Structural comparison in social representation theory: A research proposal

IOANNIS D. KATERELOS
Panteion University, Athens, Greece

ABSTRACT In recent studies, many authors have argued that social representations differ structurally and, that, this difference is fundamental. Although, these studies use several methods of comparing structures, no numerical coefficient has been reported regarding this comparison. The theoretical construction of such a coefficient presupposes a definition of a structural distance between elements (of a social representation), so that their distance within the maximal similarity tree can be determined. Having defined the structural distance of the elements, we propose a coefficient differential for social representations based on their divergence. Three empirical studies using the coefficient are reported.

Key words: Differential coefficient, Social representation, Structure.

In this paper, we propose a coefficient of comparison of structures of social representations ("structure" is used as Abric defines it in his theory of central nucleus, cf. Abric, 1987) following the logic of the similarity analysis procedure. In a variety of studies dealing with the structure of social representations, the authors face the problem of comparing heterogeneous groups or comparing different states of the same group. In these cases, reference is often made either to simple parametric statistics (i.e., comparison of means) or the data are literary explored.

Without underestimating these efforts, which have produced valuable results, a coefficient that aims to examine and measure the structural distance between representations is considered to be useful. This coefficient was developed to meet the needs of a study (part of the doctoral dissertation of the author, Katerelos 1993) concerning comparisons of representations of the same social object, in different social groups. However, it can be used, in any study addressing similar problems provided that the same approach is used. The coefficient uses the maximal similarity tree (in other words, the shortest spanning tree, cf. Kruskall, 1956 and Rosenstiel, 1967). The methodology employed to construct the tree will not be discussed in this paper. Instead, we will use the similarity analysis approach pioneered by Flament (1981). Thus far, it appears that this coefficient has proven to be very useful in the following instances:

a. When we have two or more groups to compare and have reasons to believe that group A

Address: Ioannis D. Katerelos, Department of Psychology, Panteion University, L. Syggrou 136, 176 71 Athens, Greece. E-Mail: iokat@panteion.gr

and group B differ significantly concerning the social representation of a social object.

b. When we want to compare the same group at two different time intervals and have reasons to believe that the social representation has changed because something very important has happened in the interim.

c. When we attempt to introduce changes in a social representation, e.g., by making suggestions to a group of subjects concerning the social representation of an object and wish to measure the impact of these suggestions.

The theoretical basis of the coefficient: The structural distance between elements of a social representation

In the context of similarity analysis, the maximal similarity tree is generally presented as a complete graph without circles (Κατερέλος, 1996). Thus, there is only one path between two elements in the tree.

To explain maximal similarity trees, we use the criterion of proximity between elements in the tree. We claim that two elements can be considered to "go" together if they are closely related in the tree, in other words, if the two elements are in the same "neighborhood". Accordingly, if two elements are far away from each other, they can not be considered to "go" together. Based on these two properties of the maximal similarity tree, we can define a unique and unitary distance of 1 when two elements occupy consecutive positions on the tree (In graph terminology two consecutive vertices are called "adjacent".) Thus, a first order distance establishes "contact" between two elements. If two elements are not in contact, this means that there are other elements in between, we can count the distance between these two elements as shown in Figure 1. Table 1 is the derived distance matrix.

The logic of Table 1 is simple: Between A and B, there is a first order distance (contact), between A and C, there is a second order distance (one element in between), between A and K, there is a sixth order distance (maximum distance in this tree, five elements in between), and so on and so forth. In general, we firstly assign the value i and secondly the value j to the elements of a tree, and we suppose that i and j run from 1 to v if there are v elements. The distance or data value differentiating element i from element j we represent by Dij. We arrange the values Dij in a matrix which we call Δ. For example, if v=4, then

$$\Delta = \begin{bmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{bmatrix}$$

