Classifying infants in the Strange Situation with three-way mixture method of clustering

Pieter M. Kroonenberg*

Department of Education, Leiden University, Wassenaarseweg 52, PO Box 9555, 2300 RB Leiden, The Netherlands

Kaye E. Basford

Department of Agriculture, The University of Queensland

Marion van Dam

National Court of Audit, 's-Gravenhage, The Netherlands

The quality of the attachment relationship between mother and infant is typically determined in the Strange Situation. The assignments of infants to the A (avoidant), B (secure), and C (resistant) attachment classes are largely but not exclusively based on measurements during the reunion episodes. In this paper, the measurements in the reunion episodes are used to derive a clustering of the infants via three-way mixture method of clustering, a technique especially designed for clustering three-way data (here: infants, variables and episodes). The results are compared with the A–B–C classification, and the relevance of the outcomes for attachment research are discussed. At the same time, the paper aims to demonstrate the use and usefulness of the three-way clustering procedure for data from the social and behavioural sciences.

In developmental psychology there always has been a strong interest in the attachment relationship between mother and child. One of the main instruments to assess the quality of attachment has been the Strange Situation (see Ainsworth, Blehar, Waters & Wall, 1978). Measurements are made in several episodes of this laboratory-based procedure, and the final assessment is a classification of each infant (or better, infant–mother relationship) into one of three categories, usually indicated with the letter A, B, and C (see Ainsworth *et al.*, especially pp. 334–335, 343–362). The classification is established via detailed scoring rules using the above-mentioned measurements.

The major evaluation of the classification procedure as outlined in Ainsworth *et al.* (1978, chapter 6) was to perform a *post hoc* discriminant analysis using the classification as the criterion and the measurements as predictors. A disadvantage of such a procedure is its circularity. First the measurements are used to create the classifications, which in turn are evaluated by using the measurements. This was

* Request for reprints.

recognized by the originators, but at the time no clear alternative seemed to be available. In the present paper we will attempt to provide an assessment of the classification results without using the same data twice by using a clustering procedure. Clustering procedures attempt to find groupings of subjects (objects, etc.; here: infants) on the basis of the measurements available. The resulting grouping (or *statistical*) classification can be compared with the usual *clinical* or A–B–C classification to assess the latter (see Sawyer, 1966 for a relevant discussion of clinical versus statistical prediction; the term 'clinical' is seldomly used in this context, but was inspired by the paper by Richters, Waters & Vaughn, 1988). If the groups in the A–B–C classification adequately portray the individual differences in the Strange Situation, then the classes found by the clustering procedure should correspond reasonably with A–B–C classification, especially if the Ainsworth groups are natural clusters. If there is little or no correspondence, the empirically derived clusters will contain mixtures of the Ainsworth groups (see also Lamb, Thompson, Gardner & Charnov, 1985, p. 214).

As we are dealing with three-way data, our prime clustering tool will be the *mixture method of clustering for three-way data* (Basford & McLachlan, 1985*b*; McLachlan & Basford, 1988). As there are to our knowledge no applications of the three-way mixture method of clustering outside the field of agriculture and biological sciences (see e.g. Basford, Kroonenberg & DeLacy, 1991), we have included an appendix with some of the technical details, while a more conceptual introduction including some remarks about interpretation is given in the main body of the paper. Unique to the three-way cluster method is that it can handle explicitly data which arise from the type of three-way designs that form the basis of the present data set. In particular, during the two so-called *reunion episodes* of the Strange Situation, five variables measuring the intensity of the infants' behaviours were scored from videotapes for 326 Dutch infants. Thus, the data set can be seen as a 326 by 5 by 2 three-way data array.

In this paper we will first provide an exposé of the three-way mixture method clustering, and add a short discussion of an ordination technique to present some of the results graphically, i.e. three-way replicated principal component analysis. Then the substantive background to the data will be presented. To appreciate the results of the three-way method on these data, we will dwell also on the results for two-way data, especially because several aspects of the mixture method of clustering can be more easily demonstrated on two-way data. In particular, we start with analysing the two episodes separately, and only then continue with the data set as a whole. The results of the last analyses will be compared with the clinical classification to evaluate the characteristics of both the clinical and statistical classifications.

Method

There has been a long history of 'grouping' approaches when analysing three-way data, probably starting with Tucker & Messick (1963) devising a 'points-of-view' approach where the aim was to seek to partition two-way proximity matrices from several sources into relatively homogeneous subgroups, aggregate within, and then run an analysis for each subgroup. Carroll & Arabie (1983) devised a method for non-hierarchical overlapping clustering called INDCLUS for the case of three-way proximity data with either directly measured or derived proximities. Carroll, Clark & DeSarbo (1984) developed a

methodology called INDTREES for fitting a hierarchical tree structure to obtain a discrete network representation of proximity data. De Soete & Carroll (1989) have made further extensions to this approach using ultrametric trees (for an overview see Arabie, Carroll & DeSarbo, 1987). As far as we are aware, the mixture method of clustering is the only clustering technique which handles three-way profile data (i.e. subjects by variables by conditions data) directly.

The more usual clustering approaches (to clustering individuals, say) are the hierarchical ones in which by successive fusion the number of individuals (and subsequently individuals and clusters) is reduced by one, until one large cluster remains. As there is no explicit model underlying such clustering procedures, it is extremely difficult to evaluate the optimal number of clusters and the adequacy of the cluster solution for the data. Moreover, once two individuals are allocated to the same cluster in an hierarchical clustering, they will never be separated again. This is in contrast with the mixture method of clustering where for each number of clusters a new solution is sought independent of the solution with fewer or more clusters.

Under the mixture approach to clustering (Everitt, 1980; Wolfe, 1970), it is assumed that the data at hand can be considered as a sample from a mixture of several populations in various proportions. Estimates of the parameters of the underlying distributions can be obtained using the likelihood principle, and the elements can be allocated to these populations on the basis of their estimated posterior probabilities of group membership. In this way, individual observations can be partitioned into a number of discrete, relatively homogeneous groups. However, if the posterior probability is less than a specified value the individuals concerned can remain unallocated but with known probabilities of belonging to the various groups. The three-way mixture method of clustering to be presented is a direct generalization of the two-way variant developed by Wolfe (1970), and the two methods are equivalent for a single sample. Even though the estimation of the two-way case can be solved with the three-way program, special software exists for the two-way case (see below).

Three-way mixture method of clustering: Theory

In this section we will give an outline of the three-way mixture method of clustering. We will skip some details and present a primarily intuitive and simplified introduction. A more detailed and general exposition is contained in the appendix. The clustering method is based on the assumption that each of the 326 ($= N$) infants belongs to one of g possible groups, but it is unknown to which one. Therefore, all one can do is assign it to that group to which it has the highest probability of belonging. In order to be able to do so, it is necessary to establish the characteristics of each group and the probability of each infant to belong to each group. Observations are available on 5 ($= p$) variables at 2 ($= r$) separate times or episodes. Thus the data comprise two vectors of multivariate observations on each infant, one for each episode. For infant j ($j = 1, \ldots, n$) these will be denoted by x_{j1} and x_{j2}, respectively. The vector of all observations for infant j has 10 elements, and will be denoted by x_j.

Let us first assume that there exists only one group, then the data would have come from two multivariate normal distributions, one for each episode. When there are more groups, then the normal distributions of the variables in each group are allowed to have different means in each episode, thus there are $p \times r \times g$ ($= 5 \times 2 \times g$) means to be estimated. Furthermore, it will be assumed that the multivariate distribution for each group will not change between episodes, but groups may or may not have different covariances, i.e. the groups may have a *common* (within-group) covariance matrix, or each group may have an *arbitrary* (within-group) covariance matrix. Thus there is either one covariance matrix or there are g covariance matrices to be estimated. Under the normal mixture model proposed by Basford & McLachlan (1985b) for three-way data, it is assumed that the relative sizes of the g groups $G_1, \ldots, G_g$, are given by the *mixing proportions* $\pi_1, \ldots, \pi_g$ respectively, which are also unknown, so that they, too, have to be estimated.

