WORKING PAPER 344

EXPLORING THE SIMILARITIES OF DIFFERENT RISKS

S.M. MACGILL

WORKING PAPER
School of Geography
University of Leeds

EXPLORING THE SIMILARITIES OF DIFFERENT RISKS

Multi-factorial data on risk originally collected and analysed by Fischoff *et al* (1978) using the method of factor analysis are re-analysed in this paper using Atkin's (1974) Q-analysis algorithm. This algorithm is presented as a distinctive method for summarising and revealing the internal structure of a data set. Unlike more familiar methods of factor analysis that can be used for a broadly similar purpose, the data is not distorted in any way, for example via the calculation of correlation coefficients.

The "results" which are obtained by invoking Q-analysis are contrasted with those originally obtained from factor analysis, and some comment on their difference is made. In an appendix, an exploration of the same data set via the method of Q-discrimination analysis is made. The latter method is an extension by Johnson (1981) the original Q-analysis algorithm.

S.M. Macgill
December 1982

EXPLORING THE SIMILARITIES OF DIFFERENT RISKS

INTRODUCTION

Data is sometimes presented which is portrayed as showing relatively how acceptable the risks from different activities should be by giving some measure of the number of deaths or injuries they have each given rise to (Kletz 1980). The relative acceptability of different risks on the basis of such data does not, however, seem to accord with how acceptable the different risks actually appear to many groups of people. The discrepancy between how acceptable risks "should be" and how acceptable they "appear to be" has sometimes been interpreted as an indication that some people are "irrational" in the way that they perceive risk. Other authors, however, (Otway 1980, Fischoff *et al* 1981) have argued that there is no irrationality in perceiving certain risks to be fundamentally different from others when the activities with which they are associated and intrinsic characteristics of the risks are fundamentally different. In recognition of the fact that there are many other dimensions involved than measures of fatality or injury when addressing the issue of risk acceptability (Lowrance 1976), there has been much recent interest in the multi-dimensional nature of risk. This paper is written as a contribution to this interest by exploring a new methodology for handling multi-factorial data.

Although it is possible to handle a limited amount of data on the different dimensions of different risks without having to invoke a particular analytical methodology, as the number of risks or the number of dimensions considered increases, the need for methods for summarising and interpreting the wealth of multi-dimensional information increases, if this information is to be assimilated. In this paper, data on the multi-dimensional nature of risk originally collected and analysed by Fischoff *et al* (1978) is re-examined in the light of a relatively new methodology (known as Q-analysis) for interpreting the main features of large data sets. The aim of this paper is to introduce the methodology of Q-analysis in the context of multi-dimensional risk data, and to explore the type of insights that this methodology reveals. Since this paper represents the first explorations of Q-analysis to risk data and thus concentrates on the nature of the insights that may be offered, only the more basic facets of the method will be explored, and little justification for the rather distinctive origin of the data set (the League of Women Voters in Oregon, USA) will be given.* Some preliminary remarks about the existing methods that can be used for a similar purpose will first be made.

As far as the present author is aware, factor analytic methods are usually favoured for this purpose (see, for example, Fischoff *et al* (1978), Slovic *et al* (1981), Slovic *et al* (1980). Factor analysis (or rather, the collection of methods known by that name) embodies a particular set of assumptions and characteristics, and these flavour the nature of the "results" obtained. An alternative range of existing methods which could serve a broadly similar purpose as factor analysis are those of cluster analysis. The choice of the latter (cluster analysis) changes the emphasis away from a summary based on reducing the original range of dimensions (based in turn on intercorrelations between the dimensions) to a much smaller number (maybe two or three) underlying the original range, towards a summary made by grouping directly on the basis of the original dimensions (see, for example, Duran and Odell (1974), Everitt (1974), Goddard and Kirby (1976) for reviews of these methods). The range of approaches under each of the headings "factor analysis" and "cluster analysis" may well each give different results in any particular context. In view of this diversity, any chosen method needs careful justification.

In view of the evident clustering powers of the approach to be explored below, Q-analysis may at first sight appear to be an addition to the eixsting wealth of clustering methods (it may even be an existing method in some heavy disguise)** The wider ideals of Q-analysis go far beyond (are in a sense more ambitious than) those of traditional clustering methods, the clusters it generates being only one of many measures it produces, and forming a springboard for rather than an end-point of data analysis and interpretation.

---

* Notwithstanding this limited objective, it is important to note that some significantly different results to those obtained by a more widely used method are obtained.

** See Macgill (1982c)

Taken purely for its clustering powers, Q-analysis may be found attractive (over existing methods) because it uses the data in its original form (rather than transforming it in some way or using proxy representations)* and exposes the internal fabric of the data particularly well. This can result in a more natural and yet finer discrimination than that obtained from existing methods and may perhaps make it attractive to researchers who have hitherto eschewed the coarseness and unnatural assumptions of existing clustering methods.

Taken for its deeper and more distinctive ideals (see, for example, Atkin (1974) (1980), Gould (1980), Macgill (1982a)), Q-analysis embodies a particular philosophy and theory for the interpretation of data about social systems. Its philosophy is that there are latent structures in all fields of inquiry which can often usefully be revealed. Its theory suggests that these structures play a role in influencing activity (mental or physical) analogous to the way in which three-dimensional (physical) space in the "real world" constrains (in terms of available 'routes') or influences (in terms of "forces") physical movement or activity. In the present paper, it is mainly the clustering powers of Q-analysis that will be invoked, though we can hint at its distinctive interpretations.

Q-ANALYSIS

Q-analysis, as a methodology of data analysis, provides a means of interpreting data in a matrix $\{M_{ij}\}$, say, which in turn represents how the elements of two sets $\{e_i\}$ and $\{d_j\}$, say, are related to each other. In the context of the present paper, $\{e_i\}$ will be risky activities, $\{d_j\}$ will be different dimensions of risk, and $\{M_{ij}\}$ will be the weight with which dimension (or risk descriptor) d_j applies to element (risky activity) e_i.

The method has some affinity with hierarchical clustering though various other summary measures and indicators are also generated. The basis of the clustering is simply the number of descriptors each element shares with others, and the hierarchical levels are determined by different numbers of descriptors. The algorithm involved is geared to exploiting

---

* The latter distort the original data. Moreover the dependence of factor analysis and many existing clustering methods on correlation coefficients can raise the problem of method and objective being incompatible because, if correlation measures are to be judged as significant indicators, independence in the data observations is required.

a binary matrix (a matrix of zeroes and ones). Where the original data matrix $\{M_{ij}\}$ is not binary (as in the present case) it is reduced to a binary matrix in a preliminary step by setting to zero all values below a certain level of significance (below some suitable cut-off) and all others to one. Justification for this preliminary step will be given below. A Q-analysis is performed on the binary matrix so obtained, and (due to a possible arbitrariness involved in choosing the first cut-off level) repeated for all other arguably suitable levels. Thus we have a succession of data cut-offs each to be subjected to an augmented hierarchical clustering method. The data cut-off may at first appear rather clumsy, but it may be rationalised by the desire to examine only the most significant portion of the full data set. It may be additionally justified by arguing that an inherently complex set of data cannot necessarily be most usefully analysed by simple methods. Thus rather than a single cluster pattern and set of summary measures, we will be given several such patterns and sets. Fortunately, the philosophy and theory of Q-analysis also gives guidelines on how to interpret all this.

THE STANDARD Q-ANALYSIS ALGORITHM

As noted above the starting point of the Q-analysis algorithm is a binary matrix relating two sets, in the present context risky activities to descriptors. In Table 1 some hypothetical data are given, an entry of 1 denoting that descriptor d_j applies to element e_i, an entry of 0 denoting that it does not. From this matrix we may identify how many

TABLE 1

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}
e_1	0	0	1	1	0	0	0	0	0	1	1	0
e_2	1	1	0	0	0	0	0	0	0	0	0	0
e_3	0	1	0	1	0	0	1	0	1	0	0	1
e_4	0	0	0	1	1	0	0	0	0	0	0	1
e_5	0	0	0	1	0	0	0	0	0	0	1	0
e_6	0	0	0	0	0	0	0	1	1	0	1	0
e_7	0	0	0	1	0	1	1	1	1	0	0	0

descriptors each activity has in common with all others. This is given in Table 2. It is a convention of Q-analysis that rather than using the information in Table 2 directly, we subtract 1 from all terms. (The reason for this lies in the pre-occupation of Q-analysis with the dimension of "objects" and "spaces", an object in two dimensional space requiring three vertices, an object in three dimensional space requiring four vertices, an object in n dimensional space requiring n+1 vertices, and so on (see Atkin 1981).) Table 3 is the result of the subtraction. In the terminology of Q-analysis, it is called a shared face matrix.

TABLE 2

	e_1	e_2	e_3	e_4	e_5	e_6	e_7
e_1	4	0	1	1	2	1	1
e_2	0	2	1	0	0	0	0
e_3	1	1	5	2	1	1	3
e_4	1	0	2	3	1	0	1
e_5	2	0	1	1	2	1	1
e_6	1	0	1	0	1	3	2
e_7	1	0	3	1	1	2	5

TABLE 3

	e_1	e_2	e_3	e_4	e_5	e_6	e_7
e_1	3	-1	0	0	1	0	0
e_2	-1	1	0	-1	-1	-1	-1
e_3	0	0	4	1	0	0	2
e_4	0	-1	1	2	0	-1	0
e_5	1	-1	0	0	1	0	0
e_6	0	-1	0	-1	0	2	1
e_7	0	-1	2	0	0	1	4

A more subtle representation of the similarities (in terms of risk characteristics) of activities is to list for descending values of an integer q (say) all those activities that have at least q + 1 descriptors; in compiling the listing, any activities that mutually share at least q + 1 descriptors are grouped together. The results of this listing are given in Table 4. (The listing may be obtained numerically by reading from the diagonal in Table 3 either upwards and to the left or to the

TABLE 4 The Q-analysis for the data in Table 1

$q = 4$ (e_3) (e_7)

$q = 3$ (e_1) (e_3) (e_7)

$q = 2$ (e_1) (e_4) (e_6) (e_3e_7)

$q = 1$ (e_1e_5) (e_2) $(e_3e_4e_6e_7)$

$q = 0$ $(e_1e_2e_3e_4e_5e_6e_7)$

right and downwards, looking on each "round" for successive integer values of q in descending order.) A so-called structure vector summarises the number of components at each level of q. For the present example this is given by $\{\overset{4}{2}\ 3\ 4\ 3\ \overset{0}{1}\}$. A measure known as an eccentricity, definable for each activity e_i, indicates its overall similarity to (or distinctiveness from) others ($\hat{q}$ gives its dimension, $\check{q}$ the dimension at which it can be grouped with others). The eccentricities are listed in Table 5. These are the main numerical outputs of the standard Q-analysis algorithm.

TABLE 5 $\hat{q}$, $\check{q}$ and eccentricities

	$\hat{q}$	$\check{q}$	$e = \dfrac{\hat{q} - \check{q}}{\check{q} + 1}$
e_1	3	1	1
e_2	1	0	1
e_3	4	2	0.666
e_4	2	1	0.5
e_5	1	1	0
e_6	2	1	0.5
e_7	4	2	0.666

In the next section, a Q-analysis is worked through and, more importantly, interpreted for the risk data given in Fischoff *et al* (1978). Whereas in the above hypothetical example, a single binary matrix underpinned the analysis, for the data below the analysis will be repeated for several binary matrices, each representing a different level of significance in the original data set.