It is easily understood that the matrix is nothing more than a way of re-writing the tree. A distance specific to this tree defines the proximity of each pair of elements. In addition, this distance complies with the three properties of a metric measure. If x, y and z are elements of a maximal tree and D(x,y), D(x,z), and D(y,z) are the structural distances between them, it follows that:

a) $D(x,y) = D(y,x)$

b) $D(x,y) = 0 \Leftrightarrow x = y$

c) $D(x,y) \leq D(x,z) + D(y,z)$

These relations mean that: a) the distance between elements is symmetrical, b) if the distance between two elements is zero then these two elements are the same element, and c) the direct distance between two elements is always smaller than the distance between these two elements if there is another element in-between. Thus, the main diagonal of the matrix (from upper left to upper right) consists of 0s, and the matrix is symmetrical. Most commonly, only a halfmatrix of distances is collected, since in most cases there is no important distinction between Dij and Dji.

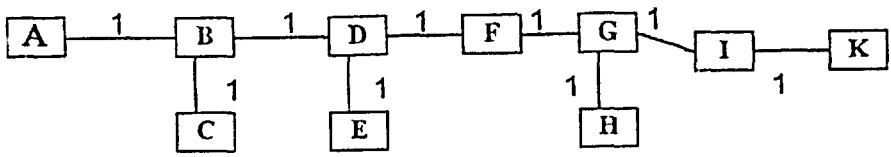


Figure 1: A maximal tree of 10 elements randomly created

Table 1:
Distance matrix between elements in the tree of Figure 1

A	0									
B	1	0								
C	2	1	0							
D	2	1	2	0						
E	3	2	3	1	0					
F	3	2	3	1	2	0				
G	4	3	4	2	3	1	0			
H	5	4	5	3	4	2	1	0		
I	5	4	5	3	4	2	1	2	0	
K	6	5	6	4	5	3	2	3	1	0
	A	B	C	D	E	F	G	H	I	K

Distance between two maximal trees

Suppose that we have two trees (A and B) which have the same elements but different structures. If a distance specific to a tree (Tree A) defines each pair of elements, then in an another tree (Tree B) this same pair of elements would be defined by either a different or the same distance. In this way, if all distances are the same, then the two trees are identical. The bigger the difference between the pairs of elements between the trees, the lower the chances that the interpretation of the two trees will be similar. Although we can use many ways of comparing two series of values, and the distances of pairs in the two trees, we shall present the one that appears more often in the literature. This is the city-block distance, a metric which is used as a measure of similarity between groups being well known.

The city block distance. The city-block distance (CBD, also called Manhattan Distance) is the sum of absolute differences between the series of data values (Verges, 1994). The choice of this index is based on the fact it is frequently used in similarity analysis studies within the theory of social representations. In this case

there is no effective difference in meaning between Dij and Dji and there is not at all difference for Dii (equal to 0); thus, the data values do not form the entire matrix on each tree, but only a part of one. The number of pairs of a symmetrical distance, like the structural distance, for a given number of elements (v) (except for the values of distance of each element with itself which equals to 0), is:

$$N = \frac{v(v-1)}{2}$$

Thus: (Formula A)

$$CBD\ (TA, TB) = \sum_{i=1}^{v(v-1)/2} (|DTA_i - DTB_i|)$$

Where:

DTAi: Distance between elements of the pair i in the Tree A

DTBi: Distance between elements of the pair i in the Tree B

The CBD, being a measure of distance between trees, can vary between 0 and infinity (depending on v). If the trees are "similar", we can logically expect that the distances between elements in both trees will also be "similar". In this case, the sum of differences between distances will be minimal (equal to 0 if they are exactly the same trees): same distances for the same pair of elements in both trees.

To the extent the trees are not "similar", we can expect this measure to reach maximal levels. If elements which are close to each other in one tree and, thus, can be interpreted globally, are far away in the other tree (in which case we can not find a common interpretation by neighborhood), then the sum of differences between the two trees will be maximal.