The estimation is carried with maximum likelihood procedures, and once a solution has been obtained, one can derive the so-called *posterior probabilities* that an infant j, ($j = 1, \ldots, n$), with observation vector x_j, belongs to group G_i ($i = 1, \ldots, g$) (see equation 5 in the Appendix). Each infant is then assigned to the group to which it has the highest estimated posterior probability of belonging. In this way, the individual infants are partitioned into a number of discrete, relatively homogeneous groups.

A tendency has been observed for the derived clusters to be of roughly equal size when the covariance matrices are specified to be equal (see e.g. Gordon, 1981, p. 52). Clearly, if the model is incorrect, for instance, if the parent populations are decidely non-normal, the method might be far from optimal (see e.g. Everitt, 1980, section 5.2). Whether in real life this is a problem depends on the aim of the clustering, i.e. whether one is seeking for natural clusters, or wants to dissect continuous observations. In the latter case, the multinormality assumption seems a reasonable one to make, in the former case it is very much an empirical matter whether the method is appropriate.

Testing for the number of components g in a mixture is an important but very difficult problem which has not been completely resolved (McLachlan & Basford, 1988). An obvious way of approaching the problem is to use the likelihood ratio statistic λ, as discussed by Wolfe (1970, 1971), to test for the smallest value of g compatible with the data. Many authors, including Wolfe, have noted that unfortunately with mixture models, regularity conditions do not hold for $-2\ln\lambda$ to have its usual asymptotic null distribution of chi squared with degrees of freedom equal to the difference in the number of parameters in the two hypotheses.

McLachlan & Basford (1988) recommended that in general the outcome of Wolfe's likelihood ratio test should not be rigidly interpreted, but rather used as a guide to the possible number of underlying groups. They suggested that use also be made of the estimates of posterior probabilities of group membership in the choice of g. They can be examined for values of g near to the value accepted according to the likelihood ratio test, and may therefore be of assistance in leading to a final decision as to the number of underlying groups.

When the number of components in the mixture is known (as may be the case here with $g = 3$, i.e. equal to the number of clinical classification groups), Everitt (1981) stated that the parameters in the model 'may be estimated by maximum likelihood methods although problems may arise due to singularities in the likelihood function unless some constraint is placed on the variance–covariance matrices, the most natural being that these are the same for all components' (p. 172). Note, however, that although it may be natural to assume equal covariance matrices, in many applied cases (including here) the data do not really conform with that assumption. A detailed discussion of the problem is given in Duda & Hart (1973, pp. 198–201).

The mixture model used here assumes that the measurements taken on infants during the separate episodes are independent of each other, in that the likelihood was obtained by multiplying together the probability density function for each infant in each episode. There is no concern about the infants being independent of each other, but the same infants were measured during both episodes. Therefore, we are really assuming independence of measurements on the same infants over time. Independence was a valid assumption for the agricultural data to which the model had been previously applied (the same genotypes, but separate plants, were grown in each location), but may be open to criticism with social science data of this type. Treating the separate episodes as independent measurements is a compromise which enables some of the structure of the design of the experiment to be accommodated, i.e. the same five variables are measured each time.

Note that the exploratory nature of the clustering methodology is being stressed here. Although we probably have as much experience as anyone in the application of the three-way mixture method of clustering, we do not know about the robustness of the method to the violation of this independence assumption. It is hoped that by application to repeated observation data, such insight will be gained. Other clustering techniques simply stack the observations for the two episodes together and consider the data array to be two-way (infants by variables). If the outcome of the three-way mixture approach is supported by the outcome of the complementary ordination procedure discussed below, then the independence assumption cannot be too rigid. We are using an alternative methodology on a particular data set to look at a substantive problem, rather than choosing a data set to best demonstrate the methodology.

Ordination as a complement to clustering

In the cluster analysis the variables and episodes are jointly used to find an optimal separation of the infants into groups. The mean values of the groups on the variables can be used to evaluate the characteristics of the groups with respect to each other in either episode. The differences between the

infants are only described through the clusters, but not on an individual level other than their *a posteriori* probabilities of belonging to the groups. By using ordination techniques in conjunction with cluster analysis, a more detailed analysis of the individuals and of the clusters can be obtained. Within social and behavioural sciences, principal component analysis is probably the most common, ordination technique for two-way data. For three-way data, such as we have here, several variants of principal component analysis exist, such as three-mode principal component analysis (see e.g. Kroonenberg, 1983), and parallel factor analysis (see Harshman & Lundy, 1984). As in the present case we only have two episodes, we will use a very simple model for the three-way data, sometimes called *replicated (or weighted) principal component model*. This model is a special case of both three-mode principal component and parallel factor analysis models (see Ten Berge, De Leeuw & Kroonenberg, 1987).

In replicated principal component analysis, the data from two or more occasions or episodes are assumed to have the same configuration for the variables and for the infants for each occasion, but the relative size or importance is allowed to vary from one occasion to the next. Note that this assumption fits very nicely with the assumption of the mixture method of clustering that per group the covariance matrix is the same for each occasion. The cluster technique allows interaction between the variables and the occasions by modelling a different mean for each variable in each group at each occasion. In the replicated component model as used here, the overall means of the variables for each occasion are modelled separately, and do not form an integral part of the three-way model. As we will see the differences across occasions between the means of the groups are not extremely large, so that a reasonable concordance between the results of the cluster analysis and the component analysis should be possible.

The replicated principal component model for three-way data can be written as

$$x_{jlk} = c_k \sum_{s=1}^{s} a_{js} b_{ls} + e_{jlk} \, (j = 1, \ldots, n; \, l = 1, \ldots, p; \, k = 1, \ldots, r), \tag{11}$$

where c_k is the weight or relative contribution for episode k, the a_{ls} are called component scores and b_{ls} the component loadings, and j is the index of the infants, l the index for the variables, and S the number of components; e_{jlk} is the error of approximation. Generally, one is only interested in a small number of components, say two or three. This partially depends on the dimensionality of the space in which the clusters can be shown to their greatest advantage. The components from this technique will be used to make a simultaneous plot of the component scores, the component loadings, and the clusters. In this way, insight can be acquired about the distinctness and tightness of the clusters in relation to the variables and the individual infants.

Computer programs

The three-way mixture method of clustering is implemented in the computer program called MIXCLUS3, and can be obtained from the second author (as can the two-way version of the program, MIXCLUS2). An earlier version of this program was published as an appendix in McLachlan & Basford (1988). The replicated principal component analyses have been carried out with a program for three-mode principal component analysis (TUCKALS3) developed by the first author (Kroonenberg, 1994; Kroonenberg & Brouwer, 1993).

Substantive background

Strange Situation procedure

The Strange Situation procedure was designed to assess the quality of infant–mother attachment, and is considered appropriate for infants between 12 and 18 months (for details, see Ainsworth *et al.*, 1978). This laboratory-based procedure consists of seven three-minute episodes arranged to continuously increase the stress on the infant, so that its attachment system with respect to the mother is activated. To be specific, mother and infant are taken through a series of episodes each lasting three minutes in a (to the infant unfamiliar) room at a laboratory or institute. The crucial episodes, the *reunion episodes* (R 1

and R2), are those in which the mother returns after having been away, and the infant was left alone with a stranger. On the basis of the infant's behaviour during the procedure, the quality of the infant–mother attachment relationship is categorized as *insecure–avoidant* (A), *secure* (B), or *insecure–resistant* (C). It should be noted that a fourth category has recently been added, the D or disorganized classification (see Main & Solomon, 1990). However, this classification category will not be considered here as it has not yet been coded for our data.