APPLICATION TO RISK DATA

Participants in the study reported by Fischoff, *et al* (1978) were asked to rate each of thirty activities or technologies (the $\{e_i\}$'s) on nine descriptive features (the $\{d_j\}$'s), each according to a seven-point scale. Each of the dimensions has been hypothesised elsewhere (see, for example, Lowrance) as influencing perceptions of actual or acceptable risk. The nine scales are described in Table 6, and the participants in the study rated all 30 activities and technologies on each scale before proceeding to the next. The responses are given in Table 7. This is

TABLE 6 The nine risk descriptors (Fischoff *et al* 1978)

1. Voluntariness of risk: Do people get into these risk situations voluntarily? If for a single item some of the risks are voluntarily undertaken and some are not, mark an appropriate spot towards the center of the scale. (The scale was labelled: 1 = voluntary; 7 = involuntary.)

2. Immediacy of effect: To what extent is the risk of death immediate - or is death likely to occur at some later time? (1 = immediate; 7 = delayed.)

3. Knowledge about risk: To what extent are the risks known precisely by the persons who are exposed to those risks? (1 = known precisely; 7 = not known.)

4. Knowledge about risk: To what extent are the risks known to science? (1 = known precisely; 7 = not known.)

5. Control over risk: If you are exposed to the risk of each activity or technology, to what extent can you, by personal skill or diligence, avoid death while engaging in the activity? (1 = uncontrollable; 7 = controllable.)

6. Newness: Are these risks new, novel ones or old, familiar ones? (1 = new; 7 = old.)

TABLE 6 (contd.)

7. Chronic-catastrophic: Is this a risk that kills people one at a time (chronic risk) or a risk that kills large numbers of people at once (catastrophic risk)? (1 = chronic; 7 = catastrophic.)

8. Common-dread: Is this a risk that people have learned to live with and can think about reasonably calmly, or is it one that people have great dread for - on the level of a gut reaction? (1 = common; 7 = dread.)

9. Severity of consequences: When the risk from the activity is realized in the form of a mishap or illness, how likely is it that the consequence will be fatal? (1 = certain not to be fatal; 7 = certain to be fatal.) Green (1974) has referred to this as the "sporting chance" factor.

the matrix $\{M_{ij}\}$ which forms the raw data for the following analysis.

It may be readily observed that each descriptor embodies two mutually complementary characteristics and the 1-7 scale may be interpreted as dividing each descriptor into its two parts, values between 1 and 4 relating to one basic characteristic of that descriptor, values between 4 and 7 relating to its opposite. The first step is to draw out from Table 7 which part of the scale is appropriate for each activity. This is done in defining Table 8 from Table 7. Table 8 will play a role in the present analysis analogous to the role of Table 1 in the previous section. It is a crude matching of activities to descriptors, but refinements will be introduced later.

The second step is to identify, on the basis of the information in Table 8, the natural groupings of the data at each dimensional level. The shared face matrix, Q-analysis results, eccentricities and structure vector are given in Tables 9, 10 and 11 and Figure 1. The fact that all activities are grouped together at Q = 5 (Table 9), reflects the fact that each activity has at least six basic features of risk in common with at least one other activity. At Q = 6 we see that nuclear power and pesticides have seven features of risk in common with each other (a set of features which they share with no other activities). That contraceptives, food colouring, food preservatives, prescriptions, spray cans and x-rays are all grouped together means that each of these has at least seven features in common with

TABLE 7 Mean ratings for nine characteristics of risk (Fischoff et al. 1978)

		$d_1, \hat{d}_1$ Voluntariness 1=voluntary	$d_2, \hat{d}_2$ Immediacy 1=immediate	$d_3, \hat{d}_3$ Known to exposed 1=precisely	$d_4, \hat{d}_4$ Known to science 1=precisely	$d_5, \hat{d}_5$ Controllability 1=can't be controlled	$d_6, \hat{d}_6$ Newness 1=new	$d_7, \hat{d}_7$ Chronic-Catastrophic 1=chronic	$d_8, \hat{d}_8$ Common-Dread 1=common	Severity of consequences 1=certain not to be fatal
e_1.	Alcoholic beverages	2.10	5.34	3.77	1.98	5.57	6.61	1.79	1.92	4.40
e_2.	Bicycles	1.90	2.82	3.27	2.80	4.99	5.19	1.30	1.74	3.77
e_3.	Commercial aviation	2.80	1.05	3.24	2.12	2.18	4.24	6.09	3.39	5.72
e_4.	Contraceptives	2.74	5.69	4.66	3.88	3.11	2.25	1.49	3.14	4.08
e_5.	Electric power	4.40	2.82	3.98	2.68	4.25	5.09	2.66	1.72	4.52
e_6.	Fire fighting	2.40	2.33	1.98	2.25	4.03	6.01	2.84	2.62	4.42
e_7.	Food colouring	5.86	6.26	6.40	4.77	2.70	2.66	2.82	3.24	3.59
e_8.	Food preservatives	5.65	6.18	6.39	4.76	2.70	2.73	2.82	3.32	3.66
e_9.	General aviation	2.20	1.66	2.96	2.60	3.99	4.08	3.40	3.15	5.63
e_{10}.	Handguns	3.42	1.65	2.64	2.41	4.05	5.69	2.10	4.40	5.67
e_{11}.	H.S. and College football	1.90	3.52	3.66	3.11	4.15	4.78	1.40	1.95	3.15
e_{12}.	Home appliances	3.61	2.97	4.47	2.90	4.85	4.39	1.38	1.43	3.08
e_{13}.	Hunting	2.01	1.66	2.62	2.64	4.45	6.14	1.59	2.79	4.91
e_{14}.	Large construction	3.07	2.23	2.77	2.51	3.91	5.04	3.04	2.61	4.77
e_{15}.	Motorcycles	1.87	1.76	2.69	2.17	4.08	4.31	1.59	3.02	5.19
e_{16}.	Motor vehicles	4.04	2.33	3.14	2.31	4.19	4.73	3.28	3.04	4.57
e_{17}.	Mountain climbing	1.15	1.78	1.83	2.49	4.98	5.63	1.32	2.57	4.80
e_{18}.	Nuclear power	6.51	5.08	5.85	4.83	1.36	1.35	6.43	6.42	5.98
e_{19}.	Pesticides	5.77	5.57	5.50	4.41	2.14	2.22	4.75	5.21	4.87
e_{20}.	Power mowers	2.23	2.99	3.31	2.60	5.13	3.70	1.16	1.75	2.75
e_{21}.	Police work	2.44	2.14	2.05	2.25	3.76	5.50	2.07	3.05	4.35
e_{22}.	Prescription antibiotics	4.44	4.33	5.40	3.91	2.77	2.87	2.35	2.19	3.82
e_{23}.	Railroads	3.42	2.91	3.66	2.68	3.22	5.49	4.49	1.75	3.60
e_{24}.	Skiing	1.28	2.45	2.47	2.51	4.73	4.69	1.06	1.92	3.15
e_{25}.	Smoking	1.85	6.11	2.86	2.15	4.43	5.04	1.68	2.89	5.01
e_{26}.	Spray cans	3.80	6.06	5.43	4.16	3.60	1.89	3.82	3.62	4.27
e_{27}.	Surgery	4.28	2.71	3.84	2.86	2.39	4.95	1.14	4.04	4.68
e_{28}.	Swimming	1.64	1.76	2.87	2.68	5.17	6.50	1.16	1.89	4.78
e_{29}.	Vaccinations	3.82	3.71	4.84	2.82	2.53	4.50	1.88	2.03	3.62
e_{30}.	X-rays	4.38	6.15	5.05	3.28	2.37	4.02	1.99	2.58	4.20

TABLE 8 The most basic relationship between risk descriptors and risky activities

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_3$	$\hat{d}_4$	$\hat{d}_5$	$\hat{d}_6$	$\hat{d}_7$	$\hat{d}_8$	$\hat{d}_9$
e_1	1	0	1	1	0	0	1	1	0	0	1	0	0	1	1	0	0	1
e_2	1	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	0	0
e_3	1	1	1	1	1	0	0	1	0	0	0	0	0	0	1	1	0	1
e_4	1	0	0	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1
e_5	0	1	1	1	0	0	1	1	0	1	0	0	0	1	1	0	0	1
e_6	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1
e_7	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
e_8	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0
e_9	1	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1
e_{10}	1	1	1	1	0	0	1	0	0	0	0	0	0	1	1	0	1	1
e_{11}	1	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	0	0
e_{12}	1	1	0	1	0	0	1	1	1	0	0	1	0	1	1	0	0	0
e_{13}	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1
e_{14}	1	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1
e_{15}	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1
e_{16}	0	1	1	1	0	0	1	1	0	1	0	0	0	1	1	0	0	1
e_{17}	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1
e_{18}	0	0	0	0	1	1	0	0	0	1	1	1	1	0	0	1	1	1
e_{19}	0	0	0	0	1	1	0	0	0	1	1	1	1	0	0	1	1	1
e_{20}	1	1	1	1	0	1	0	0	0	0	0	0	0	1	0	1	1	1
e_{21}	1	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1
e_{22}	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
e_{23}	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0
e_{24}	1	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	0	0
e_{25}	1	0	1	1	0	0	1	1	0	0	1	0	0	1	1	0	0	1
e_{26}	1	0	0	0	1	1	1	1	0	0	1	1	1	0	0	0	0	1
e_{27}	0	1	1	1	1	0	1	0	0	1	0	0	0	0	1	0	1	1
e_{28}	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	1
e_{29}	1	1	0	1	1	0	1	1	1	0	0	1	0	0	1	0	0	0
e_{30}	0	0	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0	1