However, the traditional way to describe a desired relationship is by means of a single number, which shows how well (or how poorly) the two trees fit each other; this number should vary between 0 and 1. This brings us to the preliminary phase of the computational procedure, namely, finding the longest possible distance (Distance Maximal, DM) between two trees of v elements. We define this distance to be:

$$DM = \max DT(v)$$

where DT(v): distance between two trees of v elements

Calculating Distance Maximal (DM)

Although the choice we applied can be supported by many parallels in statistics and other fields as well as by intuitive reasoning, we will not elaborate it because the issue falls outside the main goal of this paper. Essentially, one procedure for finding the DM is the intuitive definition of two "opposite" trees with v elements. Searching for the two opposite configurations with the maximum sum of differences entails that we seek to find all the trees that can be made with a definite number of elements. Nevertheless, since it is impossible to get an overview of all the trees, it is like looking for *the* difference blindfolded. We need, therefore, a starting point for the search. Sometimes we pick random matrices of distance for this purpose: it is like a scuba diver diving from a helicopter into the sea on a dark night. In practice, the possibility of finding a local maximum other than the global maximum is real. Suppose that we have two principal types of trees: the *star*, in which there is one central element and all the others are around, and the *line*, in which all the elements are aligned (see Figure 2).

By using these two extreme configurations, it is like assuming that all other trees can be a mixed type of these two. Rather than going into detail regarding the choice of these two types of configuration, we shall simply start with the derived structural distance matrices between elements.

Type I: Star. This basic type of tree can be described as follows: If element 1 is the center of the star and the tree has five elements, then

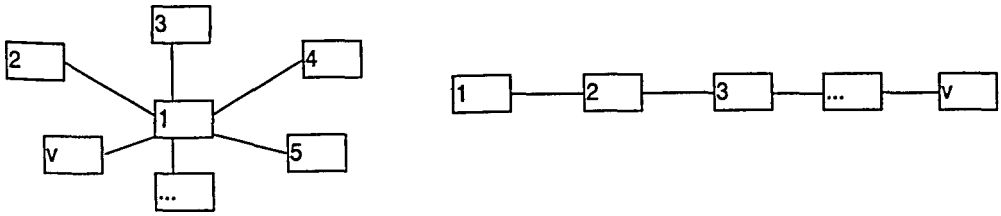


Figure 2
Two types (star and aligned) of trees with v elements.

$$\Delta star = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 2 \\ 1 & 2 & 0 & 2 & 2 \\ 1 & 2 & 2 & 0 & 2 \\ 1 & 2 & 2 & 2 & 0 \end{bmatrix}$$

In general, in all trees of this type, the central element has a distance of 1 with all other elements. All the other distances between elements are equal to 2.

Type II: Aligned. This type of tree can be rewritten in a distance matrix such as: If element 1 is the first element of the "chain" and the tree has five elements, then

$$\Delta aligned = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{bmatrix}$$

For the most part, in all trees of this type, the distances between the first element of the chain and all the others (v-1) is 1 to v-1 respectively. Thus, the following element is in a distance of 1, the element after that is in a distance of 2, ..., the last element is in a distance of v-1. The same rule can be applied to the rest of the elements.

We then apply formula A in order to measure the expected DM. Without describing in detail the computational algorithm, we obtain the DM=f(v) where v is the number of elements in each tree:

$$DM = \sum_{i=1}^{v-1} |1 - i| + \sum_{i=2}^{v} \sum_{i=1}^{v-i} |2 - i| \quad \text{(Formula B)}$$

Thus, the distance between the configurations expressed by the trees TA and TB, or the structural distance (SD) between two trees can be estimated as follows:

$$SD = \frac{Formula\ A}{Formula\ B}$$

or

$$SD = \frac{CBD\ (TA, TB) = \sum_{i=1}^{v(v-1)/2} (|DTA_i - DTB_i|)}{DM = \sum_{i=1}^{v-1} |1 - i| + \sum_{i=2}^{v} \sum_{i=1}^{v-i} |2 - i|}$$

(Formula C)

This structural distance (SD) between two trees can vary between 0 ("similar") and 1 ("unsimilar"). If this distance is SD then the Structural Similarity Coefficient (SSC) is given beneath:

$$SSC(TA, TB) = 1\text{-}SD(TA,TB)$$

and, obviously SSC measures the "similarity" between two trees with the same elements but different structures and can vary between 0 ("unsimilar") and 1 ("similar").

