NOTES

1. This requirement can be relaxed, but if overlapping categories are allowed, the partition cannot then be used to represent the sorting, and the resulting mathematics and the statistics become more complicated.

2. See Jacoby (1991) in this series for a useful comparative overview of theories of data.

3. Struhsaker, 1967, p. 110; cited in Arabie and Boorman (1973), who also reproduce the basic data (pp. 180–182).

4. Because each artifact/object will occur in several graves/categories, this example strictly speaking does not count as a partition. But it is naturally occuring example of an overlapping free-sorting.

5. The omitted words are: "all on me same level of contrast," which might roughly be translated as "at the same level of generality."

6. In free-association tasks, it is a reliable finding that adults generally reply with a word of the same grammatical class or part of speech (e.g., a noun for a noun), whereas young children typically do not do so (Anglin, 1970, p. 13). This is referred to as the syntagmic-paradigmatic shift.

7. The Interviewers' Handbook (Lockhart and McPherson, 1973) of the Project on Occupational Cognition (Coxon and Jones, 1978, 1979a, b; Coxon et al., 1986) contains the following note on free-listing:

> Pay especial attention to where pauses come, to what principles of clustering seem to be used, and to the possible occurrence of linking terms such as "... and then there's... ." Where possible, the Subject's list is tape-recorded, because it is very likely that interviewers miss this significant information, and also because it is difficult to keep up with a subject when he is going fast (Lockhart and McPherson, 1973, p. 4-1).

8. Borgatti (1998) suggests that free-listing data are arrayed as a $(0, 1)$ subject/row by instance/column matrix which can be reduced to a co-occurrence matrix, $\mathbf{X}$, where the entry $x(i, j)$ gives the number of subjects who produce both instance i and j in their listings. This matrix can be analyzed by methods outlined in Section 4.2.1 below.

9. "Is a kind of" is a basic set-inclusion relation. Other semantic relations are also appropriate, such as "is a place in" (spatial), "is a cause/result of" (causal), "is a reason for" (rationale), "is used for" (function), and "is a characteristic of" (attribution) (Spradley, 1979, p. 111).

10. This technique features as a routine and important preliminary component to the use of free-sorting a set of procedures known as "concept mapping" (Trochim, 1989).

11. The "objects" will often be nouns, but by no means have to be; many components of the subjective lexicon are not just object names, but rather predications: descriptions of and statements about the objects. These can readily be built into a sorting operation (Boster and Johnson, 1989; Coxon and Jones, 1979a, c). More complex objects can also be used, such as the photos, adjectives, and sexual harassment scenarios reported in Rhodes and Stern (1994).

12. The desire to have categories as clear-cut and distinct as possible explains why logic attempts to impose necessary and sufficient conditions as the hallmark of category membership [which "marks virtually all definitions in the tradition of Western reason" as Rosch (1978, p. 35) comments], but human categorizations resist it with equal stubbornness. As Wittgenstein (1958) argued, what is crucial is to conceive the category in terms of clear (prototypical) cases rather than in terms of its boundaries. This explains why emphasis is laid in this text on three consequent points: (1) asking subjects about "best representatives" of their categories, not least because there is often overwhelming agreement about what count as best examples, while there can be wide disagreement about the location of boundaries, (2) examining the most central elements of clustering solutions, and (3) determining why actual fuzziness of boundaries does not necessarily require representation by overlapping models. Concentrating on the most stable, central elements of a sorting category obviates these boundary disputes.

13. In this text "similarity," "dissimilarity," and "proximity" are used interchangeably, with "proximity" used as the generic term to cover both similarity and dissimilarity. This avoids the ugly "dis/similarity" usage.

14. In some applications, where both visual and verbal descriptions are used, the same card can carry both forms on either side (Naveh, 1998).

15. The steps of p(airing), c(haining), and m(erging) can be used to describe the sequence by which the sorting (and the hierarchy variant) is produced (Coxon and Jones, 1979b, U3.16, pp. 202–207) and usually consist of an initial pairing, followed by a stage of chaining to the existing groups, followed by a stage of joining.

16. Many programs for the analysis of sorting data are written for a MS-DOS rather than a Windows™ environment and make specific requirements for the data format. The PDF suggested here is compatible with most programs specified in Appendix 4.

17. Verbal names and their pictorial representation will not necessarily produce identical sorting, or even the same bases for classification. Boster and Johnson (1989) use both photographs and names of fish in a similarity sorting task and show that it is the more obvious morphological characteristics that are given attention in *visual* stimuli, compared with more subtle cultural characteristics (taste, how it is caught) elicited by the *names* of objects.

18. There is some dispute among Q methodologists about whether the forced or unforced variant is preferable (Block, 1956).

19. The agglomerative hierarchies task was described originally in Fillenbaum and Rapoport (1972), who refer to it as "tree-construction," and in the Occupational Cognition Project (Coxon and Jones, 1979a, b, c; Coxon et al., 1986). The task is also referred to as a "successive pile-sort" by Borgatti (1992, pp. 10–13) and Boster (1994). The Q methodology equivalent, dubbed MOSS (the method of successive sorting), was developed by Block (1961).