Three important claims have been made with respect to the reliability and validity of the A–B–C typology, i.e. the clinical classification. First, different patterns of behaviour in the Strange Situation arise from different previous patterns of infant–mother interaction. In particular, it is the mother's sensitivity to the behaviour of their infants that leads to secure attachment. Second, infants seen more than once in the Strange Situation tend to behave in the same fashion each time they are measured. It should be mentioned that attachment classifications remain stable over a period of one to six months provided family circumstances are stable. Third, individual differences in Strange Situation behaviour predict behavioural differences in other contexts up to several years later, again provided there has been stability in the family circumstances (see among others, Lamb *et al.*, 1985; and Van Dam, 1993, for further references). The above suggests that the clinical classification has both a certain amount of reliability and validity.

According to the classification instructions (see Ainsworth *et al.*, 1978, pp. 59–63) the scores on five seven-point scales in the two reunion episodes play a crucial role in the clinical classification, i.e. proximity seeking (PS), contact maintaining (CM), resistance (RS), avoidance (AV), and distance interaction (DI), where high scores on avoidance are especially indicative for an A classification, and high scores on resistance for a C classification. Therefore, we will also restrict ourselves to these variables (see also Lamb *et al.*, 1985, pp. 209, and Richters *et al.*, 1988).

The only other clustering of strange situation measurements known to us (Lamb *et al.*, 1985, pp. 214–221) was also restricted to these same 10 variables, but Lamb *et al.* used hierarchical two-way cluster methods on a sample consisting of Swedish and American infants. Connell (1977) also claims to have carried out a cluster analysis, but closer inspection shows that in fact he used an ordination technique.

Data

A total of 326 infants, or rather infant–mother pairs, are included in our analyses. They originate from five different studies conducted at the Centre for Child and Family Studies of the Department of Education, Leiden University. The primary references for these studies, which also contain the detailed information on procedural questions, are Goossens (1986; see also Van IJzendoorn, Goossens, Kroonenberg & Tavecchio, 1985), Goossens & Van IJzendoorn (1990), Hubbard & Van IJzendoorn (1991), Lambermon & Van IJzendoorn (1989), and Van Dam & Van IJzendoorn (1988); a comprehensive description can be found in Van Dam (1993).

As the data originate from several samples, which were collected for different purposes, the reliability information is not entirely complete for all of the scales in all studies. A report containing all available details can be obtained from the first author. A typical example of the reliability of the scoring of the five variables can be found in Van IJzendoorn *et al.* (1985, p. 441). The inter-coder reliabilities varied from .73–.97, and similar values have been found in the other subsamples. With respect to the A–B–C classification, a subset of the same study was independently reclassified which gave a correct classification rate of .96, and again similar values were found in other subsamples. In as far as was feasible, the classifications were done by other persons than the scorers of the five variables. Correlations between the same variables across episodes for the whole sample are given in Table 6, and these values range from .35 (avoidance) to .71 (distance interaction).

Attachment: A continuous or discrete construct?

Ainsworth intended the Strange Situation to be primarily a classification procedure, with which individual differences in attachment could be assessed in terms of organized patterns of behaviour (see e.g. Ainsworth, 1990). The A–B–C typology wass originally empirically developed, and there seems to

be little theoretical reason to expect that individual differences in Strange Situation are discrete rather than continuous. Working with the protocols of 23 infants, Ainsworth *et al.* (1978) developed the classification system by grouping infants in clusters on the basis of perceived behavioural similarities in the strange situation. Similarities between the resulting (seven) clusters were then used to achieve a further condensation to three main groups, i.e. the A, B and C classifications. After the classification of infants, Ainsworth *et al.* (1978) identified aspects of behaviour that seemed crucial in distinguishing the various groups and subgroups in the classification. These aspects are primarily those which are included in our analyses. Thus, the classification instructions were the result of a purely informal empirical exercise without strong *a priori* theoretical reasons for a typology. As Connell & Goldsmith (1982) noted 'unless typologies are derived by appropriate empirical means (e.g. cluster analytic techniques), they are unlikely to exhibit the same predictive capacity and internal structure in subsequent applications' (p. 219). Several researchers question whether individual behavioural differences in the Strange Situation are adequately represented by the discrete categorization, and they consequently recommend the consideration of continuous measures (Connell & Goldsmith, 1982; Kroonenberg & Van IJzendoorn, 1987; Lamb *et al.*, 1985).

Irrespective of the discreteness of the attachment construct, cluster methods will produce clusters, either by dissecting continuous dimensions or by seeking for natural clusters. Only afterwards, can one discern the nature of the clusters, for instance by using ordination methods. Results of the cluster analyses therefore will not provide a definite answer to the (still unresolved) question whether attachment is a discrete or a continuous construct. However, if the groupings derived by the cluster method correspond to the clinical classifications, this will at least give further support to the A–B–C typology. If not, we will at least gain further insight into the individual differences in the Strange Situation.

Results

Overview of analyses

Even though from a substantive view our main aim is to compare a statistical classification with the clinical classification, we will present several more in-depth analyses to show how we arrived at the final statistical classification. This is done with an eye to our secondary objective, i.e. demonstrating how mixture methods of clustering work. First, we will illustrate the kind of decisions that have to be considered in using the mixture method of clustering, such as the optimal number of clusters and whether arbitrary or common covariance matrices should be specified. The measurements of the first reunion episode will be used for this purpose. Secondly, for each of the episodes the infants will be clustered on the basis of the five variables using the two-way mixture method of clustering with arbitrary covariance matrices. This will provide both an insight into the information content of the episodes and starting values for the three-way clustering. Thirdly, the three-way mixture method of clustering will be applied to the complete $326 \times 2 \times 5$ data set with arbitrary covariance matrices. Fourthly, the evaluation of the statistically derived optimal classification from the three-way clustering procedure with respect to the original clinical classification based on the Ainsworth coding scheme will be addressed. One may cross-classify the different groupings, assess the percentage agreement between the partitionings, and compute the concordance with measures like Cohen's kappa (Cohen, 1960). Such a procedure, however, needs to be supplemented by information on how the clustering technique partitioned the subjects. To do this, we will depict the groupings in the space defined by the first two components derived from the replicated principal component analysis. Some

additional information about the basis for the clustering can be gleaned from two-way discriminant analyses based on all 10 variables.

Analysis decisions

Number of clusters. To demonstrate the considerations that go into choosing an optimal number of clusters, consider the first reunion episode under the common covariance matrix model. One can take some guidance from values of the likelihood ratio test for comparing a cluster solution with g groups with the solution g-1 groups (see McLachlan & Basford, 1988, p. 23; Wolfe, 1971). Although the log likelihood increases monotonically with the number of clusters (Table 1), much smaller gains are made with the addition of more than three clusters (the $-2\ln\lambda$ values were 66, 102, and 42 for $g = 5$ to 7, respectively). In this subsection, information on five to seven groups is discussed, although not presented.

Table 1. Results of two-way mixture method of clustering first reunion episode—R1 (common covariance matrix model)

No. of groups	Log likelihood of best solution	Log likelihoods with hierarchical starts					$-2\ln\lambda$
		H1	H2	H3	H4	H5	
1	−2987						
2	−2831	−2831	−2831	−2831	−2831	−2831	312
3	−2743	−2769	−2817	−2795	**−2743**	**−2743**	176
4	−2685	−2745	−2808	−2708	**−2685**	**−2685**	116

Notes. Lambda is the ratio for g groups and $g-1$ groups. The degrees of freedom for the approximate chi-squared test is twice the difference in number of parameters in the two models. This is 2*5 for the additional mean vector for the common covariance matrix model and would be 2*(5 means $+\frac{1}{2}$*5*6) for the additional mean vector and covariance matrix for the arbitrary covariance model. **Bold** type indicates the best solutions. H1 to H5 are the hierarchical clustering techniques using group average, median, centroid, flexible sorting, and incremental sum of squares, respectively, as the classification strategy.