TABLE 9 Shared face matrix corresponding to Table 8

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	9	6	5	5	6	7	2	2	6	6	6	5	7	6	7	6	7	1	1	5	6	3	4	6	8	4	4	7	4	5
2	6	9	5	3	6	7	2	2	6	6	9	7	7	6	7	6	7	-1	-1	7	6	3	6	8	6	2	4	7	5	3
3	5	5	8	4	5	6	1	1	7	5	5	4	6	7	6	5	5	2	2	4	7	2	7	5	5	3	5	6	5	4
4	5	3	4	8	3	4	5	5	5	3	3	4	4	5	4	3	4	4	4	4	5	6	3	3	5	7	3	4	5	6
5	6	6	5	3	9	7	2	2	6	6	6	5	7	6	7	8	7	1	1	5	6	3	4	6	6	2	6	7	4	5
6	7	7	6	4	7	8	1	1	7	7	7	6	8	7	8	7	9	0	0	6	7	2	5	7	7	3	5	8	5	4
7	2	2	1	5	2	1	8	8	2	0	2	3	1	2	1	2	1	5	5	3	2	7	2	2	2	6	2	1	4	5
8	2	2	1	5	2	1	8	8	2	0	2	3	1	2	1	2	1	5	5	3	2	7	2	2	2	6	2	1	4	5
9	6	6	7	5	6	7	2	2	8	6	6	5	7	9	7	6	7	1	1	5	8	3	6	6	6	4	6	7	6	5
10	6	6	5	3	6	7	0	0	6	9	6	5	7	6	7	6	7	1	1	5	6	1	4	5	6	2	6	7	4	3
11	6	9	5	3	6	7	2	2	6	6	9	7	7	6	7	6	7	-1	-1	7	6	3	6	8	6	2	4	7	5	3
12	5	7	4	4	5	6	3	3	5	5	7	8	6	5	6	5	5	0	0	6	5	4	5	7	5	3	3	6	7	4
13	7	7	6	4	7	8	1	1	7	7	7	5	9	7	8	7	8	0	0	6	7	2	5	7	7	3	5	8	5	4
14	6	6	7	5	6	7	2	2	8	6	6	5	7	8	7	6	7	1	1	5	9	3	6	6	6	4	6	7	6	5
15	7	7	6	4	7	8	1	1	7	7	7	5	8	7	8	7	9	0	0	6	7	2	5	7	7	3	5	8	5	4
16	6	5	5	3	9	7	2	2	6	6	6	5	7	6	7	8	7	1	1	5	6	3	4	6	6	2	6	7	4	5
17	7	7	6	4	7	8	1	1	7	7	7	6	8	7	8	7	8	0	0	6	7	2	5	7	7	3	5	8	5	4
18	1	-1	2	4	1	0	5	5	1	1	-1	0	0	1	0	1	0	8	9	0	1	4	1	-1	1	5	3	0	1	4
19	1	-1	2	4	1	0	5	5	1	1	-1	0	0	1	0	1	0	8	9	0	1	4	1	-1	1	5	3	0	1	4
20	5	7	4	4	5	6	3	3	5	5	7	6	6	5	6	5	5	0	0	8	5	4	5	7	5	3	3	6	5	2
21	6	6	7	5	6	7	2	2	8	6	6	5	7	9	7	6	7	1	1	5	9	3	6	6	6	4	6	7	5	5
22	3	3	2	6	3	2	7	7	3	1	3	4	2	3	2	3	2	4	4	4	3	8	3	3	3	5	3	2	5	6
23	4	6	7	3	4	5	2	2	6	4	6	5	5	5	5	4	5	1	1	5	6	3	8	6	4	2	4	5	6	3
24	6	8	5	3	6	7	2	2	6	6	9	7	7	6	7	6	7	-1	-1	7	6	3	6	8	6	2	4	7	6	3
25	9	6	5	5	6	7	2	2	6	6	6	5	7	6	7	6	7	1	1	5	6	3	4	6	8	4	4	7	[illegible]	5
26	4	2	3	7	2	3	6	6	4	2	2	3	3	4	3	2	3	5	5	3	4	5	2	2	4	8	2	3	4	5
27	4	4	5	3	6	5	2	2	6	6	4	3	5	6	5	6	5	3	3	3	5	3	4	4	4	2	8	5	4	5
28	7	7	6	4	7	8	1	1	7	7	7	6	8	7	8	7	9	0	0	6	7	2	5	7	7	3	5	8	5	4
29	4	5	5	5	4	5	4	4	6	4	6	7	5	6	5	4	5	1	1	5	6	5	6	6	4	4	4	5	8	5
30	5	3	4	6	5	4	5	5	5	3	3	4	4	5	4	5	4	4	4	2	5	6	3	3	5	5	5	4	5	8

TABLE 10 Results from the Q-analysis algorithm given the data in Table 8

```
  (Q=8)
COMPONENT  - (  1)
                  ALCBEV, SMOKING
             *** NEW SIMPLEX ***   ALCBEV
             *** NEW SIMPLEX ***   SMOKING
COMPONENT  - (  2)
                  BIKES, FTBLL, SKIING
             *** NEW SIMPLEX ***   BIKES
             *** NEW SIMPLEX ***   FTBLL
             *** NEW SIMPLEX ***   SKIING
COMPONENT  - (  3)
                  COMMAV
             *** NEW SIMPLEX ***   COMMAV
COMPONENT  - (  4)
                  CONTRAS
             *** NEW SIMPLEX ***   CONTRAS
COMPONENT  - (  5)
                  ELECPWR, MCARS
             *** NEW SIMPLEX ***   ELECPWR
             *** NEW SIMPLEX ***   MCARS
COMPONENT  - (  6)
                  FIREFT, HUNTNG, MBIKES, MTNCLING, SWIMG
             *** NEW SIMPLEX ***   FIREFT
             *** NEW SIMPLEX ***   HUNTNG
             *** NEW SIMPLEX ***   MBIKES
             *** NEW SIMPLEX ***   MTNCLING
             *** NEW SIMPLEX ***   SWIMG
COMPONENT  - (  7)
                  FOODCOL, FOODPRES
             *** NEW SIMPLEX ***   FOODCOL
             *** NEW SIMPLEX ***   FOODPRES
COMPONENT  - (  8)
                  GENAV, LGECONS, POLICE
             *** NEW SIMPLEX ***   GENAV
             *** NEW SIMPLEX ***   LGECONS
             *** NEW SIMPLEX ***   POLICE
COMPONENT  - (  9)
                  HANDGUNS
             *** NEW SIMPLEX ***   HANDGUNS
COMPONENT  - ( 10)
                  HOMEAPPL
             *** NEW SIMPLEX ***   HOMEAPPL
COMPONENT  - ( 11)
                  NUCLEAR, PESTS
             *** NEW SIMPLEX ***   NUCLEAR
             *** NEW SIMPLEX ***   PESTS
COMPONENT  - ( 12)
                  POWERMWRS
             *** NEW SIMPLEX ***   POWERMWRS
COMPONENT  - ( 13)
                  PRESCPTNS
             *** NEW SIMPLEX ***   PRESCPTNS
COMPONENT  - ( 14)
                  RAILRDS
             *** NEW SIMPLEX ***   RAILRDS
COMPONENT  - ( 15)
                  SPRAYCNS
             *** NEW SIMPLEX ***   SPRAYCNS
COMPONENT  - ( 16)
                  SURGERY
             *** NEW SIMPLEX ***   SURGERY
COMPONENT  - ( 17)
                  VACCNS
             *** NEW SIMPLEX ***   VACCNS
COMPONENT  - ( 18)
                  XRAYS
             *** NEW SIMPLEX ***   XRAYS
  (Q=7)
COMPONENT  - (  1)
                  ALCBEV, BIKES, COMMAV, ELECPWR, FIREFT, GENAV, HANDGUNS, FTBLL, HOMEAPPL, HUNTNG, LGECONS, MBIKES,
                  MCARS, MTNCLING, POWERMWRS, POLICE, RAILRDS, SKIING, SMOKING, SWIMG, VACCNS
COMPONENT  - (  2)
                  FOODCOL, FOODPRES, PRESCPTNS
COMPONENT  - (  3)
                  CONTRAS, SPRAYCNS
COMPONENT  - (  4)
                  NUCLEAR, PESTS
COMPONENT  - (  5)
                  SURGERY
COMPONENT  - (  6)
                  XRAYS
  (Q=6)
COMPONENT  - (  1)
                  ALCBEV, BIKES, COMMAV, ELECPWR, FIREFT, GENAV, HANDGUNS, FTBLL, HOMEAPPL, HUNTNG, LGECONS, MBIKES,
                  MCARS, MTNCLING, POWERMWRS, POLICE, RAILRDS, SKIING, SMOKING, SURGERY, SWIMG, VACCNS
COMPONENT  - (  2)
                  CONTRAS, FOODCOL, FOODPRES, PRESCPTNS, SPRAYCNS, XRAYS
COMPONENT  - (  3)
                  NUCLEAR, PESTS
  (Q=5,4,3,2,1,0)
COMPONENT  - (  1)
                  ALCBEV, BIKES, COMMAV, CONTRAS, ELECPWR, FIREFT, FOODCOL, FOODPRES, GENAV, HANDGUNS, FTBLL,
                  HOMEAPPL, HUNTNG, LGECONS, MBIKES, MCARS, MTNCLING, NUCLEAR, PESTS, POWERMWRS, POLICE, PRESCPTNS,
                  RAILRDS, SKIING, SMOKING, SPRAYCNS, SURGERY, SWIMG, VACCNS, XRAYS
```

TABLE 11 Eccentricities for Table 10 and Figure 1

8 0
(obstruction vector is (18 6 3 1 1 1 1 1 1)

ECCENTRICITIES NAME	FORMULA I (TOPO+1)-(BOTO+1) (BOTO+1)	FORMULA II (TOPO+1)-BOTO+1) (TOPO+1)	TOPO	BOTO
ALCBEV	0.000	0.000	8	8
BIKES	0.000	0.000	8	8
COMMAV	0.125	0.111	8	7
CONTRAS	0.125	0.111	8	7
ELECPWR	0.000	0.000	8	8
FIREFT	0.000	0.000	8	8
FOODCOL	0.000	0.000	8	8
FOODPRES	0.000	0.000	8	8
GENAV	0.000	0.000	8	8
HANDGUNS	0.125	0.111	8	7
FTBALL	0.000	0.000	8	8
HOMEAPPL	0.125	0.111	8	7
HUNTNG	0.000	0.000	8	8
LGECONS	0.000	0.000	8	8
MBIKES	0.000	0.000	8	8
MCARS	0.000	0.000	8	8
MTNCLING	0.000	0.000	8	8
NUCLEAR	0.000	0.000	8	8
PESTS	0.000	0.000	8	8
POWERMWS	0.125	0.111	8	7
POLICE	0.000	0.000	8	8
PRESCPTNS	0.125	0.111	8	7
RAILRDS	0.125	0.111	8	7
SKIING	0.000	0.000	8	8
SMOKING	0.000	0.000	8	8
SPRAYCNS	0.125	0.111	8	7
SURGERY	0.285	0.222	8	6
SWIMG	0.000	0.000	8	8
VACCNS	0.125	0.111	8	7
XRAYS	0.285	0.222	8	6

TABLE 12 Eccentricities for Figure 5

ECCENTRICITIES NAME	FORMULA I (TOPC+1-(ECTC+1)) / (BOTC+1)	FORMULA II (TOPO+1)-BOTO+1) / (TOPO+1)	TOPO	BOTO
ALCREV	0.000	0.000	0	0
BIKES	0.000	0.000	0	0
COMSAV	********	********	-1	-1
CONTRAS	0.000	0.000	0	0
ELECPWR	********	********	-1	-1
FISEPT	********	********	-1	-1
FCCDCOL	********	********	-1	-1
FOODPRES	********	********	-1	-1
GENAV	********	********	-1	-1
HANDGUNS	********	********	-1	-1
FTBLL	0.000	0.000	0	0
HCRSAPPL	1.000	0.500	1	0
HUNTNG	********	********	-1	-1
LGECONS	********	********	-1	-1
MBIKES	********	********	-1	-1
ECARS	********	********	-1	-1
STRCLING	0.000	0.000	1	1
NUCLEAR	********	1.000	2	-1
PESTS	********	********	-1	-1
POWERMWPS	0.000	0.000	0	0
POLICE	********	********	-1	-1
PRESCPTNS	********	********	-1	-1
RAILRDS	********	********	-1	-1
SKIING	0.000	0.000	1	1
SMOKING	********	********	-1	-1
SPRAYCNS	********	********	-1	-1
SURGERY	0.000	0.000	0	0
SWIMG	1.000	0.500	1	0
VACCNS	********	********	-1	-1
XRAYS	********	********	-1	-1