Applications

Application 1: The representation of teachers in a state of "certainly predicted reversibility"

In a previous study (Katerelos, 1993), we had four groups of teachers working in "schools with difficult pupils". In such schools, teachers face many special difficulties. Thus, many teachers want to be transferred. "Certainly predicted reversibility" states that if a social subject (individual, group) is "submerged" in conditions contradictory to his/her representation, then, normally, transformation mechanisms are activated in order to adapt him/herself to context. The activation of these mechanisms is postponed, or cancelled, if the subject believes that these conditions will change again to the status quo ante: thus, there is no essential need to change... The four groups studied were:

- The Group 1 wanted to be transferred from this school but they could not because the administration did not allow it yet.
- The Group 2 were newcomers (less than 1 year of service) in this school but they wanted to be transferred.
- The Group 3 wanted to stay in this school.
- The Group 4 were newcomers (less than 1 year in service) in this school and they wanted to stay.

The task of all participants was to categorize twenty items (e.g., affection, transfer of knowledge, one-way communication, etc.) regarding the educational relationship, in five categories (from "very relative to the educational relationship" to "very distant from the educational relationship") of four items each. A similarity analysis was applied and four maximal trees were constructed from the data (Figure 3).

The central hypothesis was, that due to certainty predicted reversibility, the Groups 1 and 3 would present the same behavioral pattern potential: they both believe that they will be transferred, so there is no need to transform their representations in order to adapt themselves to the present context. Note that this period of time can last for as long as twenty years.

According to Figure 4, we can claim that our hypothesis was verified. Groups 1 and 3 were closely related relative to the other groups. There is much more distance between Groups 2 and 4, because these groups must transform their representations in order to comply with the context. Obviously, Group 2 is more motivated for staying at this type of school than Group 1. However, Groups 1 and 2 have a close similarity when they are both newcomers. As time passes, the distance between them is maximized (Group 3- Group 4).

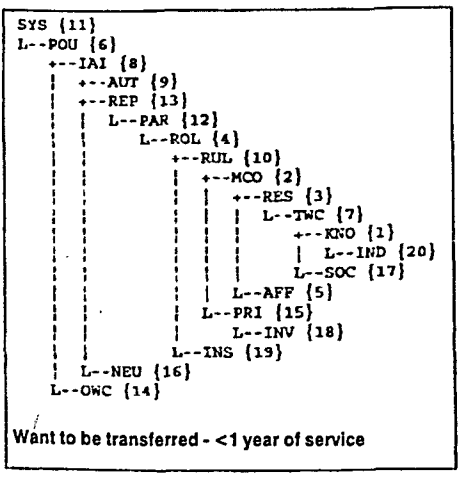
Application 2: The social representation of mental illness

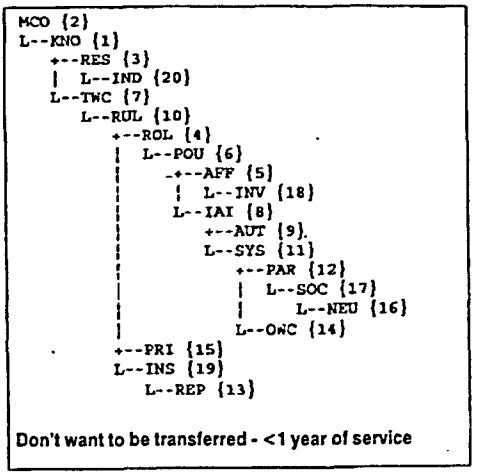
In a study by Pradeilles (1992), three categories of participants responded to a common questionnaire of twenty items about mental illness (dangerous, contagious, it can happen to everybody, incurable, etc.). The three categories of participants were:

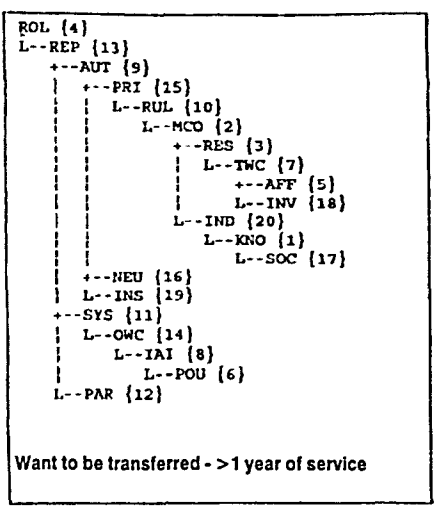
- Nurses who work in hospitals treating mental illness.
- Adults (over a certain age).
- Young people (under a certain age).

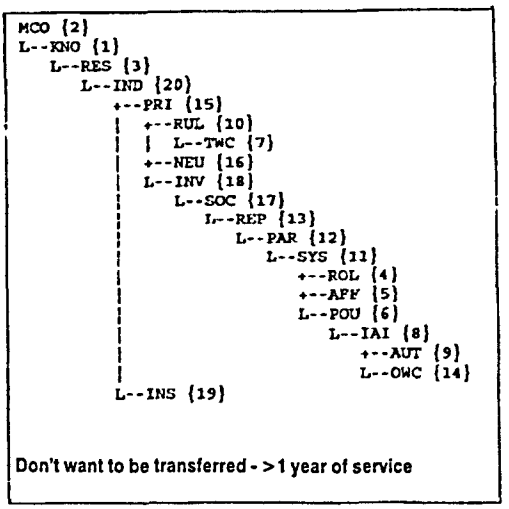
The main hypothesis was that among young people a new element would appear in the categorization of mental illness, namely, "Nervous Breakdown". This element is treated neither like "brain damage" nor like "neural damage" (Jodelet, 1989). The "nervous breakdown" exists also within the other representations but it does not occupy a central position in them. Somehow, the "nervous breakdown" is a new "modern" category of mental disease that is curable and can happen to anyone.

We notice that all three values of the coefficient are relatively low. In studies concerning the social representation of teachers the values range from .51 to .67. In the Pradeilles' study the values ranged from .42 to .48. Since the denominator of the fraction (Formula C) is constant (both representations were examined by *means of twenty*









