WORKING PAPER 345

AN APPRAISAL OF Q-ANALYSIS

S.M. MACGILL

WORKING PAPER

School of Geography

University of Leeds

AN APPRAISAL OF Q-ANALYSIS

Contents

December 1982

AN APPRAISAL OF Q-ANALYSIS

1. Introduction

Q-analysis is a relatively new methodology and framework for thought for social scientists. Despite its outward appearance - a rather uninformative title and heavy disguise of idiosyncratic notation and technical jargon - the ideal of the approach is quite simple, namely clear and unambiguous description of those aspects of the real world that are relevant to any particular field of inquiry. This paper is written as an attempt to provide an appraisal of Q-analysis for an audience who may be curious about the approach but as yet unable to penetrate its disguise, and to begin to inquire into the effectiveness of Q-analysis (viz-a-viz other, more traditional methods) for addressing certain types of problem.

The development of Q-analysis has been primarily the work of Atkin (see, for example, Atkin, 1974, 1977). Much of the initial exposition of Q-analysis was difficult and beyond the grasp of many social scientists. This hardly depicted it as an aid to understanding. More recently some more readable expositions of some aspects have appeared (Atkin, 1981; Gould, 1980; Chapman, 1981; Beaumont and Gatrell, 1982). This essay presents a further view about what Q-analysis appears to offer social science, in particular, human geography. As an additional pedagogic review this paper is justified firstly purely selfishly - writing down a latecomer's understanding, as far as it goes, for the sake of clarification. A secondary reason is then to share this understanding with others. Such justification seems appropriate for an approach about which many people have heard, apparently enthusiastic claims have been made and further developments are proceeding apace (a spate of articles in the *International Journal of Man-Machine Studies* and in *Environment and Planning B*, for instance) but many remain unaware about what the approach actually offers. A third reason is to seek to break down the insularity of Q-analysis from more traditional methods. This needs to be done in order to establish a position from which to choose whether Q-analysis is an appropriate alternative or additional approach in any particular area of 'problem solving'. A fourth is to try to be more comprehensive in covering the different attributes of Q-analysis than the other pedagogic reviews cited above.

It is difficult to convey the essence of the approach in only a few words, but the following are summarised as what the present author interprets as the main facets of the approach.

(i) an explicit philosophy. This is based on two underlying principles, namely that hard description must come first in any area of 'problem solving', and that there are latent 'structures' in all fields of inquiry which can often usefully be revealed. Hard description is a non-trivial requirement, and is achieved by observing strict rules of set membership. Structure arises via the relations that exist between the sets that are thereby defined.

<u>and</u> (ii) a general theory. This is that the multidimensional structures latent in social systems play a role in constraining or influencing 'activity' (what Atkin calls traffic; it could be monetary expenditure, physiological stress, disease, or whatever) analogous to the way in which 3-dimensional space in the 'real world' constrains (in terms of available 'routes') or influences (in terms of forces) physical movement or activity (vehicles, particles, people, etc.). Such activity or 'traffic' may also exert reciprocal influences on the structure.

<u>and</u> (iii) an algorithm and a set of indicators for representing all or part of the aforementioned structures and traffic in a numerical and, if desired, graphical way. Aspects of global structure are depicted, for instance, in the standard listing of so-called q-components and in the obstruction vector; aspects of local structure are found by examining individual components and looking at so-called 'stars' and 'chains of connection' (spaces and tunnels in the structure), and 'eccentricities'. Also, more recently, there are apparently algorithms for depicting traffic patterns.

<u>and</u> (iv) a language for representing 'structure' and 'traffic' and related concepts, and for exploring further ideas. This draws (heavily to a social scientist, but lightly to a professional mathematician) on topological notions, with Atkin's (and others) own modifications and additions. The language is still developing. A sparing use of formal mathematical notation will be made below.