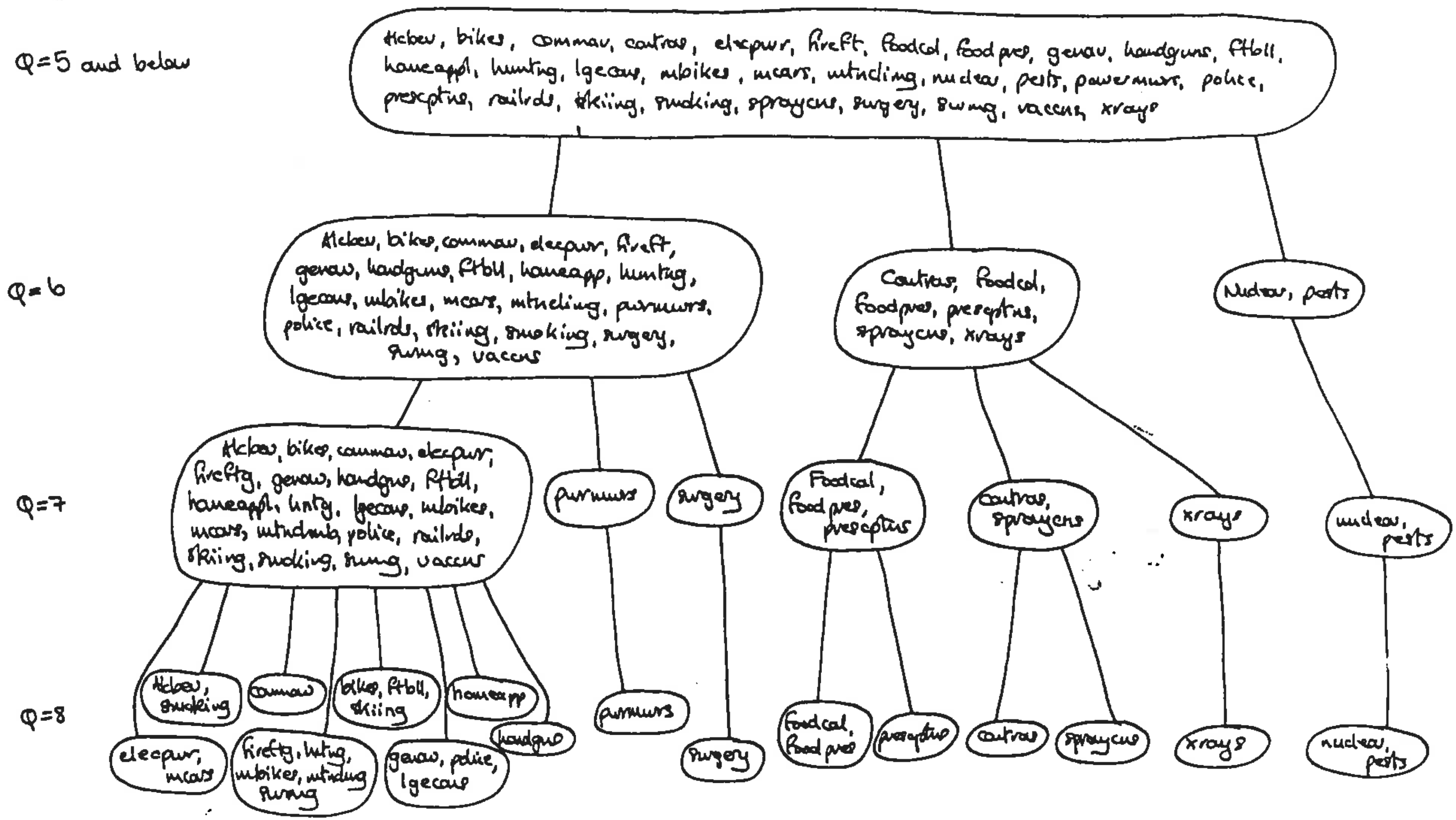


Figure 1 Groupings of risky activities generated by Q analysis algorithm, all values of Mij

some other activity in that group. The third and largest group at that level is interpreted in a similar way. At the next level, seven groups have been identified, and these would be interpreted in a corresponding way, depicting activities linked by eight risk features. The other level depicted is Q = 8. This is the last level because there were only nine basic risk features. (Although the original scale was augmented from nine to eighteen descriptors, each activity has just nine descriptors.) Activities that are not discriminated at this level are not at all distinctive from others, that is, they all register the same nine risk attributes as each other. Thus nuclear power and pesticides have the same nine basic risk features as each other; alcoholic beverages and smoking have the same nine (obviously, a different nine in this case); bikes, football and skiing all have the same nine risk features as each other. But activities or technologies such as prescriptions, handguns, commercial aviation and others that stand in isolation each have a set of nine risk features that is not shared by any other activity or technology. Particularly notable in Figure 1 is the relatively isolated position of nuclear power and pesticides. This isolation is, however, not shown up in the eccentricities, since these reflect the distinctiveness (or otherwise) of individual activities, not of pairs. (Thus, the lower the eccentricity, the less distinctive a given activity is from others; the higher the eccentricity, the more distinctive it is).

The Q-analysis given above thus provides a means of re-interpreting the original data, the clusters identifying immediately which activities are most similar to each other, on the basis of the risk descriptors used (and held in common). Although the "results" do not specify which particular sets of risk characteristics are involved in each case, this can be readily found from the original data. Q-analysis has suggested what to look for and where to look. Some elaboration will be given below as to why clusters based on numbers of shared descriptors is worthwhile. A further means of interpreting the above results will first be indicated.

Rather than focussing on the clusters in Figure 1, it may be worthwhile to examine one particular activity, and see what it has in common with others. The shared face matrix is a useful starting point here. For the chosen activity (for example, nuclear power) we may read upwards and to the right (or downwards and to the left) to see how many risk descriptors it has in common with others. It is then necessary to refer back to the original data to see what these descriptors are.

The Q-analysis has so far been based on a fundamental cut-off between the risk characteristics, and thus consequently somewhat coarse. Due to the nature of the problem, scale values in the middle of the 1-7 range (values of 3, 4 and 5, say) are in a sense much less significant than values towards the ends (values of 1, 2, 6 and 7). A second application of the Q-analysis algorithm will now be explored which discards the less significant values (specifically, values $3 < M_{ij} < 5$ will be "sliced out").* The clusters in this case are given in Figure 2. The significance of the underlining in Figure 2 is to denote the highest Q-level at which each activity appears. Note in this connection that since the middle of the range of scale values for risk descriptors has now been discarded, all activities no longer register on all nine descriptors; some register only on as few as 2 or 3 (motor cars, football, spray cans).

The Q-values in this case cover the lower range, the highest level at which <u>all</u> activities are identified being Q = 1. (The analysis is now more discriminating, there is a more stringent rule to be satisfied for a descriptor to be deemed significant, namely that it must register a value greater than 5 or less than 3). It can be seen from Figure 2 that when the less significant range of scale values are sliced at, 2 is the greatest number of descriptors that all activities can register; all except "motor cars" have 3 "significant" descriptors (see Q = 2 level). Since they are all in one component at this level, they each have at least 3 descriptors in common with some other. At Q = 3 much more discrimination is seen. Vaccinations, surgery and contraceptives have four descriptors (and only four, as indicated by the underlining), in each case these four being shared by no other activity. Food colouring, food preservatives, nuclear power, pesticides, prescriptions and x-rays all pairwise share four descriptors (since they in fact have more than four, they are not underlined); a similar interpretation may be made of the larger component at that level. At Q = 4, prescriptions and x-rays have five descriptors (the underlining denotes that at least some are distinctive to them). A corresponding interpretation of others at this and other levels can hopefully by now be left to the reader. At the richest level, Q = 7, nuclear power, smoking and swimming each have their own set of 8 descriptors (some may be mutually shared, others not).

---

* Values within the range ($3 < M_{ij} < 5$) will be replaced by 0; value outside this range will be replaced by 1; the Q-analysis algorithm is applied to the resulting matrix of zeroes and ones.

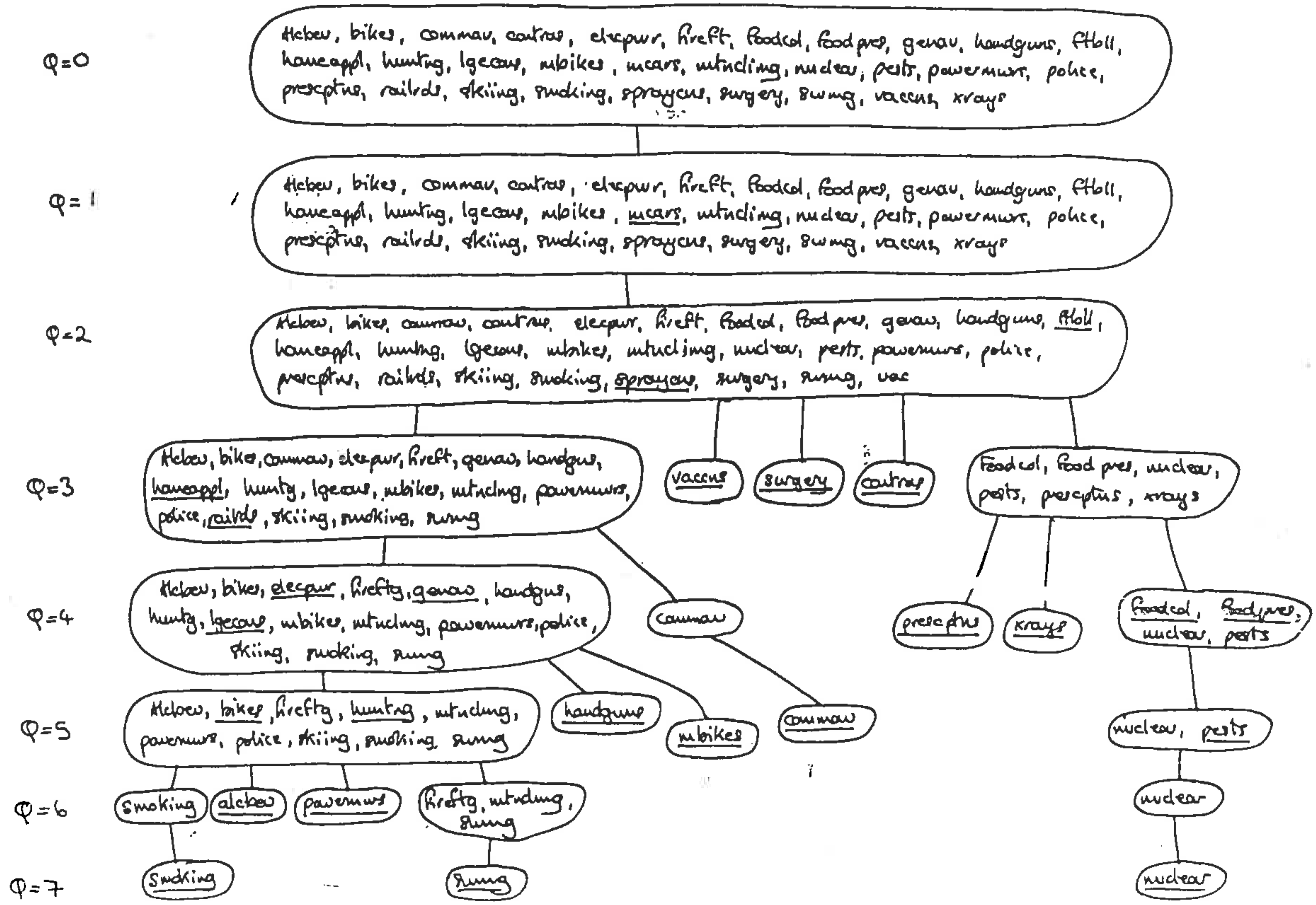


Figure ? [illegible] activities generated by Φ analysis algorithm, values $3 < M_{ij} < 5$ discarded

Thus, comparing Figures 1 and 2, the pattern of clustering (not surprisingly) is different when the middle range is sliced out. It is now appropriate to see what happens when successively stricter levels of significance are imposed. In Figures 3, 4 and 5, the patterns of clusters resulting from various different cut-offs are presented. In Figure 3, the range 2.5 to 5.5 is discarded; in Figure 4, the range 2 to 6 is discarded; in Figure 5, the range 1.5 to 6.5 is discarded. Thus as we move through towards this last figure, only those descriptors that register very strongly (right at the ends of the scale values) are incorporated in the analysis. It is hopefully self-evident that there is a sense in which these are relatively more interesting than the middle ranges.