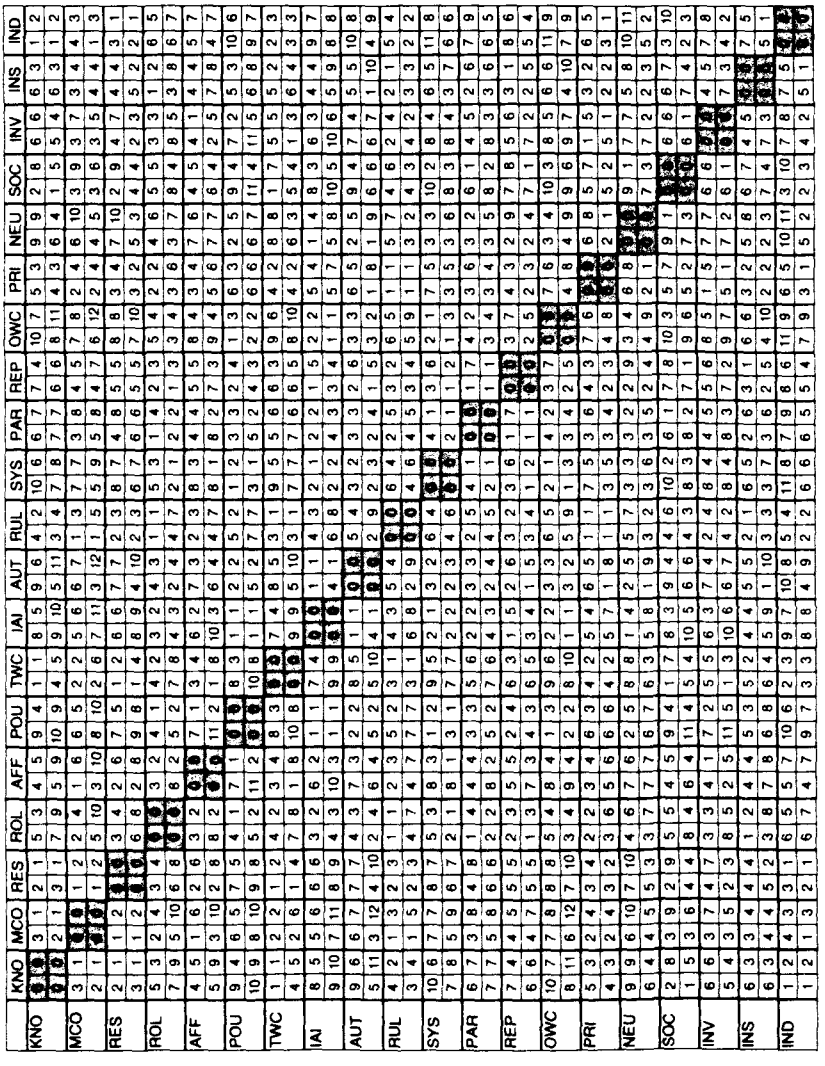
Abbreviations Index:

AFF	An affective relation	RUL	Respect of mutual rules
IAI	One individual (teacher) "builds" another individual (pupil)	PRI	A privileged relation between youth & adults
TWC	Two-way communication	POU	An exerted domination
OWC	One-way communication	REP	A socio-cultural reproduction of society
MCO	Mutual confidence between the protagonists	RES	A mutual respect between the protagonists
PAR	A parental relation	ROL	Playing institutional roles
INS	A relation of instruction	KNO	Transmission of knowledge
INV	An investment by both the protagonists	SOC	A socialization process
IND	A relation aiming the pupil's independence	SYS	An omnipotent system of social authority
NEU	Personal neutrality towards knowledge	AUT	Installation of automatism in the pupil's mind

Figure 3

The four maximal similarity trees extracted for the social representation of educational relationship.

Table 2
Difference matrices concerning the four groups of teachers



Note: The abbreviations index of the elements is given in Figure 3.
Reading instructions:

	Element 2	
Element 1	*Distance on tree:* Want to be transferred - < 1year of service	*Distance on tree:* Don' t want to be transferred - < 1year of service
	Distance on tree: Want to be transferred - > 1year of service	*Distance on tree:* Don't want to be transferred - > 1year of service

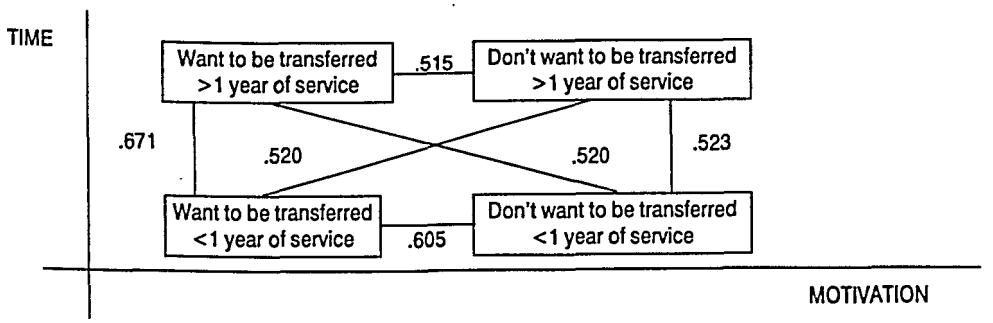


Figure 4
Structural similarities between the representations of teachers.
(Vertical dimension is time serving in this school, horizontal dimension is «motivation».)

