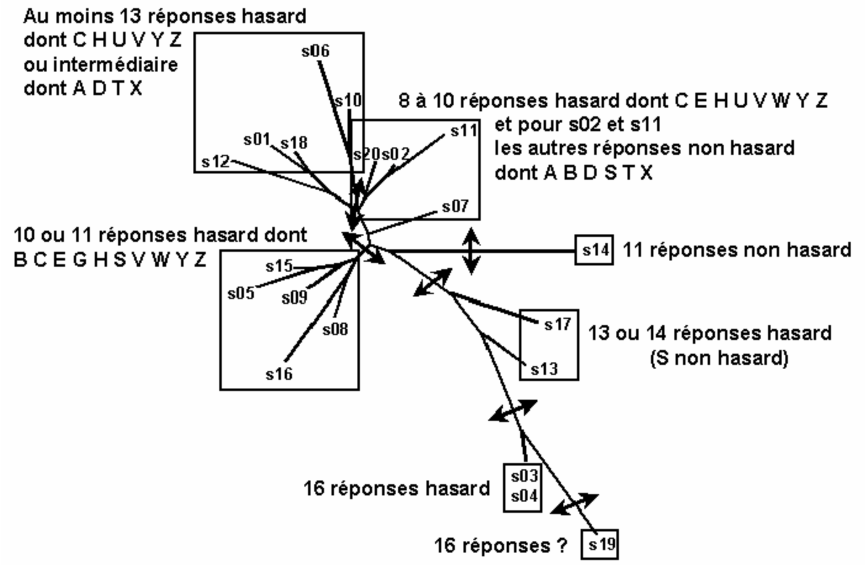

Figure 9 – Mathématiciens : arbre associé aux sujets dans la catégorisation contrainte (avec ajout d'objets fictifs)

II.4 Comparaison des arbres des deux catégorisations

La comparaison des structures d'arbres pour les situations montre un point commun, à savoir la séparation entre les situations réelles et les situations stochastiques ; néanmoins cette distinction est sensiblement atténuée dans la phase 2. Cependant la comparaison révèle également plusieurs différences dont les principales sont les suivantes. Pour les situations stochastiques le type d'issue « symétrique » ou « non symétrique » intervient beaucoup moins. Pour les situations réelles, les mathématiciens ne font plus la distinction entre « avec » ou « sans » implication ; pour ces situations encore, les situations A (Ami) et B (Loto) sont plus proches des situations stochastiques.

Ces différences dans les structures d'arbre montrent des changements notables d'une phase à l'autre. On peut supposer que le fait de rendre explicite la notion de hasard conduit le sujet à activer des connaissances différentes ce qui entraîne un changement de point de vue sur les situations considérées. Ceci renvoie à l'aspect dynamique de la catégorisation et à la flexibilité des représentations du sujet.

III. Interprétations : quelques représentations du hasard

Dans la situation de catégorisation avec contrainte, les résultats ont été complétés et éclairés notamment par une analyse des patrons de réponses individuels, s'appuyant sur les justifications données par les sujets. Cette analyse, non présentée ici, permet de dégager cinq interprétations dominantes: deux interprétations globales rendant compte de l'ensemble des 16 réponses et trois interprétations locales rendant compte d'une partie des réponses

La première interprétation globale dominante est « il y a hasard quand il y a un raisonnement probabiliste » ; ceci revient à considérer que le hasard intervient toujours. Cette interprétation est majoritaire, surtout chez les mathématiciens, ce qui n'est pas surprenant. Cependant, on observe (surtout chez les mathématiciens) une distinction entre deux types de hasard, selon les situations: un hasard mathématique lié aux événements pour lesquels on peut calculer une probabilité, comme c'est le cas pour les situations stochastiques (conception objectiviste classique) ; un hasard par ignorance lié aux événements pour lesquels la probabilité est calculable mais difficilement, comme c'est le cas pour les situations réelles. Donc il y a toujours du hasard mais les situations stochastiques et réelles sont opposées selon la nature du hasard.

La deuxième interprétation globale oppose également les situations stochastiques et les situations réelles mais selon la présence ou l'absence du hasard. Les sujets ont un point de vue déterministe sur les événements réels (donc « le hasard n'intervient pas ») ; par contre, on peut calculer une probabilité pour les items stochastiques (donc « le hasard intervient »). C'est chez les collégiens qu'on rencontre le plus cette interprétation ; il est intéressant de noter qu'avant tout enseignement des probabilités les collégiens repèrent les situations facilement probabilisables.

Les trois interprétations locales, dont chacune rend compte d'une partie des réponses sont les suivantes. (1) Un événement relève typiquement du hasard quand toutes les issues ont la même probabilité, ce qui se traduit par la réponse que le hasard intervient au moins pour les quatre situations stochastiques à issues symétriques ; ce type d'interprétation est à rapprocher du « biais d'équiprobabilité » (Lecoutre, 1992) selon lequel des événements aléatoires sont considérés comme équiprobables *par nature*. (2) Le degré d'intervention du hasard est lié à la valeur des probabilités associées aux issues, le hasard intervenant d'autant plus que ces valeurs sont faibles. (3) Enfin, un phénomène relève du hasard quand on ne peut pas faire de prédiction soit parce qu'on ne peut pas identifier de facteur causal, soit parce qu'on n'a aucun contrôle sur la situation.