In summarising each of these four features as integral aspects of Q-analysis, it is hopefully possible to begin to appreciate why Q-analysis has captured the imagination of some social scientists, left others cold, with yet others somewhere in between. The four features are intimately interrelated (for instance the language used and theory generated deliberately reflect the underlying philosophy) though to avoid having to talk about everything at once they will be discussed sequentially below. The philosophical foundations of the approach will be considered first (in section 2), then the standard algorithm (in section 3). The immediate practical utility of the type of results obtained from applying the Q-analysis algorithm will be discussed in section 4, and the insights stemming from the theory of the approach will be invoked to give a deeper interpretation of these results in section 5. Qualitative features about the language of Q-analysis are discussed in section 6 before some general conclusions are drawn in section 7. Various relationships between Q-analysis and probably more familiar existing approaches will be indicated as they arise.

A warning for the impatient: the essence of Q-analysis is rather more modest than that of many more widely recognised mathematical and statistical approaches, being primarily concerned with description rather than, say, explanation, prediction or optimisation. The strength of the approach may lie in rather than in spite of this modesty. This is because its descriptive powers may force a degree of rigour not previously contemplated. Thus Q-analysis has not evolved as a black-box method which can be applied to data which are somehow 'just there' and produce conspicuous and obviously useful results, but as an approach within which the basic nature of available information or data is to be critically examined.* In the author's view it may yet be some time until the true utility of this approach can be evaluated (in the light of extensive applications), therefore the emphasis in the present paper is on exploring rather than judging the approach.

*The examination may remain descriptive, or may lead to the formation of theories and hypotheses about the aspect of the "real world" being represented.

2. Q-analysis: philosophical elements

As suggested above, the underlying philosophy of Q-analysis may encourage, in terms of description and data categorisation, a degree of rigour and precision not otherwise obtained. In terms revealing hitherto unrecognised structure in various fields of inquiry, Q-analysis seems to provide an operational framework for a certain type of structuralist inquiry. Whether or not this is considered appropriate may depend in part on one's own assessment of the currently fashionable structuralist debate in various disciplines. It is considered here to be at least worth exploring.

In this section three propositions and a corollary will be discussed with the aim of capturing the underlying foundations of Q-analysis. Although presented as a starting point for the approach, the propositions of this section are not starting blocks which should be left too hastily, as they carry some profound implications. A reader finding this to be unfamiliar material might give them more than passing attention, whether or not the further aspects of Q-analysis are pursued. Conversely, anyone for whom the material is familiar can anticipate the further aspects of the approach (in later sections) as a (unique?) means of pursuing the propositions in an operational framework. Due caution is warranted of published articles purporting to invoke the framework of Q-analysis, but in fact only implementing its algorithm (ie. one aspect of it, and may be not the most important aspect).

As an introduction to the first proposition it will be asserted that clear and unambiguous data may be described as being 'hard'. Those that are fuzzy, value-laden or otherwise ambiguous may correspondingly be called 'soft'. It is believed here that 'hard' data are inherently worthier than soft data in terms of an understanding of the real world.

Proposition I (Atkin, 1974) Hard data are the result of observing set membership. (By default, then, all other data or information, whether or not purporting to be hard or 'scientific' - whatever that is taken to mean - will here be called soft data.)

In appraising this proposition, it is first necessary to understand what is meant by set membership. A set is a collection of objects or entities for which it is possible to answer either 'yes' or 'no' to the key question 'Does ... belong to the collection?' (where ... can be anything at all). The conventional ways to define a set are either (i) to list all elements or (ii) to specify some property which all elements must possess. In the latter case, any article or entity possessing that property must belong to the set. Both ways enable an unambiguous answer to the key question to be given. The lack of ambiguity will be genuine of course, only if this answer is agreed by anyone involved with the piece of work in which the sets arise - whether author, client or reader. Some examples of sets are listed in Figure 1, indicating whether they are of type (i) or type (ii), and posing the key question about possible members. See, for example, Stewart (1975) for further exposition about set membership.

Identification of set membership is not in itself a revolutionary concept. All data gathering exercises and modelling and statistical work rests (or at least, should rest) implicitly on well defined sets (sets of zones, sets of characteristics ...). Intelligible data presentation and mathematical and statistical notation would be impossible if this were not so.* The point to recognise, then, is not the novelty of the concept but its generality, its applicability in hitherto uncontemplated contexts and its ability to supply a definition of 'hard', thus quietly opening a door to the further riches which will be discussed below.