Basford & McLachlan (1985*a*; McLachlan & Basford, 1988, chapter 5) also look at the *overall correct allocation rate and the correct allocation rate for each cluster*, where allocation rates are defined as weighted sums of the *a posteriori* allocation probabilities of entities (here infants) to clusters. In the present case they do not seem to be very informative with respect to the number of clusters, as the overall values are already very high for each solution from two to seven clusters (.998, .979, .974, .964, .929, and .987, respectively). A further possibility might be to look at the average absolute within-group correlations to evaluate how well we have succeeded in creating homogeneous groups. These values for one to seven clusters are .41, .19, .21, .14, .14, .13, and .08, respectively. It seems that the big gains in diminishing the heterogeneity are obtained up to four groups, but little is gained afterwards.

Further information on the stability of the division of the infants into groups can be gained by cross-tabulation of the various partitionings. To this end, we show the cross-tabulation of the three-cluster solution against the other solutions (Table 2). A

Table 2. Cross-tabulations of cluster solutions against the three-cluster solution first reunion episode (common covariance matrix model)

Group	2		3			4			
	1	2	1	*2*	3	1	*2*	*3*	4
1	**140**	0	**140**	0	0	**139**	1	0	0
2	*109*	0	0	*109*	0	0	*84*	*25*	0
3	0	**77**	0	0	**77**	0	1	**6**	**70**
T	*259*	**77**	**140**	*109*	**77**	**139**	**85**	*31*	**70**

Note. T = Totals.

fair amount of nesting of solutions occurs and this continues even when a larger number of clusters is examined. Apparently there exist fairly definite and stable boundaries between the groups. This illustrates one of the strengths of the mixture cluster method, i.e. with increasing (decreasing) numbers of groups, solutions are not necessarily a hierarchy, and it is an empirical, rather than a method-dependent, issue whether nesting takes place.

Assembling all the information on the various cluster solutions and using the stability argument presented below, it seems that either a two-group or three-group solution (given common covariance matrices) is optimal for the first reunion episode. As there are three groups in the clinical typology, it was decided to restrict subsequent cluster analyses to three groups.

Arbitrary versus common covariance matrices. As explained above, the mixture method is an iterative procedure which uses maximum likelihood estimation. Because such a procedure is only assured to converge to a local maximum, one has to use several different starting allocations to (hopefully) find the global maximum. In the present case, these were obtained by using the grouping at the appropriate level from each of several different hierachical clustering methods from the statistical package SAS CLUSTER (SAS Institute, 1985), in particular, group average (H1), median (H2), centroid clustering (H3), flexible sorting with beta equal to −0.25 (H4), and Ward's method (Ward, 1963) or incremental sums of squares (H5). The starting allocations for the division of infants into groups presented in Table 1 were obtained without standardizing the variables.

For all solutions with more than two clusters, the likelihood function has indeed multiple maxima under the common covariance model for clustering the data from the first reunion episode. There is no guarantee that the best maximum is also the global maximum, be it that in general more confidence is inspired by a solution when

the iterative procedure produces the same solution from different starting allocations. A starting allocation obtained via Ward's method very often leads to a solution with the highest log likelihood. This is not completely surprising as there exist close links between Ward's method and the mixture method of clustering (see e.g. Gordon, 1981, p. 50). It seems that for the common covariance case only the two-cluster solution has good stability (see Table 1).

For this data set, the three-cluster solution with arbitrary covariance matrices is very stable as all starts converged to the same solution. Even when using the clinical classification as a starting allocation, the algorithm converged to the same maximum. This indicates that the clinical classification is suboptimal for the mixture modelling of this data.

At the three-cluster level (as at other levels), the log likelihood of the arbitrary covariance solution by far outstrips that of the common covariance one (6517 versus −2743), in other words the assumption of a common covariance matrix is not appropriate here. Both models produced a solution with generally unambiguous allocations of infants to groups although the partitions were quite different (Table 3).

Table 3. Cross-tabulations of three-cluster solutions

R1 Arbitrary covariances	R1 Common covariances			R2 Arbitrary covariances			
Group	1	2	3	1	2	3	T
1	**111**	74	0	**112**	36	37	185
2	29	**10**	6	10	**23**	12	45
3	0	25	**71**	2	6	**88**	96
T	140	109	77	124	65	137	326

Note. T = Totals.

Given the more consistent convergence of the arbitrary covariance matrix model to a particular solution and the above information on more definite allocation into groups, it was decided to continue with the arbitrary covariance model for all other analyses.

Separate analyses for the reunion episodes (arbitrary covariance matrices)

Given that the same number of infants and the same variables were available, the smaller size of the log likelihood for the second reunion episode (5864 versus 6517) could suggest that the cluster structure might be less clear for this episode. The comparison between the clusterings of the first and second reunion episodes (Table 3) revealed, to our surprise, that the clusterings hardly agree. The percentage agreement is .68 while the chance corrected agreement measured by Cohen's kappa (Cohen, 1960) is .50, with .60 considered to be an absolute minimum for reasonable agreement. While the different distributions of the infants over groups suggest that

the clusters are rather different, superficial inspection of the means for each group indicates that the groups are not unalike (Table 4). We have used the two mixture solutions from the separate episodes as additional starting allocations for the three-way clustering to see whether they give rise to different solutions in the combined analysis.

Table 4. Estimated means for three-cluster solutions for first and second reunion episodes (arbitrary covariance matrix model)

Group	PS	CM	RS	AV	DI	N
First reunion episode						
1	2.2	1.0	1.7	**3.7**	**4.9**	185
2	**4.0**	2.8	1.8	2.8	**5.3**	45
3	**5.6**	**4.8**	2.8	1.7	1.0	96
Second reunion episode						
1	1.9	1.0	1.9	**3.6**	**5.3**	124
2	**4.2**	**3.6**	2.4	2.8	**4.5**	65
3	**5.5**	**5.4**	3.1	1.8	1.0	137

Note. All means greater than 3.5 are indicated in **bold**. PS = proximity seeking; CM = contact maintaining; RS = resistance; AV = avoidance; DI = distance interaction.

First and second reunion episodes jointly

For the single episode analyses we had no obvious starting values for group membership, so had to make do with those from several hierarchical clustering procedures. For the combined analysis, we can supplement the starting allocations obtained from hierarchical clustering procedures with the groupings obtained as solutions from the mixture analysis of the separate episodes. As it turned out, *all* starting allocations (for which the program converged to a solution) converged to the same three-way cluster solution with arbitrary covariance matrices. Given the log likelihood value for one group was -6041, the best two-cluster solution had a log

Table 5. Cross-tabulations of cluster solutions against joint solution for first + second reunion episodes (arbitrary covariance matrices model)

	R1			R2			
Group	1	2	3	1	2	3	T
1	**114**	0	0	**112**	0	2	114
2	69	**45**	8	12	**65**	45	122
3	2	0	**88**	0	0	**90**	90
T	185	45	96	124	65	137	326

Note. T = Totals.

likelihood of 336, while the log likelihood for the three-cluster solution was 5811. Solutions with a larger number of clusters only produced marginal increases in the likelihood. In comparison, the three-cluster solution for the common covariance matrix model had a likelihood of -5733. Thus, the three-cluster solution with arbitrary covariance matrices is clearly the preferred solution. This solution was also satisfactory from the point of view of *correct allocation rates* as these were all equal to 1.00, indicating that all infants were allocated to groups with an *a posteriori* probability of 1.00.