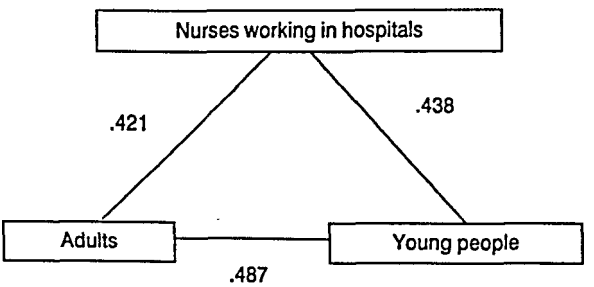


Figure 5
Structural similarities between the representations of three populations regarding the menatl disease.

elements), this means that the consensus regarding the representation of mental disease is lower than that of the representation of educational relationship in difficult schools: the representation of mental disease presents many more differences between the populations studied. On the other hand, we can say that nurses are highly differentiated from the general population, because they are professionally involved in that context.

Application 3: The social representation of hunting

In a study by Guimelli (1989), three categories of participants (all hunters) responded to a questionnaire of twenty items about hunting (legally using a gun, an outdoor sport, respect for animals, financing protection of nature projects, etc.).

The participants were aware of the fact that several years ago, a disease (myxomatose) killed most of the hares in France. At the beginning, the hunters thought that the high rate mortality in the hare population was temporary and, therefore, there was no need for them to change their representations by incorporating (in the central nucleus) ecological concerns (e.g., protecting wild animals or creating special sites for their proliferation and survival). However, later on,

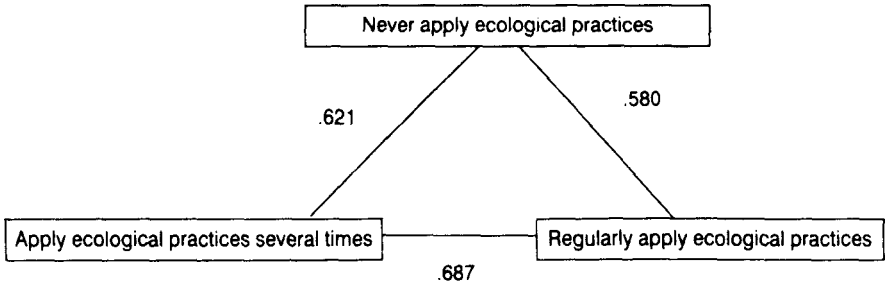


Figure 6
Structural similarities between the representations of three groups regarding hunting and ecological practices.

when they realized that this situation is irreversible, most of the hunters should operate within the context of ecology.

The three groups of participants were:

- Hunters who never operated within the context of ecology.
- Hunters who have operated within the context of ecology several times.
- Hunters who regularly operated within the context of ecology.

In the Guimelli's study, the central hypothesis was that during the years from the onset of the disease to the present, the social representation of hunting underwent a slow and continuous change towards a more ecological behavior. This means that the ecology schemata, which were latent and selectively applied, were largely mobilized and, at the end, occupied an important place in the representation. In Figure 6, we notice that selectively applying ecological practices seems to be the first step for engaging in ecological practices regularly.

Discussion

In this paper we described a coefficient which attempts to deal with some questions regarding the comparison of social representations by defining structural distance between elements within the context of similarity analysis. Of course, we are aware of some major weaknesses of similarity analysis, which reflect on the construction of the maximal tree:

- The treatment of equal values: Since the algorithm (for constructing the tree) ranks all values of similarity between elements in ascending or descending order, the program, which has been constructed for estimating SSC, simply takes the one, which comes first in the file.

- The treatment of almost equal values: For the same reason, if two values differ even in the third decimal, the algorithm will use the bigger value. This means that: a) significant values of similarity are neglected, and b) the maximal tree is extremely sensitive and unstable. The fact that the whole tree changes by changing one answer of one subject supports evidence for the previous comment.