---

* In many cases the sets are of type (i); it may be a revealing exercise in certain contexts to see whether the same sets would arise under the more exacting type (ii) rule.

Figure 1. Some sets

Type			Possible members x	Does x belong to the set?
Set 1	(i)	{cat, pencil, jelly, thunderstorm}	cat dog	Yes No
Set 2	(ii)	{whole numbers}	1 2 101 -3.3	Yes Yes Yes No
Set 3	(ii)	{types of fruit}	apple tomato banana	Yes Yes[1] Yes
Set 4	(i)	{Christaller, Losh, Harton, Haggett, Chorley}	Berry Haggett	No Yes
Set 5	(ii)	{objects topologically equivalent to a solid doughnut}[2]	teacup teapot milk jug ring	Yes No Yes Yes
Set 6	(ii)	{the objects currently on my desk}	telephone paper cat	Yes Yes No
Set 7	(ii)	{local authority district councils in the UK}	Derwentside Dunfermline Duddan	Yes Yes No
Set 8	(ii)	{types of vegetable}	bean tomato carrot banana	Yes Yes[1] Yes No

Notes on Figure 1:

1 It does not matter that "tomato" has been assigned to two different sets. It is probably more difficult to specify entities and phenomena that are members of only one set than it is to specify those that are members of many.

2 See, for example, Stewart (1975, p. 146)

Figure 2. Some soft statements

Manchester is a successful city.

The New Statesman is a socialist newspaper.

10^{-6} is an acceptable level of risk.

The social sciences are not scientific.

All reasonably practicable measures will be taken.

Figure 3.

1. The relationship between distance and time (speed).
2. The relationship between a line on the ground and mark on a calibrated scale (length).
3. The relationship between people and shopping zones.
4. The relationship between people and activities,
5. The relationship between exported commodities and countries.
6. The relationship between countries and regions.
7. The relationship between students and option courses.
8. The relationship between industries and industries.

Figure 4a. Tree-type inter-set relations (partitions) between the sets {X,Y} and {A,B,C,D,E}

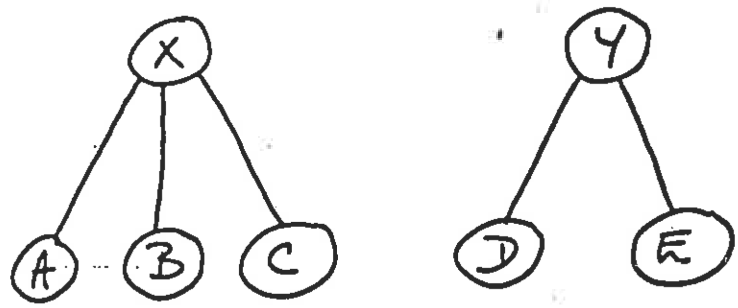


Figure 4b. More general inter-set relations between the sets {X,Y} and {A,B,C,D,E}

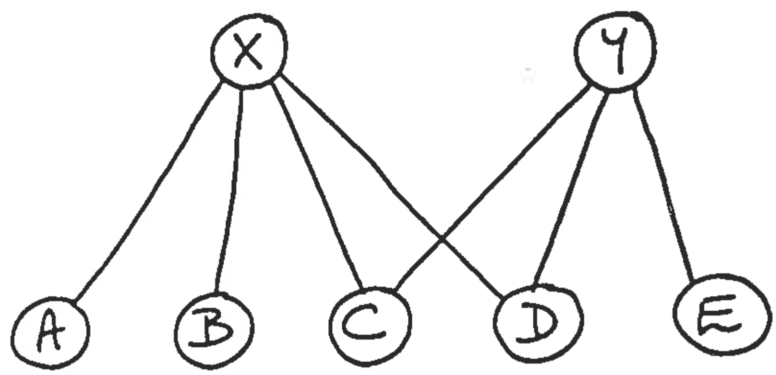


Figure 4c. A latent structure derived from Figure 4b (the two pieces may be joined by glueing at D and C)