20. This method is called a "clustering technique" by Bricker et al. (1969) and by Sherman (1972) and Borgatti (1992) and a "merge method" by Clark (1968).

21. The number of possible partitions of a set of p objects is known as the Bell Number (Moser and Wyman, 1955; Arabie and Boorman, 1973). For 6 objects, there are 203 possible partitions, and for 12 there are already over 4 million possible partitions.

22. See Appendix 1 for definition.

23. If set A is a *proper subset* of (or is *strictly included* in) set B (A ⊂ ⊂ B) then A ≠ B and there is at least one element that is in B and not in A, and A is hence smaller than B.

24. Intersection is not restricted to two-way comparison and can be extended to *n*-way comparison. However, it is highly unlikely that identical categories will be found in any large number of sortings unless the categories are highly stereotypical.

25. Kendall's S (the numerator of the ordinal correlation measure τ) can be interpreted as the number of element-wise changes necessary to convert one rank-ordering into another (Kendall, 1962, p. 8).

26. Intuitively, the Arabie-Boorman family of dissimilarity measures for individuals' sorting are related by the nobon of the number of admissible transformations that convert a given partition into another [see Day (1981) for a systematic review; Arabie and Boorman, 1973; Boorman and Olivier, 1973].

27. Implemented as a computer program by Fisher and Hoffman (1988) and by Saltstone and Strange (1996).

28. The form of the Arabie and Hubert Adjusted Rand Index is: $(a + d - n_c)/(a + b + c + d - n_c)$. It outbehaves the Jaccard and the "raw" Rand Index in recovering a "true" (known) cluster structure from error-perturbed data (see Milligan and Cooper (1986), who also reproduce the definition of the n_c term in the Adjusted Rand Index.

29. For example, a *dis*similarity measure is the required form in the command menu of the Windows™ SPSS screen version of ALSCAL, a program often used to analyze such data. SPSS/ALSCAL can only be made to accept a *similarity* measure by using the supplemental SYNTAX file options.

30. Borg and Groenen (1997, p. 103) argue that individual differences in discrimination ought to be built into data modeling criteria rather than "building them into the data." It is a curious objection, not least because it is the differences that constitute the data.

31. In information theory, the more frequent a code is, the less information it contains. Consequently, a code that only occurs infrequently contains considerably more information, and in that sense "surprises" the receiver.

32. This formula is also sometimes given in terms of its square root, a change that has no effect on nonmetric analyses such as multidimensional scaling and hierarchical clustering scheme solutions but will have a nonlinear effect in metric analyses.

33. If objects A and B occur in a group because a subject judges them as similar, and B occurs with C also because of their similarity, then A and C must necessarily co-occur in the group, even if their similarity is low.

34. Compare Miller (1969, p. 170): $N = 48$ nouns, range = 6–26; mean = 14.3 groups and SD = 5.0.

35. Arabie and Hubert (1992, p. 179) cite the result that number of distinct partitions of n objects into m categories increases approximately as: $m^n/m!$.

36. This and a range of other approaches have been discussed under the heading of "Consensus Classifications," a problem first addressed by Régnier in 1965, and a special issue of the *Journal of Classification* has been devoted to the subject [Day, 1986; see also Day (1988) for a historical review].

37. This is the default option in most simple correspondence analysis programs. However, options also exist (as in SPSS) to perform a column-conditional or unconditional analysis and to use a Euclidean distance measure in preference to a chi-square distance.

38. Strictly speaking, direct comparison between the row and column categories is not possible, except by projection, a feature shared with the parallel MDPREF model (Weller and Romney, 1990; Borg and Groenen, 1997, p. 416), making it necessary to interpret correspondence analysis solutions with circumspection.

39. An extended review of these methods is given in Belbin (1987).

40. Variants exist that allow the number of categories to be changed (Ball and Hall, 1967) and even for putting badly fitting objects into a residual category (Wishart, 1978). However, all such partitioning methods are susceptible to local optima problems and depend heavily on the choice of initial partition. Arabie and Hubert (1994, p. 169) advise caution in using other partitioning solutions as starting positions for iterative methods.

41. See Gordon (1996) for an important recent summary of hierarchical classification and its algorithmic implementation.

42. Miller (1969, p. 181) comments that "the connectedness method tends to emphasize the smallest and probably unreliable values of [the co-occurrence frequencies]; the diameter method tends to suppress them."

43. To find a nonhierarchical dustering of p objects that best represent a triangular matrix involves searching through astronomically large number of ways in which the objects might be clustered: 2,147 million for 5 objects and 10^{76} for 8 objects. The mathematical programming approach actually used to search this space in MAPCLUS, the program implementing the model, is described in Arabie and Carroll (1980a, b), and uses a user-specified (small) number of clusters in which to search for an optimal solution.