The cross-tabulation of the R1 + R2 three-cluster solution with that of the first and second reunion episodes (Table 5) gave a percentage agreement and Cohen's kappa of 76 per cent and $\kappa = .64$ with R1, and 82 per cent and $\kappa = .73$ with R2. This mediocre level of agreement might be disconcerting for natural clusters, but not necessarily so when the clusters are the result of dissections of continuous dimensions. In that case, it is easy to imagine that even small shifts in the variables could lead to different optimal solutions for the clustering algorithm. As the three-way solution had virtually perfect allocation of the infants to clusters, we are prepared to put more trust in that solution than in each of the separate ones.

The correlation matrices for the R1 and R2 episodes and the three-group mixture solution are listed (Table 6) to provide an indication of the relationships between

Table 6. Correlation matrices

	PS	CM	RS	AV	DI	PS	CM	RS	AV	DI
Original correlation matrices										
			Reunion Episode 1					*Reunion Episode 2*		
Proximity (PS)	1.00					1.00				
Contact (CM)	.71	1.00				.71	1.00			
Resistance (RS)	.27	.41	1.00			.28	.35	1.00		
Avoidance (AV)	−.53	−.49	−.07	1.00		−.59	−.48	.01	1.00	
Distance (DI)	−.49	−.62	−.32	.19	1.00	−.60	−.69	−.38	.31	1.00
Correlations between the same variables across										
episodes						.52	.67	.40	.35	.71
Three-group solution for R1 + R2 arbitrary covariance matrices model										
			Group 1					*Group 2*		
Proximity	1.00					1.00				
Contact	.00	.00				.51	1.00			
Resistance	−.08	.00	1.00			.10	.13	1.00		
Avoidance	−.23	.00	.33	1.00		−.41	−.25	.11	1.00	
Distance	.21	.00	−.11	−.45	1.00	−.28	−.44	−.28	.01	1.00
			Group 3							
Proximity	1.00									
Contact	.17	1.00								
Resistance	.03	.27	1.00							
Avoidance	−.37	−.17	.10	1.00						
Distance	.00	.00	.00	.00	.00					

variables in the different episodes and clusters. In both the first and the third group, all infants had the same score of 1.00 (the lowest possible) for one of the variables (contact maintaining and distance interaction, respectively).

The mean differences of proximity seeking (PS), contact maintaining (CM), and distance interaction (DI) contribute most to the distinction between groups (Table 7). Resistance (RS) and avoidance (AV) are less important, be it that they follow the pattern of PS and CM, and that of DI, respectively. These means (Table 7) show that the *first* cluster is characterized by high values in both episodes for avoidance and distance interaction, and low to no proximity, resistance, and contact maintaining. The *second* cluster shows a stable low RS, but increasing PS and CM coupled with decreasing AV and DI. Finally, the *third* cluster has consistently high PS and CM scores with low to no AV and DI coupled with a comparatively high level of RS. Overall, there seems to be a single (proximity + contact) versus (avoidance + distance) dimension which the cluster method uses to define groups.

Table 7. Estimated means for three-cluster solution for first + second reunion episodes (arbitrary covariance matrices model)

Group		PS	CM	RS	AV	DI	N
1	R1	2.1	1.0	1.5	**3.6**	**5.1**	114
	R2	1.9	1.0	2.0	**3.6**	**5.2**	
2	R1	3.0	1.8	2.0	3.3	**4.7**	122
	R2	**4.4**	**3.9**	2.5	2.6	3.3	
3	R1	**5.5**	**4.8**	2.8	1.8	1.0	90
	R2	**5.7**	**5.7**	3.3	1.7	1.0	

Note. PS = proximity seeking; CM = contact maintaining; RS = resistance; AV = avoidance; DI = distance interaction.

To gain another perspective on the differences between the infants a replicated principal component analysis was performed on the two times five variables. Given that we intended to use this analysis as a support for the cluster analysis, it was decided to limit the solution to a fairly easily presentable one with two components. In this way, we could easily provide a graphical display of the results. The two-component solution accounted for 63 per cent of the variability with a first component accounting for 47 per cent and the second for 15 per cent. The relative weights of the two reunion episodes were .66 and .75, indicating that the second episode carried a bit more weight in the solution. A discriminant analysis with the groups from the three-way cluster analysis as the dependent variable and the first component as predictor yielded a canonical correlation coefficient of .89. The first component provides for a 83 per cent correct allocation of the infants to the groups found by the cluster analysis. The loadings for the five variables on the first component are .84 (CM), 0.77 (PS), 0.47 (RS), −0.55 (AV), and −0.74 (DI). This confirms the above statement about the importance of the PS and CM versus AV and
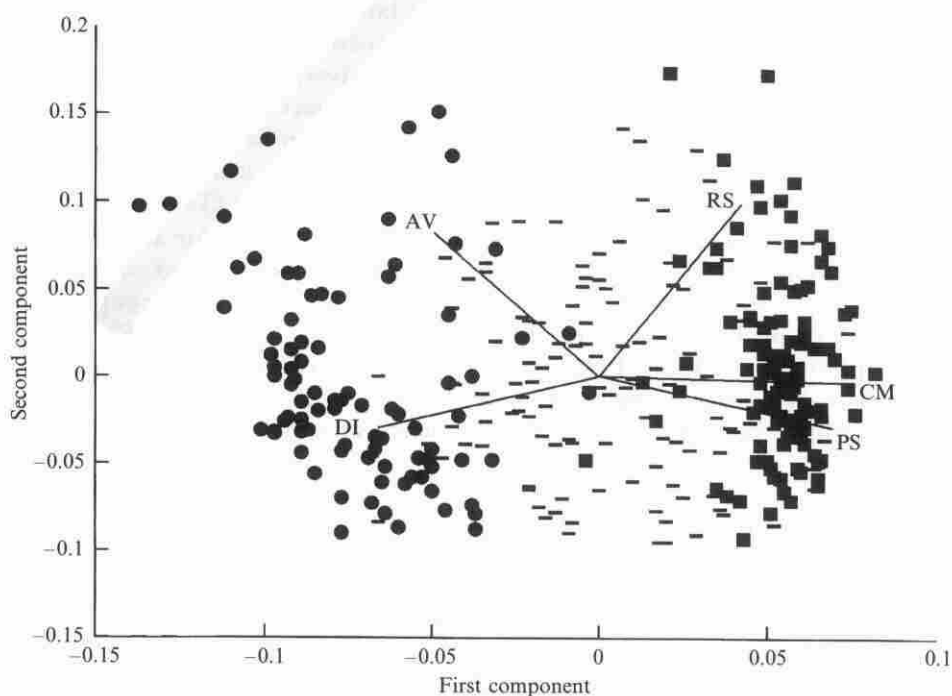
Figure 1. Joint representation of the infants and variables on the components of the replicated component analysis. Infants are labelled according to their statistically derived cluster membership from the three-way mixture method of clustering with the arbitrary covariances model (● = cluster 1; — = cluster 2; ■ = cluster 3; AV = avoidance; CM = contact maintaining; DI = distance inter-action; PS = proximity seeking; RS = resistance).

DI contrast in the cluster analysis. In Fig. 1 the two components are presented, and each of the infants is labelled according to its cluster membership. In the figure, we have also drawn the vectors of component loadings for the five variables.

Comparison with clinical classification

As mentioned in the introduction, the main object of the paper is to evaluate clinical classification which was constructed using the guidelines set out by Ainsworth *et al.* (1978). To do so, we have crossed the clinical classification with the clusters from the three-way mixture method for the case of arbitrary covariance matrices (Table 8). It is apparent that the clustering method and the clinical classification procedures work differently as they create entirely different clusters. Whichever way one wants to arrange the table, the highest percentage agreement one can get is 51 per cent, a very low figure indeed. Neither the cluster sizes, nor the allocation of individuals to clusters show much agreement. From the point of view of the cluster analysis, the clinical classification is clearly suboptimal. Even with the clinical classification as a starting allocation for the clustering, the algorithm will converge to the solution already presented.