Moving from Figure 1 through to Figure 5, note that successively fewer activities have been identified. The most extreme case, Figure 5, only 3 Q-levels occur and a very sparse distribution of activities. These are the activities that register on characteristics that people feel most strongly about. Nuclear power stands out over all others, not only because it has the greatest number of descriptors registering at this level, but also because it does not connect to any of the other activities at any level. A similar feature of nuclear power is seen in Figure 4, but the weaker cut-off point in this case enables it to connect at Q = 0 (that is, there is a descriptor that some other activity also shares at this level). Further similar interpretation of Figures 3 and 4 will be left to the reader.

FURTHER INSIGHTS

The interpretations so far have rested on using Q-analysis as a clustering device. A richer interpretation of the results obtained may be derived by appealing to the wider philosophy and theory of Q-analysis, and by contrasting the Q-analysis results with the factor space representation originally given by Fischoff *et al* (1978). The latter is reproduced in Figure 6.

The original nine risk dimensions have been aggregated into two basic underlying factors, "the first factor correlated highly with all characteristics except severity of consequences. The second factor was

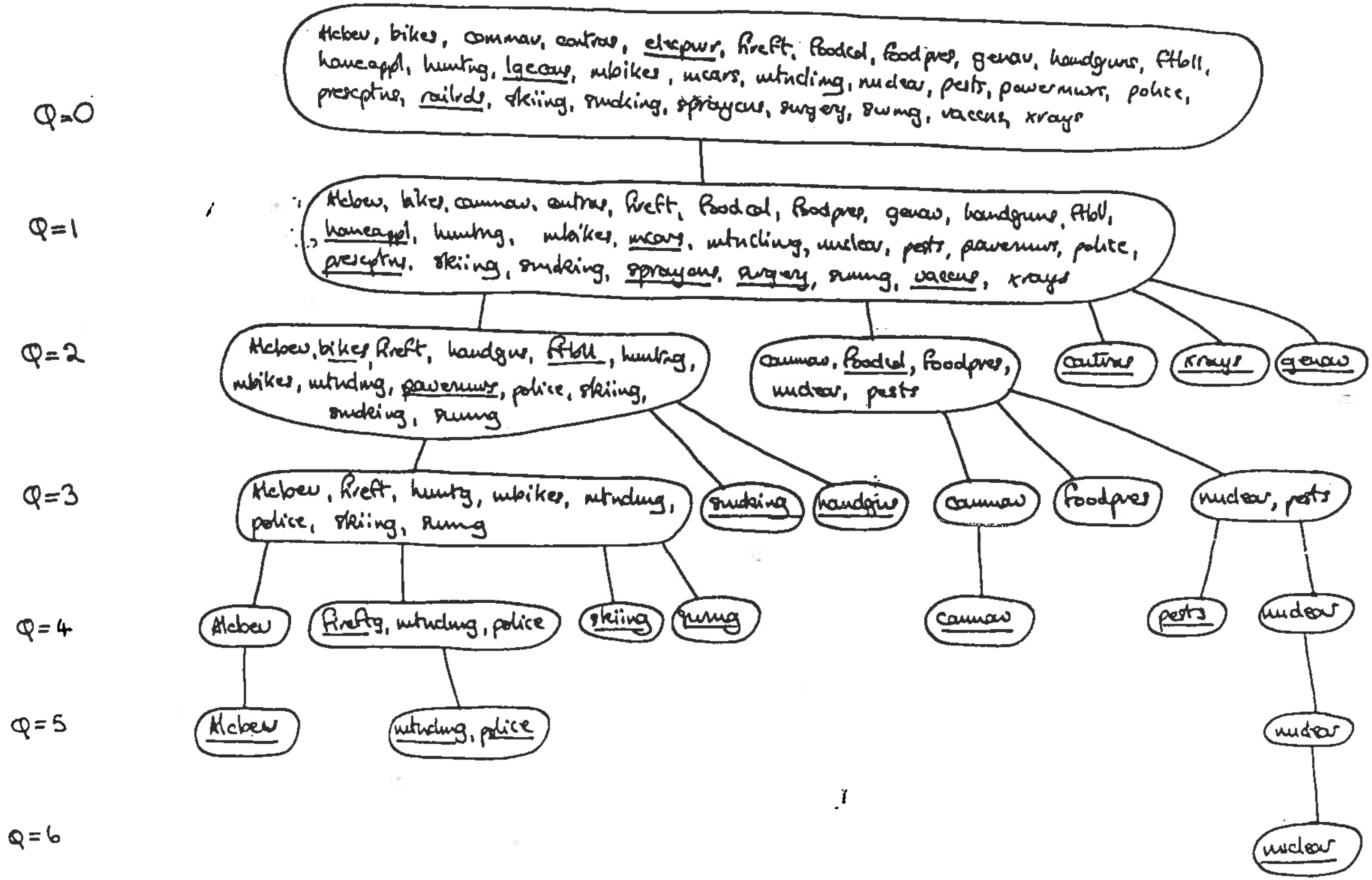


Figure 3 Groupings of risky activities generated by Q-analysis algorithm; values $2.5 < M_{ij} < 5.5$ discarded

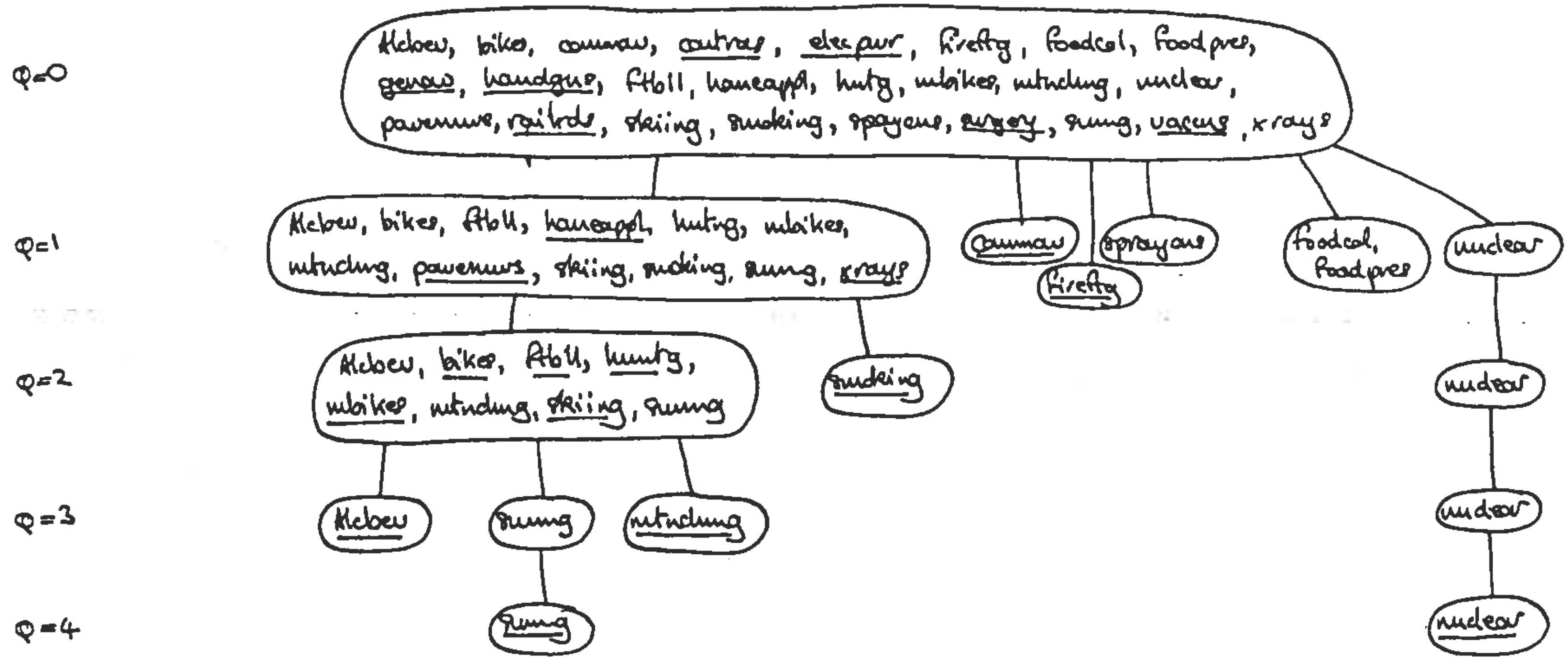


Figure 4 Groupings of risky activities generated by Q analysis algorithm, values 2 < Mij < 6 discarded