On notera encore que les enseignants-chercheurs en psychologie se différencient par le fait que 3 sujets disent que le hasard n'intervient jamais ; c'est une interprétation qui n'est pas rencontrée dans les deux autres groupes. De plus, il faut noter que la justification suivante a été également trouvée uniquement chez les psychologues : « dès que l'on peut calculer une probabilité il n'y a plus de hasard » ; on peut se demander dans quelle mesure ceci n'est pas lié à leur pratique des tests de signification (*cf.* notamment Lecoutre, Lecoutre et Poitevineau, 2001), où le calcul d'un seuil est utilisé pour rejeter l'hypothèse du hasard.

IV. Conclusion : intérêt et limites de la méthode

La méthode des arbres additifs s'est révélée utile. Elle permet une structuration plus fine que celle que l'on aurait pu obtenir par une analyse des proximités (Cox et Cox, 2001). Ainsi ici elle nous a notamment permis de mettre en évidence la flexibilité des représentations suivant l'explicitation ou non de la notion de hasard.

Cependant les arbres obtenus sont, en quelque sorte, des arbres « moyens » calculés à partir des données agrégées dans chaque groupe de sujets, ce qui limite leur interprétation ; néanmoins celle-ci peut être affinée, dans la catégorisation contrainte, en examinant les patrons de réponses individuels. La construction d'arbres individuels nécessiterait une autre procédure de recueil des données pour pouvoir obtenir une matrice de dissimilarités pour chaque sujet. Sans que cela soit impossible, une telle procédure serait irréaliste et en outre très lourde dès que le nombre d'objets en jeu est un tant soit peu élevé (ici, un sujet devrait effectuer $16 \times 15/2=120$ comparaisons par paires pour apprécier toutes les dissimilarités entre les 16 situations).

Enfin, comme pour toute méthode de statistique descriptive, les résultats obtenus ne sont, en principe, applicables qu'aux sujets examinés. La généralisation à une population plus vaste pourrait s'appuyer sur des études de simulation, mais ce travail reste à faire et la généralisation est sans doute problématique.

Références

Barsalou, L. W. (1987) The instability of graded structure : Implications for the nature of concepts, *In* U. Neisser (éd.), *Concepts and conceptual development : Ecological and intellectual factors in categorization* ,101-141, Cambridge, Cambridge University Press.

Barsalou, L. W. (1993) Flexibility, structure, and linguistic vagary in concepts : manifestations of a compositional system of perceptual symbols, *In* A. F. Collins, S. E. Gathercole, M. A. Conway & P. E. Morris (éds), *Theories of memory,* Hillsdale, N. J., Erlbaum.

Barthélémy, J. -P. & Guénoche, A. (1991) *Trees and Proximity Representations,* Wiley.

Blaye, A. & Bonthoux, F. (2001) Thematic and taxonomic relation in preschoolers : the development of flexibility in categorization choices, *British Journal of Development Psychology,* **19**, 395-412.

Cox, T. F. & Cox, M. A. A. (2001) *Multidimensional Scaling. 2nd edition,* Chapman & Hall.

Guénoche, A. & Garreta, H. (2001) Can we have confidence in a tree representation ? Proceedings of JOBIM'2000, *Lecture Notes in Computer Sciences*, vol. 2066, 43-53.

Konold, C., Falk, R., Lipson, A., Lohmeier, J., Pottatsek, A. & Well, A. (1991) Novices views on randomness, Paper presented at the *13th Annual Meeting on the International Group for the Psychology of Mathematics Education,* Blacksburg, VA.

Lecoutre, B., Lecoutre, M.-P. & Poitevineau, J. (2001) Uses, abuses and misuses of significance tests in the scientific community : won't the Bayesian choice be unavoidable? *International Statistical Review,* **69**, 399-417.

Lecoutre, M.-P. (1992) Cognitive models and problem spaces in « purely random » situations, *Educational Studies in Mathematics*, **23**, 557-568.

Medin, D. L., Lynch, E. B. & Coley, J. D. (1997) Categorization and reasoning among tree experts : do all roads led to Rome ? *Cognitive Psychology*, **32**, 49-96.

Nisbett, R. E., Krantz, D. H., Jepson, C. & Kunda, Z. (1983) The use of statistical heuristics in everyday inductive reasoning, *Psychological Review*, **90**, 339-363.

Robinson, D. R. & Foulds, L. R. (1981) Comparison of phylogenetic trees, *Mathematical Biosciences*, **53**, 131-147.

Sattath, S. & Tversky, A. (1977) Additive similarity trees, *Psychometrika*, **42**, 319-345.