Table 8. Clinical classification versus cluster classification first and second reunion episodes combined (arbitrary covariance matrices model)

Group	1	2	3	T
A	**58**	19	3	80
B	66	**41**	102	209
C	0	5	**32**	37
T	124	65	137	326

Note. T = Totals.

To evaluate why the clusterings are so different, one may look at the means for the A, B, and C groups (Table 9). The means of avoidance and resistance are notable when compared with their role in the cluster classification (Table 7) where they were less disparate. Further insight can be gained by performing discriminant analyses on all 10 variables with the clinical classification and the mixture classification as dependent variables, respectively. The concordance between the mixture cluster analysis and the discriminant analysis need not be exact. The cluster method treats the data as multivariate measurements on the same infants at two separate (independent) times, whereas the discriminant analysis assumes (a larger set of) multivariate observations on the infants (at one time) only. A further point is that discriminant analysis assumes equal covariance matrices in the three groups, while we have shown that the groups found by the cluster analysis have quite different covariance matrices. In principle, one would need a quadratic discriminant analysis to do the arbitrary covariance matrices solution justice. However, in the present case this option is not available as the covariance matrices of both group 1 and 3 are singular, which prohibits such an analysis.

Table 9. Estimated means for clinical classification

Group		PS	CM	RS	AV	DI	N
A	R1	2.3	1.3	1.9	**4.3**	**4.4**	80
	R2	2.2	1.7	2.6	**4.7**	**4.2**	
B	R1	**3.6**	2.6	1.8	2.5	**3.8**	209
	R2	**4.3**	**3.7**	2.1	2.0	3.3	
C	R1	**4.5**	**3.6**	**4.0**	2.9	2.6	37
	R2	**5.4**	**5.3**	**4.9**	2.1	1.4	

Note. PS = proximity seeking; CM = contact maintaining; RS = resistance; AV = avoidance; DI = distance interaction.

From the standardized coefficients in the discriminant analyses (Table 10), it is evident that the two groupings weigh the variables quite differently. The clinical classification rests virtually exclusively on avoidance and resistance, and the additional analyses show that avoidance and resistance in the second reunion episode

**Table 10.** Discriminant analyses results (standardized discriminant coefficients)

	No.	1st reunion episode					2nd reunion episode					R_c	%
		AV	RS	CM	PS	DI	AV	RS	CM	PS	DI		
Clinical classification													
All Vars	1	**.4**	.1	.0	.2	.0	**.8**	.2	−.2	−.3	−.2	.78	83.7%
A+R/R12	1	.3	.1				**.9**	.2				.77	82.2%
A+R/R2	1						**1.0**	.2				.76	82.5%
A+R/R1	2	**.9**	−**.4**									.47	66.8%
All Vars	2	.1	**.5**	−.2	.2	.2	−.2	**.7**	.1	.0	−.1	.63	
A+R/R12	2	.1	**.5**				−.2	**.7**				.62	
A+R/R2	2						−.1	**1.0**				.58	
A+R/R1	1	**.5**	**.9**									.52	
Mixture method clustering (arbitrary covariance matrices model)													
All	1	−.3	.0	.3	.2	−**.5**	−.1	.1	.3	.3	−.2	.91	88.0%
All	2	.1	.1	−.3	−.1	**.6**	.2	−.1	**.6**	**.5**	.0	.57	

Notes. R_c = canonical correlation; % = percentage correct classification by both discriminant functions jointly; A+R = avoidance and resistance only; R12 = R1+R2; No. = number of discriminant function; PS = proximity seeking; CM = contact maintaining; RS = resistance; AV = avoidance; DI = distance interaction. Values greater than or equal to .4 in absolute value are shown in **bold face**.

are able to discriminate as well as all variables together, 82.5 and 83.7 per cent, respectively. The clustering solution, however, is based on the information in all variables with the least weighting on resistance. Note that while the clustering method indicates all subjects have *a posteriori* probabilities of 1.00 of belonging to a particular cluster, the discriminant analysis casts doubt about the proper allocation of 12 per cent, i.e. 39 infants. This discrepancy is probably due to the different assumptions about the covariance matrices, as mentioned above.

A final point about the clinical classification can be best illustrated by presenting again the first two components of the replicated component analysis, but now labelling the infants in the plot with their clinical classifications. Again the vectors of the five variables are displayed as well. In the section on the substantive background we mentioned the problem of continuity versus discreteness of the attachment construct, which corresponds to the question of natural clusters or dissecting continuous dimensions. Inspecting both Fig. 1 and Fig. 2, we see that in neither case is it easy to maintain that the partitions correspond to natural clusters. Both the partitioning by the cluster method and that by the clinical classification appear to dissect the continuous dimensions.
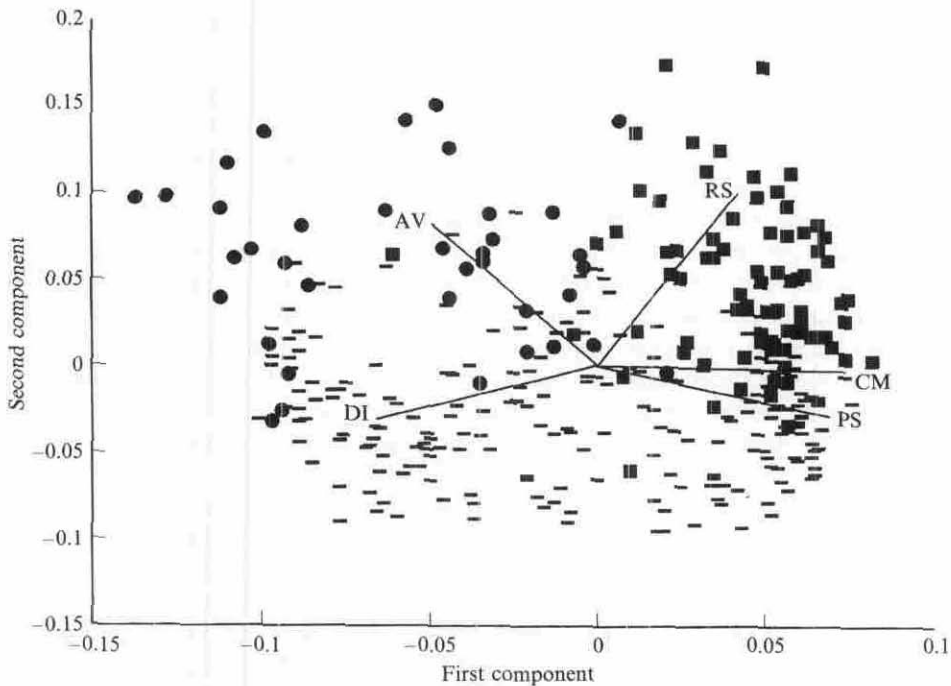
Figure 2. Joint representation of the infants and variables on the components of the replicated component analysis. Infants are labelled according to their clinical classifications ($\bullet$ = A; — = B; ■ = C; AV = avoidance; CM = contact maintaining; DI = distance interaction; PS = proximity seeking; RS = resistance).

Conclusions

Given the results presented, we have to conclude that three-way clustering does not correspond to the A–B–C typology (see also Lamb *et al.*, 1985). The three-way clustering methods and clinical classification procedures create entirely different groups of infants. In the course of analysis, these discrepancies were considered from different perspectives to gain further insight into the individual differences in the Strange Situation. From the estimated means for the cluster solutions, it could be derived that the role of avoidance and resistance is much more important to the clinical classifications than to the clusters. The same conclusion can be drawn from the results of discriminant analyses with the clinical classifications and the clusters as dependent variables. The two groupings weighted the variables in a different way; whereas the clinical groupings rest almost exclusively on avoidance and resistance, the cluster groupings are based on the information of all variables, except resistance.