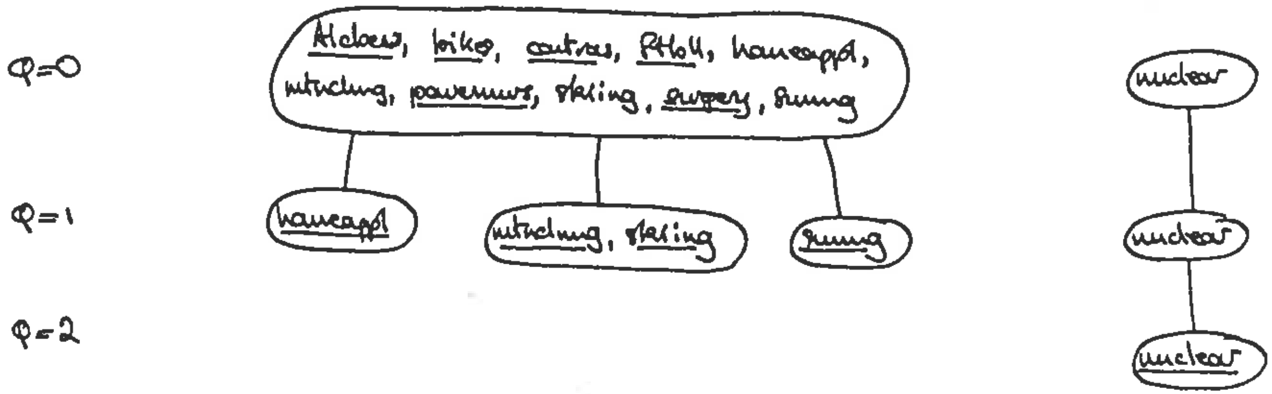


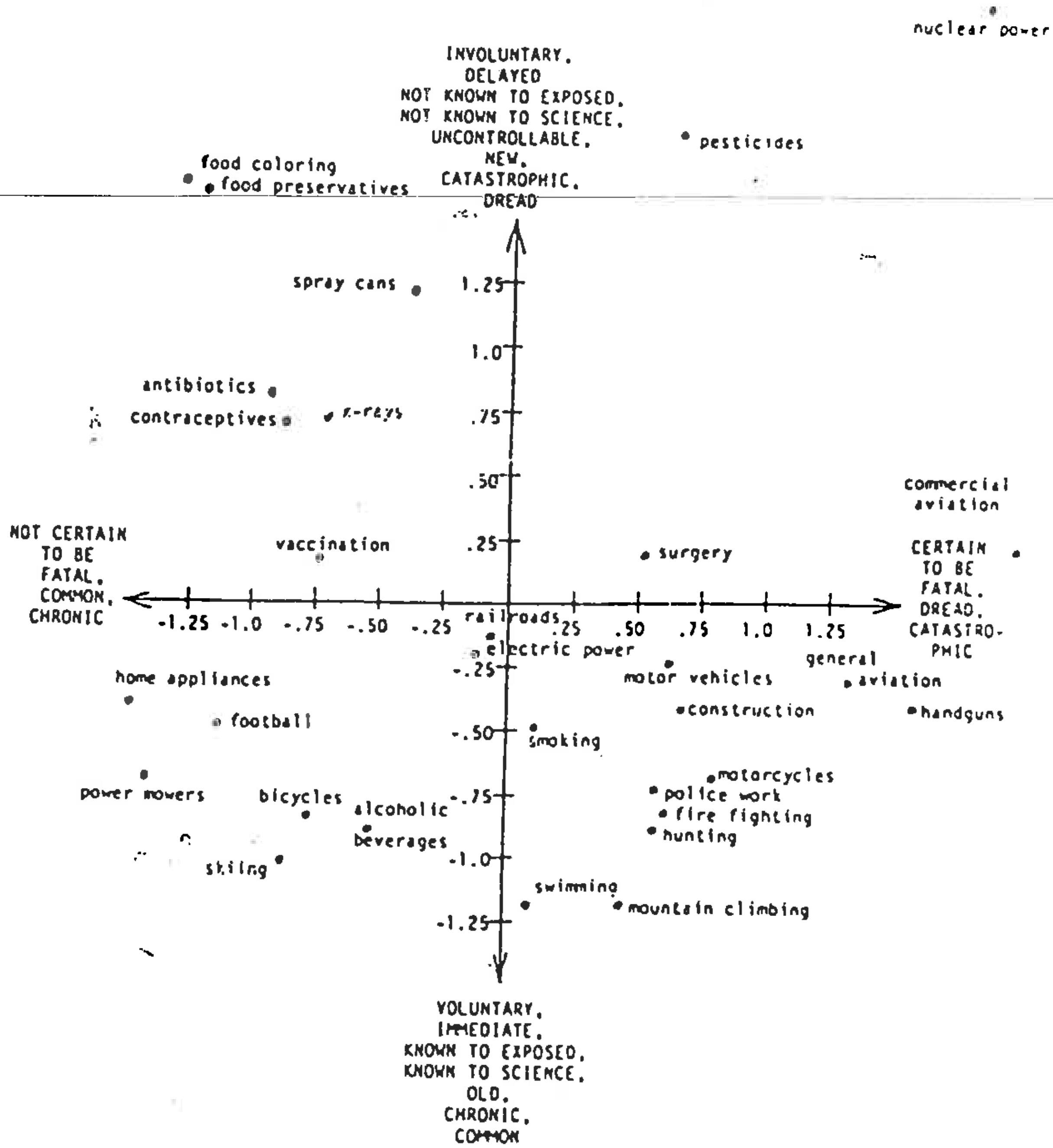
Figure 5 Groupings of activities generated by Q-analysis algorithm, values $1{\cdot}5 < M_{ij} < 6{\cdot}5$ discarded

associated with severity of consequences and, to a lesser extent, with common/dread and chronic/catastrophic. The communality index ... reflects the extent to which the two factors accounted for each of the ratings. The communalities were high, indicating that this two-factor solution did a good job of representing the ratings for the nine scales." (Fischoff *et al* , 1978, p.146). Having derived these two factors, a score for each activity on each item was obtained from the original data (Table 7). These factor scores enabled the 30 activities to be plotted in factor space.

Comparing the results of Figure 6 with those of Figures 1-5, we see in broad terms, that some general features are depicted in both representations, but it is also apparent that there are also significant discrepancies. A distinctive position of nuclear power is seen in both representations, though depicted in different ways; in factor space it is seen in an extreme position in the upper right hand quadrant thus registering more strongly than any other activity on the "unknown" (vertical) dimension, and only matched in its perceived severity on the "dread" (horizontal) dimension by commercial aviation. Of the activities depicted differently in the two representations, alcoholic beverages appears more distinctive viz.-a-viz. others in the Q-analysis representations than in the factor space representation; pesticides in general appear less distinctive in Q-analysis than in the factor space representation (for instance, although it is at a fairly extreme position in the upper right hand quadrant in the latter, it does not appear at all in the Q-analysis which exploits the most extreme perception levels). As another general observation, whereas surgery, alcoholic beverages and swimming are grouped together in the most sensitive of the Q-analytic representations (Figure 5), they are quite separated in factor space (being in the upper right, lower left and lower right quadrants, respectively).

We may now begin to relate these general observations to some of the wider ideas from Q-analysis. These are geared to examine the deeper significance of entities (in this case risky activities) having particular attributes (in this case risk characteristics) in common with others.

Figure 6 Location of risk items within a two-factor space (Fischoff et al 1978)



Considering firstly Figure 2: activities which appear grouped together in a particular cluster at Q = 8 are those which have an identical set of risk characteristics (since only nine basic characteristics have been identified). All other things being equal, people should be indifferent to the activities within any particular group. All other things are not, of course equal, firstly because risk descriptors can take scale values between 1 and 7, not just 0 or 1,*and secondly because there are other dimensions that have not been included in the above analysis. Indeed, in relation to the former point, for all other figures, there is no Q = 8 level, depicting the fact that all activities are instrinsically different from each other (once the unrealistic restriction of a 0 or 1 rating is removed).

As noted above, activities are clustered at a given level whenever they share a given number of features. This may mean that two activities in a given cluster may have a rather different combination of features, but are clustered because of sharing features with some third activity (or, more generally, with some intermediate sequence of activities). Rather than considering such "chaining" an undesirable feature, a Q-analyst would seek insight from it. The interpretations that follow will be concentrated on Figure 5 (because this figure depicts the most strongly held risk features and therefore in one sense, holds the most important information; - see below). This will hopefully convey the main flavour of possible interpretations, and the reader can subsequently consider other Figures in this light.

Part of the justification for the concentration on Figure 5 is as follows: food colouring, food preservatives and nuclear power are clustered together in Figures 2 and 3 - they share the characteristics of "newness", "involuntariness", "delayed" and "known 1". Yet from general knowledge, the risks from food colouring and food preservatives seem to be significantly more acceptable than those from nuclear power. This difference is undoubtedly partly (i) due to factors excluded from the present analysis (for example, issues of nuclear proliferation and the way society is run) but also partly (ii) because risk attributes are "felt" differently (i.e. more extremely) for nuclear power, and partly (iii) because nuclear power registers at higher discrimination levels. The first of these features cannot be overcome given the data on which the present analysis is based. The second can,

* This restriction is relaxed below.

however, be met by using more stringent slicing levels (thus considering results in Figures 4 or 5 rather than 1 or 2). For the third feature, we need to consider the underlining of activities in particular components.

In relation to points (ii) and (iii) in the previous paragraph, the grouping of alcoholic beverages and surgery in Figure 5, level Q = 0, will be considered. The former appears due to the fact that it registers strongly on "not new" and the latter due to the fact that it registers strongly on "chronic". They can be grouped together because swimming has both these attributes and therefore "chains" or "links" them together. One inference we can make from this is that *(ceteris parabus)*, someone who accepts the original data in Table 8 and accepts the risks from swimming should also accept those from surgery and alcoholic beverages (in Q-analysis jargon, risk acceptance would be 0-traffic on the given backcloth). This interpretation can be related to the distinct separation of these activities in Figure 6, where it appeared that these three activities had little in common. It is not being suggested that Figure 6 is "wrong" and Figure 5 is "right", but merely pointing to the fact that very different interpretations of a given data set may be made.

This Q-analytic interpretation can next be applied to a particular activity (again, drawing out contrasts with the factor analytic "results"). Consider (at random?) contraceptives at level Q = 0 in Figure 5 in relation to other activities in that component. Several of these other activities - surgery, mountain climbing, bikes and powermowers - appear in that component for the same reason as contraceptives (because they register strongly on the "chronic" descriptor) and yet even so, they are not close to contraceptives (or in some cases, even to each other) in factor space; (as can be seen by comparing Figures 5 and 6). Note, for instance, the "distance" between mountain climbing and contraceptives. The grouping of alcoholic beverages with swimming (at Q = 0, Figure 5) does not arise because they hold a given descriptor in common with each other (alcoholic beverages registers on "not new" not on "chronic" as noted above) but because alcoholic beverages shares "not new" with swimming and the latter shares "chronic" with contraceptives. Despite this indirect sequence of linking, alcoholic beverages and swimming are not as far apart in factor space as some of the other more directly linked activities.

Thus some risks that appear quite different in factor space may appear quite similar from another viewpoint (and vice versa). Nuclear power is seen to be the most distinctive of all activities, it being "linked" to no other activity at all in Figure 5, where the most sensitive scale values are used. Pesticides do not appear in this figure, because although they registered the same basic nine risk characteristics as nuclear power (see Figure 1) they do not register at such extreme scale values as nuclear power.

EVALUATION

The methodology of Q-analysis may seem rather clumsy at first sight - more so than some simple black-box clustering packages, though less so than some of the further reaches of factor analysis. Increasing familiarity with the method gradually dispels its initial mystique, and all steps are well-defined and therefore computable. Unlike factor analysis, it does not invent "proxies" to represent the original dimensions (this must be a particularly important attribute of the approach), and exposes the internal fabric of the original data set methodically and closely.

Explicit consideration of risk characteristics comes at a relatively late stage in Q-analysis. This may be unexpected, but is not therefore detrimental. What Q-analysis has done is to distil from the original data set some signposts back into the original data. These then lead in turn to risk characteristics. Thus, we are shown where to look, and may consequently see meaning where in Table 8 previously there were just numbers.

The preceding section highlighted some stark differences in results obtained from the two methods of factor analysis and Q-analysis. Although some of these observations may be counterintuitive, it is important to recognise that they pinpoint differences but not necessarily inconsistencies between the two methodologies. It should not be unexpected that the results from each approach are different, since the Q-analysis (Figure 5) focusses only on extreme values, whereas the factor analysis depicts some aggregate over all values. It is an open question as to whether an activity which registers at a medium level (score between 3 and 5) on several factors is more or less acceptable or significant than one that registers extremely on just one (or two). Thus the approaches are complementary, not exclusively competitive. The most important aspect for a researcher is to know the characteristics, strengths and weaknesses of each method and, most importantly, how to choose. Also, as with other techniques, much of the benefit is derived by the analyst in undertaking the analysis; the "results" are just the end-point, and not necessarily more significant than the analysis.