Although the coefficient seems to work satisfactorily, we can still pose the question of a more differentiating coefficient. Our recent efforts are oriented towards using the distances in powers of 2 or 3 or even more. For example:

$$CBD\ (TA, TB) = \sum_{i=1}^{v(v-1)/2} (DTA_i^2 - DTB_i^2)$$

or

$$CBD\ (TA, TB) = \sum_{i=1}^{v(v-1)/2} (DTA_i - DTB_i^2)$$

The results seem to be promising but nonetheless the construction of the coefficient resembles more the nonparametric x^2 than the other parametric tests. (If we admit that the scope of this paper is the construction of a parametric index.) Using powers is appropriate because if two elements stand in a distance of 1, they can be interpreted together. If they stand in a distance of 2, the common interpretation is more difficult, because many elements are in the same distance, in a distance of 3, a great deal of elements are interpreted together and, in a distance of 4, almost half the tree is composed. Thus, we can assume that the second or the third power of the distance used diminishes the power of reasonably global interpretation with respect to particularities of each pair of element. Thus, if on tree A the two elements have a distance of 1 (which permits a direct reference from one to the another) and, a distance of 3 on tree B (which almost forbids the close interpretation), then the amount of difference between the two trees must be greater than 2.

We believe that this issue is not closed: on the contrary, we think that there is a long way to go before researchers uncover the "hidden structure" of data based on verbal answers. Social scientists, who are instrumental in developing such techniques, must use them only to the limits of their internal validity.

References

Abric, J.-C. (1987). *Coopération, compétition et représentations sociales* [Co-operation, competition and social representations]. Cousset, Suisse: DelVal.

Flament, C. (1981). L'analyse de similitude: Une technique pour les recherches sur les représentations sociales [Similarities analysis: A technique for the research on Social Representations]. *Cahiers de Psychologie Cognitive, 1*, 375-385.

Guimelli, C. (1989). Pratiques nouvelles et transformation sans rupture d'une représentation sociale: La représentation de la chasse et de la nature [New practices and transformation without disruption of a social representation: The social representation of nature and hunting]. In J. L. Beauvois, R. V. Joulé, et J. M. Monteil (Eds.), *Perspectives cognitives et conduites sociales: Tome 2. Représentations et processus cognitifs* (pp. 117-138). Cousset, Suisse: DelVal.

Jodelet, D. (1989). *Folies et représentations sociales* [Madness and social representations]. Paris: Presses Universitaires de France.

Katerelos, I. (1993). *Pratiques, conditionnalité et sous-structuration au sein des représentations sociales* [Practices, conditionallity and sub-structure in social representations]. Thèse de Doctorat de l' Université de Provence, Aix-en-Provence, France.

Κατερέλος, Ι. [Katerelos, I.] (1996). Κατά συνθήκη συμπεριφορά στις κοινωνικές ανα-παραστάσεις [Conditional behavior in social represenattions theory]. *Επιθεώρηση Κοινωνικών Ερευνών, 91*, 123-144.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society, 7*, 48-50.

Pradeilles, C. (1992). *La représentation sociale de la maladie mentale en Lozère* [The social representation of mental ilness in Lozère]. Thèse de Doctorat de l'Universite de Provence, Aix-en-Provence, France.

Rosenstiehl, P. (1967). L'arbre minimum d'un graphe [The minimum tree of a graph]. In P. Rosenstiehl (Ed.), *Théorie de graphes* (pp. 357-368). Paris: Dunod.

Vergès, P. (1994), Approche du noyau central: Propriétés quantitatives et structurales [A cen-tral nucleus approach: Quantitative and structural properties]. In C. Guimelli (Ed.), *Struc-tures et transfomations des représentations sociales* (pp. 233-249). Neuchâtel, Suisse: Dela-chaux et Niestlé.