44. In a MAPCLUS soluton the inital estimates of the p_{ij} presence/absence parameters are real values and not $(0, 1)$. The mathematical programming iterative procedure initially attempts *both* to incease the variance accounted for *and* (increasingly) drive the presence/absence values to $(0, 1)$. An alternating least-squares "ping-pong" approach concentrates on fitting each cluster and its weight. The details of the MAPCLUS procedure are given in Arabie and Carroll (1980a, b).

45. The reason for scaling in dimensions 5 to 1 is so that use can be made of Spence and Graef's (1974) procedure for comparing obtained data to random data and estimating the "true" dimensionality and the random expected stress$_1$ values.

46. PRO-FIT fits a numerical property as a line vector, using either a linear or nonlinear function. As an alternative, external preference mapping (PREFMAP) fits a property either as a line (phase IV) or as an ideal-point (phase III), using either a metric/linear or ordinal/quasi-nonmetric function.

47. See Appendix 1 (III).

48. Scalings were performed for both the Euclidean and City-block distance models; only the Euclidean solutions are considered here. The scalings used stress$_1$ (with Kruskal's best-fitting monotone function quantities) as the badness-of-fit measure and the primary approach to ties.

49. PAIRBONDS measures should be scaled in two dimensions. A one-dimensional solution should *not* be interpreted, for the contour information will be mapped down onto a line, and points on the same contour will appear along its full length and mixed in with points from other contours, thus destroying the very structural information about contours. This is a situation similar to the "horseshoe" phenomenon, which is often detectable when there is sequential information in the data (Kendall, 1971; Coxon and Jones, pp. 86–90).

50. Nishisato (1981, pp. 11–19) gives an excellent review and exposition of dual scaling procedures, which originate in early statistical work to quantify categorical variables by deriving from the cross-classification of two categorical attributes new "re-scaled" numerical scores that correlate to a maximum extent—the "optimal" scores.

51. The MDSORT procedure has affinities with both Coombsian unfolding analysis (Coombs, 1964), dual scaling and correspondence analysis (Benzécri, 1973; Greenacre, 1984), and a variant of latent class anlysis called latent partition analysis (Wiley, 1967). See Takane (1980, pp. 81–85).

52. Strictly, multiple correspondence analysis (MCA) analyzes the *bivariate* relationships between a set of variables, and in this sense reduces to simple CA when there are only two (row, column) variables.

53. Homogeneity analysis is described in the SPSS *Categories*® documentation and is implemented in the Windows™ version as *Statistics* ⇒ *Data Reduction* ⇒ *Optimal Scaling*.

54. I am grateful to Dr. David Canter for help with this example.

55. This is in part because the subject weights are normalized in such a way that their length from the origin is (approximately) equal to the amount of explained variance in their data. Consequently, the difference between two subject "points" consists of both a directional element and a length element, and these become confounded if the distance between points is compared.

56. Gower (1975) describes the geometry of Procrustean rotation, and Lingoes and Borg (1978) and Borg and Groenen (1997) describe the procedures implemented in their program PINDIS, which is designed to perform increasingly complex transformations of the input configurations. See Commandeur 1991 for a critical review of the procedure.

57. This data set has also been reanalyzed by Daws (1996), using his procedure for examining triple co-occurrences.

58. Rosenberg and Kim (1975) actually calculate Rosenberg's δ as well as F, and Carroll and Arabie (1983) analyze the matrices as conditional as well as unconditional. Only the analysis of the unconditional **F** matrices are reported here.

59. For details, see Coxon and Jones (1979a, pp. 13–45), Coxon and Jones (1979b, pp. 156–178), and Coxon et al. (1986, pp. 122–139).

60. The measures were scaled in 5 to 1 dimensions and compared to Spence and Graef's (1974) simulation estimates of stress$_1$ ($\hat{d}$). The obtained stress$_1$ ($\hat{d}$) value of 0.1687 compares with Spence's random stress of 0.3444 and Arabie and Boorman's 0.275 for D-metric values for 60 partitions.

61. The terminology is varied. Normally "cell" is reserved for use when referring to the divisions (exclusive subsets or "parts") constituting the formal structure of the partition and "subgroup" or "category" for the corresponding empirical sorting structure.

62. In this text, the symbol & ("and") is normally used in preference to the symbol $\cap$ to denote intersection.

63. This appendix and notation are based on Hubert an Arabie (1985, Tables 1 and 2). The dot-notation signifies summation over the relevant subscript.

64. See Hubert and Arabie (1985) for discussion of correction for chance and methods of normalizing the Rand Index.

65. Ramsey (1988) points out that alternating least-squares programs (such as ALSCAL in SPSS and SAS) maximize s-stress, which is based upon the *squared* dissimilarity data values. On the assumption that high values contain highest error, the implications can be serious: the effect is to put most emphasis on fitting the noisiest part of the data. Other alternatives are to use an OLS or MLE program, such as those available in NETLIB and MULTISCALE (see below).