As with factor analysis and other clustering methods, the results are dependent on the factors (risk descriptors) originally identified, thus, everything is qualified by *ceteris parabus*; a different set of risk characteristics might have led to a different pattern of clusters (and distribution in factor space). In future work the present author will explore a richer data set. These richer explorations will also probe some deeper facets of Q-analysis; only the more basic notions have been used in the above explorations.

In an Appendix to this paper, the results of applying a refinement of the Q-analysis algorithm to the data set that has been analysed above is given. This refined method is known as Q-discrimination analysis, and although it produces a finer pattern of clustering than the basic algorithm, does not lend itself to the same type of interpretation as the original method.

REFERENCES

Atkin, R.H. (1974) *Mathematical structure in human affairs.* Heineman.

Atkin, R.H. (1981) *Multidimensional Man.* Penguin.

Duran, B.S. and Odell, P.L. (1974) *Cluster analysis : a survey.* Springer-Verlag.

Everitt, B. (1974) *Cluster analysis.* Heineman.

Fischoff, B., Slovic, P., Lichtenstein, S., Read, S. and Combs, B (1978) How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences 9,* 127-152.

Fischoff, B., Slovic, P. and Lichtenstein, S. (1981) Lay foibles and expert foibles in judgements about risks in T. O'Riordan and R.K. Turner *Progress in Resource Management and Environmental Planning, Vol. 3,* pp. 161-202, Wiley.

Goddard, J. and Kirby, A. (1976) An introduction to factor analysis. CATMOG 7, (Geo. Abstracts, Norwich, England).

Gould, P. (1980) Q-analysis, or a language of structure : an introduction for social scientists, geographers and planners. *International Journal of Man-Machine Studies 12,* 169-199.

Johnson, J.H. (1982) Q-discrimination analysis. *Environment and Planning B, 8,* 419-434.

Kletz, T.A. (1980) Benefits and risks : their assessment in relation to human needs. *I.A.E.A. Bulletin 22,* 2-12.

Lowrance, W.W. (1976) *Of acceptable risk.* Kaufman.

Macgill, S.M. (1982a) An appraisal of Q-analysis (forthcoming Working Paper).

Macgill, S.M. (1982b) A consideration of Johnson's Q-discrimination analysis. Working Paper 343, School of Geography, University of Leeds.

Macgill, S.M. (1982c) Cluster analysis and Q-analysis (forthcoming Working Paper).

Otway, H. (1980) A perspective on risk perception : confessions of a disillusioned analyst.

Slovic, P., Fischoff, B. and Lichtenstein, S. (1980) Facts and fears : understanding perceived risk. In R.C. Schwing and W.A. Albers (eds.) *Societal risk assessment.* Plenum.

Slovic, P., Fischoff, B. and Lichenstein, S. (1981) Characterising perceived risk, to appear in R.W. Kates and C. Hohenemser (eds.) *Technological Hazard management.*

APPENDIX A Q-DISCRIMINATION ANALYSIS OF THE SAME RISK DATA

In this appendix, the results of subjecting the data analysed in the main body of the paper to a so-called Q-discrimination analysis will be given. Q-discrimination analysis is a modification by J.H. Johnson of R.H. Atkin's Q-analysis for use in clustering and discrimination work (see Johnson 1981 and also Macgill 1982b). It is designed to achieve a finer degree of discrimination than that obtained from the basic Q-analysis algorithm. To the authors knowledge, this is the first reported application of the Q-discrimination analysis method at other than a pilot level, and the results are therefore given in the spirit of exploration rather than demonstration.

For any pair of entities (risky activities) which have a given set of risk descriptors in common (and therefore would be paired together within some cluster in a bsic Q-analysis) it can be seen that there are two sets of weights, there being one set for each risk; for activities e_1 and e_2 these weights are given, respectively, in rows 1 and 2 in Table 7. Wherever such weights are relatively similar it would seem reasonable that the activities in question should remain clustered together as would be the case with the basic Q-analysis algorithm. However, where the sets of weights are relatively different there would seem to be reasonable grounds for separating them. (Whether or not the activities would be clustered at any given level would depend on the relative similarity of the weights to the criteria of discrimination set for each hierarchical level.)

In applying the Q-discrimination method to the risk data from Table 7, an intermediate set of calculations is first required, namely a more precise measure of "distance" between each pair of risks. A crude measure of distance is already given in the shared face matrix, since this indicates how many broad descriptors any pair of risks shares (thus for any two risks j and k, cell (jk) in the shared face matrix will be relatively high whenever the risks are "close", and low when they are not). A more precise measure of distance may be found on the basis of the difference in weights $|M_{ij} - M_{ik}|$ between two risks e_j and e_k for each risk descriptor d_i they have in common. More precisely:

$$D_{jk} = \sum_{\substack{i \text{ shared} \\ \text{by } j \text{ and } k}} |M_{ij} - M_{ik}|$$

[illegible] jk [illegible]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	419	653	473	494	627	327	327	813	690	511	627	507	638	738	700	649	184	70	245	490	377	342	614	650	628	354	361	487	576
2	419	0	518	351	305	703	320	321	659	595	358	398	491	531	566	565	484	0	0	205	538	266	260	388	335	630	114	362	454	318
3	653	518	0	464	599	611	67	59	385	321	596	482	543	577	259	276	663	142	223	390	652	358	819	520	337	410	416	688	555	390
4	473	351	464	0	423	418	456	462	617	435	289	386	325	527	391	395	444	635	275	496	352	490	338	448	441	605	269	458	440	337
5	494	305	599	423	0	524	314	301	655	529	335	324	634	354	685	447	712	357	172	333	554	205	84	473	452	331	237	643	271	354
6	627	703	611	418	524	0	64	72	643	485	769	759	455	335	606	367	547	156	45	616	208	258	510	649	492	353	587	705	643	413
7	327	320	67	456	314	64	0	53	196	72	315	569	168	206	145	248	217	509	304	503	200	638	202	352	164	483	357	301	391	476
8	327	321	59	462	301	72	53	0	204	72	330	583	176	214	153	235	225	529	304	511	208	602	215	367	164	472	336	309	399	454
9	813	659	385	617	655	643	196	204	0	469	727	707	552	452	374	308	745	298	261	535	634	454	683	647	446	424	787	760	841	577
10	690	595	321	435	529	485	72	72	469	0	682	482	379	365	435	466	593	233	161	541	376	175	275	593	416	350	576	615	428	412
11	511	358	596	289	335	769	315	330	727	682	0	417	617	651	550	557	732	0	0	354	688	266	392	452	357	599	167	628	371	215
12	627	398	482	386	324	759	569	583	707	482	417	0	689	516	633	574	663	138	103	285	607	441	241	454	534	578	110	651	315	309
13	507	491	543	325	634	455	168	176	552	379	617	689	0	478	349	547	356	107	4	443	381	263	543	493	230	549	436	330	673	408
14	638	531	577	527	354	335	206	214	452	365	651	516	478	0	510	185	571	376	187	497	414	365	409	533	355	333	550	624	620	498
15	738	566	259	391	685	606	145	153	374	435	550	633	349	510	0	402	525	79	32	539	421	333	663	450	169	568	439	617	602	323
16	700	565	276	395	447	367	248	235	308	466	557	574	547	185	402	0	639	388	203	547	355	378	352	491	318	142	434	744	453	414
17	649	484	663	444	712	547	217	225	745	593	732	663	356	571	525	639	0	118	7	501	392	283	638	356	410	673	429	366	716	368
18	184	0	142	635	357	156	509	529	298	233	0	138	107	376	79	388	118	0	765	235	403	620	380	0	200	656	694	120	218	679
19	70	0	223	275	172	45	304	304	261	161	0	103	4	187	32	203	7	765	0	148	214	395	134	0	68	320	310	9	105	332
20	245	205	390	496	333	616	503	511	535	541	354	285	443	497	539	547	501	235	148	0	488	484	255	349	364	791	109	252	440	234
21	490	538	652	352	554	208	200	208	634	376	688	607	381	414	421	355	392	403	214	488	0	379	564	510	317	392	615	593	696	460
22	377	266	358	490	205	258	638	602	454	175	266	441	263	365	333	378	283	620	395	484	379	0	234	363	491	647	280	272	272	401
23	342	260	819	338	84	510	202	215	683	275	392	241	543	409	663	352	638	380	134	255	564	234	0	538	449	263	193	487	332	375
24	614	388	520	448	473	649	352	367	647	593	452	454	493	533	450	491	356	0	0	349	510	363	538	0	356	698	232	400	570	303
25	650	335	337	441	452	492	164	164	446	416	357	534	230	355	169	318	410	200	68	364	317	491	449	356	0	561	265	470	424	362
26	628	630	410	605	331	353	483	472	424	350	599	578	549	333	568	142	673	656	320	791	392	647	263	698	561	0	430	706	521	464
27	354	114	416	269	237	587	357	336	787	576	167	110	436	550	439	434	429	694	310	109	615	280	193	232	265	430	0	377	237	280
28	361	362	688	458	643	705	301	309	760	615	628	651	330	624	617	744	366	120	9	252	593	272	487	400	470	706	377	0	713	518
29	487	454	555	440	271	643	391	399	841	428	371	315	673	620	602	453	716	218	105	440	696	272	332	570	424	521	237	713	0	197
30	576	318	390	337	354	413	476	454	577	412	215	309	408	498	323	414	368	679	332	234	460	401	375	303	362	464	280	518	197	0

where D_{jk} will be called the distance between risk e_j and risk e_k. (This explicit distance measure is not given in Johnson (1981) but is required in order to overcome the weakness of the original Q-discrimination method when weighted scales are long; see Macgill 1982b.) The full matrix $\{D_{jk}\}$ based on the data given in Table 7 is given in Table A1.

On the basis of Table A1 and the shared face matrix, Table 8, it is now possible to calculate, for any risky activity e_j, the relative level of similarity of all other activities e_i, $i = 1, 30$ to e_j. In other words we can envisage a set of terms M_{ij}, depicting the level of similarity of risky activity e_i to risky activity e_j. As noted, the values of M_{ij} can range from 1 to 30 (since in this case we have 30 activities), and the value of 1 is assigned to the risky activity which is most similar to e_j (probably itself), the value of 2 assigned to the next most similar activity to e_j, the value of 3 assigned to the third most similar activity, and so on. Having found the relative levels of similarity of all activities to any chosen activity j, a corresponding procedure is applied to all other activities in turn.

The basis for determining successive levels of similarity of other risky activities to any given activity e_j rests firstly on the shared face matrix, since the number of shared descriptors is the most basic criterion of similarity between activities. For any activities that share the same number of risk descriptors with a given activity j, however, a further basis is required. For this purpose the distance matrix $\{D_{ij}\}$ is inspected in order to determine which activity is closest (the closest being that with the lowest "distance"), and all others successively close. In this way, levels of similarity of all activities to activity j can be determined. Results of this ordering are given in the matrix U_{ij} in Table A2. Only the most significant levels (between 1 and 10) have been determined, since (with the exception of cell $(M_{10,20})$) the other levels turn out to be redundant in terms of clustering the risky activities.