From the results of a discriminant analysis, which predicted the clusters using the first principal component of a replicated principal component analysis, one single dimension appeared to underlie the clustering. This dimension could be interpreted as the extent to which infants primarily seek proximity and contact with the mother (i.e. use PS and CM), or primarily stay at a distance from her (i.e. use AV and DI). This points towards different styles of behaviour of the infants, largely independent of their attachment classifications.

The differences between the two classifications point to different underlying assumptions. Clearly, there are theoretical substantive arguments why resistance and avoidance play such a dominant role and carry so much weight in the clinical classification. In the clustering method, it is the differences in the sizes of variances which determine for a large part the outcome of the analysis, and especially resistance is a variable with one of the smaller variances. One way of looking at these results is that there are more and larger differences in the strange situation than are captured by the clinical classification. On the other hand, they are apparently not the ones which are deemed the most important in the theory of attachment. Earlier, we indicated that attachment research has shown the validity of the clinical classification by relating it to several preceding and subsequent behaviours. Whether this can also be said for the differences highlighted by the clustering procedure is a matter to be investigated.

With respect to the continuity–discreteness argument, the analyses lend some support to the statement by Lamb *et al.* (1985), that the A, B, and C do not represent distinct types of infants, but that they have arisen from an underlying continuum which has been artificially trichotomized. Of course, a similar statement can be made about the statistically derived grouping. These results do not necessarily imply that a natural trichotomy does not exist, they only indicate that such a division is not strongly supported by the present empirical investigation. One should either call upon theoretical arguments or additional empirical information to substantiate the natural clusters claim.

The mixture method of clustering has allowed an effective exploratory analysis of the data which enables some aspects of the three-way structure to be incorporated. The assumption of independence of the episode measurements in the mixture model

does not appear to be too restrictive, given the consistency of the results from different analytical techniques. Thus, for the present data, we have reasonable confidence in the summarization in terms of relatively homogeneous clusters.

Acknowledgements

References

Ainsworth, M. D. S. (1990). Some considerations regarding theory and assessment relevant to attachments beyond infancy. In M. T. Greenberg, D. Cichetti & E. M. Cummings (Eds), *Attachment in the Preschool Years: Theory, Research, and Intervention*, pp. 463–488. Chicago, IL: University of Chicago Press.

Ainsworth, M. D. S., Blehar, M. C., Waters, E. & Wall, S. (1978). *Patterns of Attachment. A Psychological Study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistics*, 2nd ed. New York: Wiley.

Arabie, P., Carroll, J. D. & DeSarbo, W. S. (1987). *Three-way Scaling and Clustering*. Beverly Hills, CA: Sage.

Basford, K. E., Kroonenberg, P. M. & DeLacy, I. H. (1991). Three-way methods for multiattribute genotype × environment data: An illustrated partial survey. *Field Crops Research*, **27**, 131–157.

Basford, K. E. & McLachlan, G. J. (1985a). Estimation of allocation rates in a cluster context. *Journal of the American Statistical Association*, **80**, 286–293.

Basford, K. E. & McLachlan, G. J. (1985b). The mixture method of clustering applied to three-way data. *Journal of Classification*, **2**, 109–125.

Basford, K. E. & McLachlan, G. J. (1985c). Likelihood estimation with normal mixture methods. *Applied Statistics*, **34**, 282–289.

Carroll, J. D. & Arabie, P. (1983). INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika*, **48**, 157–169. (Reprinted in H. G. Law, C. W. Snyder Jr, J. A. Hattie & R. P. McDonald, Eds, *Research Methods for Multimode Data Analysis*, pp. 518–534. New York: Praeger.)

Carroll, J. D., Clark, L. A. & DeSarbo, W. S. (1984). The representation of three-way proximity data by single and multiple tree structure models. *Journal of Classification*, **1**, 25–74.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

Connell, D. B. (1977). Individual differences in attachment behaviour: Long-term stability and relationships to language development. Unpublished doctoral thesis. Syracuse, NY: Syracuse University.

Connell, J. P. & Goldsmith, H. H. (1982). A structural modeling approach to the study of attachment and Strange Situation behaviors. In R. N. Emde & R. J. Harmon (Eds), *The Development of Attachment and Affiliative Systems*, pp. 213–243. New York: Plenum.

De Soete, G. & Carroll, J. D. (1989). Ultrametric tree representations of three-way three-mode data. In R. Coppi & S. Bolasco (Eds), *Multiway Data Analysis*, pp. 415–426. Amsterdam: North Holland.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Duda, R. & Hart, P. (1973). *Pattern Classification and Scene Analysis*. New York: Wiley.

Everitt, B. S. (1980). *Cluster Analysis*, 2nd ed. London: Heinemann.

Everitt, B. S. (1981). A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivariate Behavioral Research*, **16**, 171–180.

Gordon, A. D. (1981). *Classification*. London: Chapman & Hall.

Goossens, F. A. (1986). *The Quality of Attachment Relationships of Two-year-olds of Working and Non-working Mothers and Some Associated Factors*. Leuven: Acco.

Goossens, F. A. & Van IJzendoorn, M. H. (1990). Quality of infants' attachments to professional caregivers: Relation to infant–parent attachment and day-care characteristics. *Child Development*, **61**, 832–837.

Harshman, R. A. & Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder Jr, J. A. Hattie & R. P. McDonald (Eds), *Research Methods for Multimode Data Analysis*, pp. 122–284. New York: Praeger.

Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Annals of Statistics*, **13**, 795–800.

Hubbard, F. O. A. & Van IJzendoorn, M. H. (1991). Maternal unresponsiveness and infant crying across the first nine months: A naturalistic longitudinal study. *Infant Behavior and Development*, **14**, 299–312.

Kiefer, J. & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimates in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, **27**, 887–906.

Kiefer, N. M. (1978). Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica*, **46**, 427–434.

Kroonenberg, P. M. (1983). *Three-mode Principal Component Analysis: Theory and Applications*. Leiden, The Netherlands: DSWO Press.

Kroonenberg, P. M. (1994). The TUCKALS line: A suite of programs for three-way analysis. *Computational Statistics and Data Analysis*, **18**, 73–96.

Kroonenberg, P. M. & Brouwer, P. (1993). *3WAYPACK User's Manual: A Package of Three-way Programs*, vol. 3: TUCKALS3 Manual. Department of Education, Leiden University.

Kroonenberg, P. M. & Van IJzendoorn, M. H. (1987). Exploring children's behavior in the Strange Situation. In L. W. C. Tavecchio & M. H. van IJzendoorn (Eds), *Attachment in Social Networks: Contributions to the Bowlby–Ainsworth Attachment Theory*, pp. 379–425. Amsterdam: North Holland.

Lamb, M. E., Thompson, R. A., Gardner, W. & Charnov, E. L. (1985). *Infant–Mother Attachment: The Origins and Developmental Significance of Individual Differences in Strange Situation Behavior*. Hillsdale, NJ: Erlbaum.

Lambermon, M. W. E. & Van IJzendoorn, M. H. (1989). Influencing mother–infant interaction through videotaped or written instruction: Evaluation of a parent education program. *Early Research Childhood Quarterly*, **4**, 449–458.

Main, M. & Solomon, J. (1990). Procedures for identifying infants as disorganised/disoriented during the Ainsworth Strange Situation. In M. T. Greenberg, D. Cicetti & E. M. Cummings (Eds), *Attachment in the Preschool Years: Theory, Research, and Intervention*. Chicago: University of Chicago Press.

McLachlan, G. J. & Basford, K. E. (1988). *Mixture Methods: Inference and Applications to Clustering*. New York: Marcel Dekker.

Richters, J. E., Waters, E. & Vaughn, B. E. (1988). Empirical classification of infant–mother relationships from interactive behavior and crying during reunion. *Child Development*, **59**, 512–522.

SAS Institute (1985). *SAS User's Guide: Statistics, Version 5 Edition*. Cary, NC: SAS Institute.

Sawyer, J. (1966). Measurement *and* prediction, clinical *and* statistical. *Psychological Bulletin*, **66**, 178–200.

Ten Berge, J. M. F., De Leeuw, J. & Kroonenberg, P. M. (1987). Some new results on principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, **52**, 183–191.

Tucker, L. R. & Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika*, **28**, 333–367.

Van Dam, M. (1993). Secundaire analyse van de Strange Situation (Secondary analyses of the Strange Situation). Doctoral thesis. Leiden, The Netherlands: Department of Education, Leiden University.

Van Dam, M. & Van IJzendoorn, M. H. (1988). Measuring attachment security: Concurrent and predictive validity of the parental attachment Q-sort. *Journal of Genetic Psychology*, **149**, 447–458.

Van IJzendoorn, M. H., Goossens, F. A., Kroonenberg, P. M. & Tavecchio, L. W. C. (1985). Dependent attachment: B4-children in the Strange Situation. *Psychological Reports*, **57**, 439–451.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**, 236–244.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, **5**, 329–350.

Wolfe, J. H. (1971). A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. Technical Report STB 72-2. San Diego: US Naval Personnel and Training Research Laboratory.

Wolfe, J. H. (1978). Comparative cluster analysis of patterns of vocational interest. *Multivariate Behavioral Research*, **13**, 33–44.

Appendix

Three-way mixture method of clustering

The fitting of a normal mixture model to three-way data is presented in the context of clustering entities on the basis of multivariate observations being recorded on each of these entities (here infants) at separate episodes or locations. Then the observation vector x_j $(j = 1, \ldots, n)$ contains the multivariate responses of the jth infant during both episodes, and is given by equation (1)

$$x_j = (x'_{j1}, \ldots, x'_{jr})', \tag{1}$$

where x_{jk} is a vector of length p giving the response of infant j in episode k for each of the same p attributes measured during each episode $(k = 1, \ldots, r)$. The vectors x_{jk} $(j = 1, \ldots, n; k = 1, \ldots, r)$ are taken to be independently distributed.

This method is model based, in that the form of the density of an observation in each of the underlying populations has to be specified. A common approach is to take the component densities to be multivariate normal distributions with (in principle) different mean vectors and either equal or arbitrary covariance matrices. Under the normal mixture model proposed by Basford & McLachlan (1985b) for such three-way data, it is assumed that each infant belongs to one of g possible groups $G_1, \ldots, G_g$, in proportion $\pi_1, \ldots, \pi_g$ respectively, so that during a given episode k,

$$x_{jk} \sim N(\mu_{ik}, V_i) \text{ in } G_i \text{ with probability } \pi_i \quad (i = 1, \ldots, g). \tag{2}$$

The within-group covariance matrix V_i is taken not to depend on the episode. From equation (2), the density of the full observation vector x_j conditional on the jth individual belonging to G_i is equal to

$$f_i(x_j; \theta) = (2\pi)^{-\frac{rp}{2}} |V_i|^{-\frac{r}{2}} \exp\left[-\frac{1}{2} \sum_{k=1}^{r} (x_{jk} - \mu_{ik})' V_i^{-1}(x_{jk} - \mu_{ik}) \right] \tag{3}$$

for $i = 1, \ldots, g$, where the vector θ of unknown parameters contains the elements of μ_{ik} $(i = 1, \ldots, g; k = 1, \ldots, r)$ and the distinct elements of V_i $(i = 1, \ldots, g)$.

Estimates of the parameters in this model can be obtained using the likelihood principle, whereby the log likelihood of (π, θ) given by

$$\ln L(\pi, \theta) = \sum_{j=1}^{n} \ln \left[\sum_{i=1}^{g} \pi_i f_i(x_j, \theta) \right] \tag{4}$$

is maximized.

Once the likelihood solution of (π, θ) has been obtained, estimates of the posterior probabilities of population membership can be formed for each x_j (really the entity, here infant, with observation x_j) to give a probabilistic clustering. The posterior probability that x_j belongs to G_i is given by

$$\tau_{ij} = \pi_i f_i(x_j, \theta) \bigg/ \sum_{u=1}^{g} \pi_u f_u(x_j, \theta) \tag{5}$$

Each x_j is then assigned to the population or group to which it has the highest estimated posterior probability of belonging, i.e. to G_u if

$$\hat{\tau}_{uj} > \hat{\tau}_{ij} \quad (i = 1, \ldots, g; i \neq u). \tag{6}$$

In this way, individual observations (here the individual infants) are partitioned into a number of discrete, relatively homogeneous groups. If (π, θ) were known, the allocation rule based on equation (5) would be the optimal or Bayes rule (Anderson, 1984, chapter 6) which minimizes the overall error rate. For the above model, the likelihood estimate of (π, θ) satisfies

$$\hat{\pi}_i = \frac{1}{n} \sum_{j=1}^{n} \hat{\tau}_{ij} \quad (i = 1, \ldots, g) \tag{7}$$

$$\hat{\mu}_{ik} = \sum_{j=1}^{n} \hat{\tau}_{ij} x_{jk} \bigg/ \sum_{j=1}^{n} \hat{\tau}_{ij} \quad (i = 1, \ldots, g) \tag{8}$$

and

$$\hat{V}_i = \sum_{j=1}^{n} \sum_{k=1}^{r} \hat{\tau}_{ij} (x_{jk} - \hat{\mu}_{ik})' (x_{jk} - \hat{\mu}_{ik}) / r \sum_{j=1}^{n} \hat{\tau}_{ij} \quad (i = 1, \ldots, g). \tag{9}$$

The posterior probability that the jth entity belongs to G_i, given the observation x_j on it, is given by

$$\tau_{ij} = \frac{\pi_i |V_i|^{-\frac{r}{2}} \exp \left[-\frac{1}{2} \sum_{k=1}^{r} (x_{jk} - \mu_{ik})' V_i^{-1} (x_{jk} - \mu_{ik}) \right]}{\sum_{u=1}^{g} \pi_u |V_u|^{-\frac{r}{2}} \exp \left[-\frac{1}{2} \sum_{k=1}^{r} (x_{jk} - \mu_{uk})' V_u^{-1} (x_{jk} - \mu_{uk}) \right]}. \tag{10}$$

Equations (7) to (10) are solved iteratively by using some initial estimates of the unknown parameter vector (π, θ). These iterative estimates can be identified with those obtained by directly applying the EM algorithm of Dempster, Laird & Rubin (1977) to solve the likelihood equation. Then provided the likelihood is bounded above, convergence to some local maximum is assured.

Unfortunately with mixture models, the likelihood equation has multiple roots, so there is the problem of which root to choose. This problem is discussed in detail in Basford & McLachlan (1985c) and McLachlan & Basford (1988). With equal covariance matrices for the groups and the normality assumption, the maximum likelihood estimator of (π, θ) does exist and is strongly consistent. Therefore the largest of the local maxima (assuming all have been located) is the one to choose. In the arbitrary covariance matrix case, the likelihood is unbounded and so the maximum likelihood estimator of (π, θ) does not exist (Kiefer & Wolfowitz, 1956). Kiefer (1978), however, verified for univariate data that there is a sequence of roots of the likelihood equation which is consistent, asymptotically normal and efficient. With probability tending to one, these roots correspond to local maxima of the likelihood. McLachlan & Basford (1988) postulated that it would be surprising if the univariate result, that it is the sequence of roots corresponding to the largest of the local maxima which is consistent (see Hathaway, 1985), does not carry over to mixtures of multivariate normal distributions.