On the basis of Table A2, risky activities can be clustered according to their relative level of similarity. Any pair of activities r and s will be grouped together at a given level (I,J) whenever $(M_{rs}, M_{sr}) < (I,J)$ (ie. whenever the level of similarity of r to s is less than I (or J) and the level of similarity of s to r is less than J (or I respectively)). Level (I,J) will be regarded to be equivalent to level (K,L) whenever

TABLE A2 $\{U_{ij}\}$, the level of similarity of risky activity i to risky activity j

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	*	*	10	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	2	*	*	6	*	*
2	8	1	*	*	9	*	*	*	*	*	2	2	10	*	*	*	10	*	*	*	*	*	3	2	10	*	*	7	5	*
3	*	*	1	*	*	*	*	*	5	*	*	*	*	8	*	*	*	*	*	4	9	*	2	*	*	*	10	*	*	*
4	*	*	*	1	*	*	6	8	*	*	*	*	*	*	*	*	*	7	6	*	*	5	*	*	*	2	*	*	*	3
5	10	*	*	*	1	*	*	*	*	*	*	*	*	*	*	2	*	*	*	*	*	*	*	*	*	*	2	*	*	6
6	5	9	8	*	3	1	*	*	7	4	9	10	5	4	4	7	5	*	*	8	4	*	*	9	7	*	*	5	*	*
7	*	*	*	8	*	*	1	2	*	*	*	*	*	*	*	*	*	3	3	*	*	3	*	*	*	3	*	*	*	*
8	*	*	*	9	*	*	2	1	*	*	*	*	*	*	*	*	*	4	3	*	*	2	*	*	*	4	*	*	9	9
9	*	*	2	*	*	*	*	*	1	*	*	*	*	3	7	9	6	*	*	*	3	*	9	*	*	*	7	*	*	*
10	*	*	*	*	*	9	*	*	*	1	*	*	7	*	10	*	*	*	*	2	*	*	*	*	*	*	5	*	*	*
11	*	2	*	*	10	*	*	*	*	*	1	3	*	*	*	*	*	*	*	*	*	*	5	3	*	*	*	*	4	*
12	*	5	*	*	*	*	10	10	*	*	4	1	*	*	*	*	*	*	*	*	*	9	*	7	*	*	*	*	2	*
13	3	7	7	*	4	2	*	*	6	2	6	9	1	5	2	4	2	*	*	5	5	*	*	8	4	*	*	2	*	*
14	*	*	3	*	*	7	*	*	2	7	*	*	9	1	*	8	*	*	*	*	2	*	6	*	*	*	4	*	7	*
15	7	8	6	*	6	4	*	*	4	3	5	6	3	6	1	6	4	*	*	7	7	*	*	6	3	*	*	4	*	*
16	*	*	*	*	2	8	*	*	*	*	*	*	*	*	8	1	*	*	*	*	*	*	*	*	9	*	3	*	*	8
17	6	6	9	*	7	3	*	*	8	5	8	8	4	7	3	3	1	*	*	6	6	*	*	4	5	*	*	3	*	*
18	*	*	*	*	*	*	8	7	*	*	*	*	*	*	*	*	*	1	2	*	*	10	*	*	*	8	*	*	*	*
19	*	*	*	*	*	*	5	5	*	*	*	*	*	*	*	*	*	2	1	*	*	8	*	*	*	5	*	*	*	*
20	*	*	*	*	*	*	*	*	*	13	*	*	*	*	*	*	*	9	9	1	*	*	*	*	*	*	*	*	*	*
21	9	*	4	5	*	6	*	*	3	8	*	*	8	2	9	*	8	*	*	*	1	*	8	*	8	7	6	10	8	10
22	*	*	*	4	*	*	3	3	*	*	*	*	*	*	*	*	*	6	8	*	*	1	*	*	*	*	*	*	*	2
23	*	10	5	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	1	*	*	*	*	*	3	*
24	*	3	*	*	*	*	*	*	*	*	3	4	*	*	*	*	7	*	*	*	*	*	7	1	*	*	*	8	6	*
25	2	*	*	7	*	*	*	*	*	9	*	*	6	*	6	10	9	*	*	*	*	*	*	*	1	*	*	9	*	7
26	*	*	*	2	*	*	4	4	*	*	*	*	*	*	*	*	*	5	5	*	*	7	*	*	*	1	*	*	*	*
27	*	*	*	*	8	*	*	*	*	*	*	*	*	*	*	*	*	10	10	*	*	*	*	*	*	*	1	*	*	5
28	3	4	10	*	5	5	*	*	9	6	7	7	2	9	5	5	3	*	*	3	8	*	*	5	6	*	9	1	*	*
29	*	*	*	6	*	*	9	9	*	*	*	5	*	*	*	*	*	*	*	*	*	6	4	*	*	*	*	*	1	4
30	*	*	*	3	*	*	7	6	*	*	*	*	*	*	*	*	*	8	7	*	*	4	*	*	*	6	8	*	*	1

I + J = K + L (This rule is needed in order to overcome an incompleteness in the partial order relation given in Figure 6 in Johnson (1981). The clusters generated from Table A2 on the basis of this procedure are given in Figure A1.

By way of interpretation, Figure A1 reveals the relative similarity of different risky activities directly in terms of the weight people assign to them: activities that are most similar in terms of these weights are grouped together at the finest level (level(2,2)) and they gradually fuse with others (through levels (2,3), (2,4) and so on) as the criterionfor similarity is slackened. It is seen on this basis that activity 20 (powermowers) is most distinctive. The next most distinctive activities are 27 (surgery) and 10 (handguns). Nuclear power is less distinctive on this basis being paired with pesticides at all but the lowest level and fusing with many other activities at level (2,6).

Not surprisingly the pattern of clusters given in Figure A1 is clearly more discriminating than the basic Q-analysis partitions in Figure 1 and a considerable refinement over the traditional matching scores method. The latter would calculate a single similarity matrix S_{jk}, say from $$S_{jk} = \sum_{\text{all } i} |M_{ij} - M_{ik}|$$ rather than ordering the similarity on the basis of both (i) the shared face matrix (Table 8) and (ii) the distance matrix based strictly on risk descriptors held in common (Figure A1). Thus Q-discrimination analysis is a refined matching scores method.

In Johnson (1981) it is suggested that the results from the Q-discrimination analysis (Figure A1) should be amalgamated with those from the basic Q-analysis (Figure 1). Such amalgamation in the present context will be left to the reader with the observation that although it will yield a yet finer pattern of clustering than that already given in either Figure 1 or Figure A1 above, it may be difficult to interpret the result as a whole, because different parts of the resulting hierarchy of clusters will have been derived on different assumptions. What it would draw out, by juxtaposing the two sets of results, is the nature of the discrimination achieved by each method. However, once this is done, there may be some careful justification for confining the comparison to only two rather than any of the other methods of cluster analysis. Other methods may be explored by the author in future work.

FIGURE A1 Groupings of risky activities generated by Q-discrimination analysis

9. Level (2,13)

8. Levels (2,8)(3,7)
(4,6)(5,5)

7. Levels (2,7)
and (3,6)
and (4,5)

6. Levels (2,6)
and {3,5}
and [4,4]

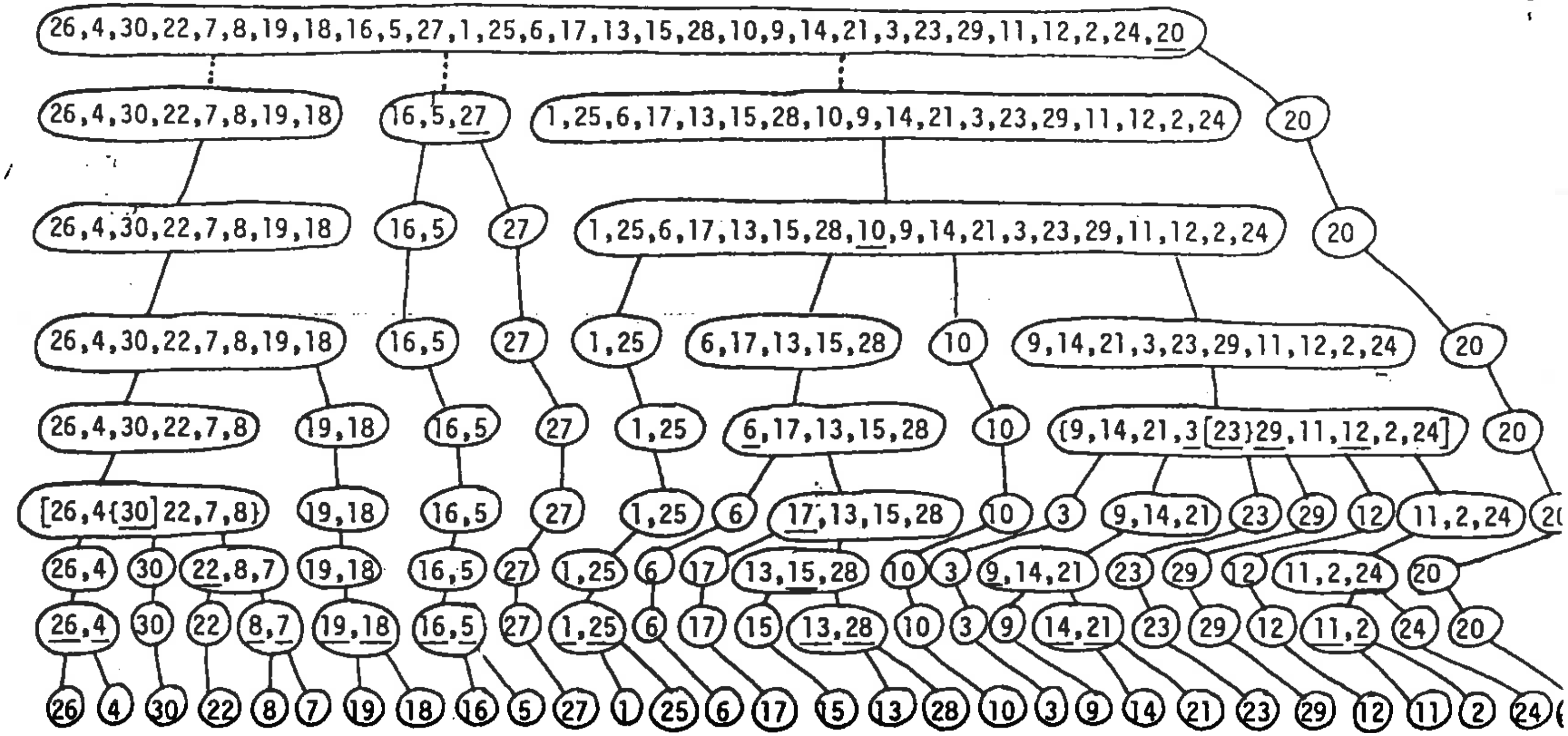
5. Levels {2,5}
and [3,5]

4. Levels {2,4}
and [3,3]

3. Level (2,3)

2. Level (2,2)

1. Levels (1,1)
and (1,2)



Notes for Figure A1

1. For conciseness, activities are identified by number not by name in this Figure; see Table 7 for identification.
2. Underlining denotes the level at which the activity underlined fuses with some other activity.
3. The dots between lines 8 and 9 denote the omission of intermediate levels. Activity 20 does not fuse with others until level (2,13), but all other activities are joined in a single cluster by then.
4. The different parentheses { } [ ] are used to denote the particular criterion that is significant in clustering activities at a particular line.