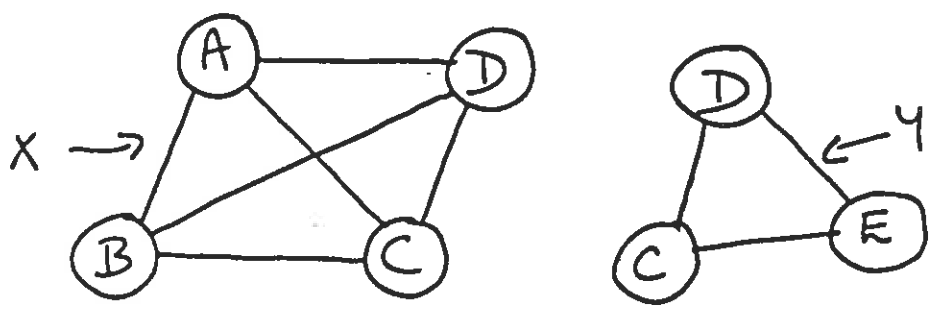
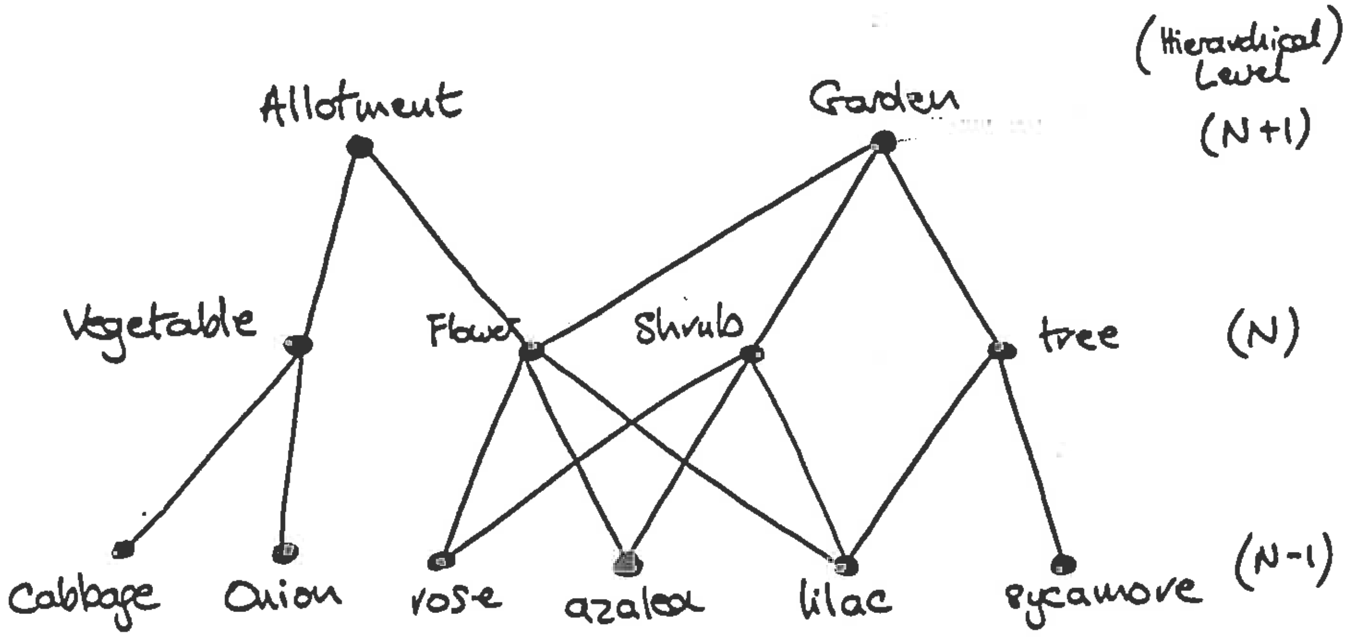


Figure 5. Some strict hierarchies (after Atkin 1981 and Chapman 1981)



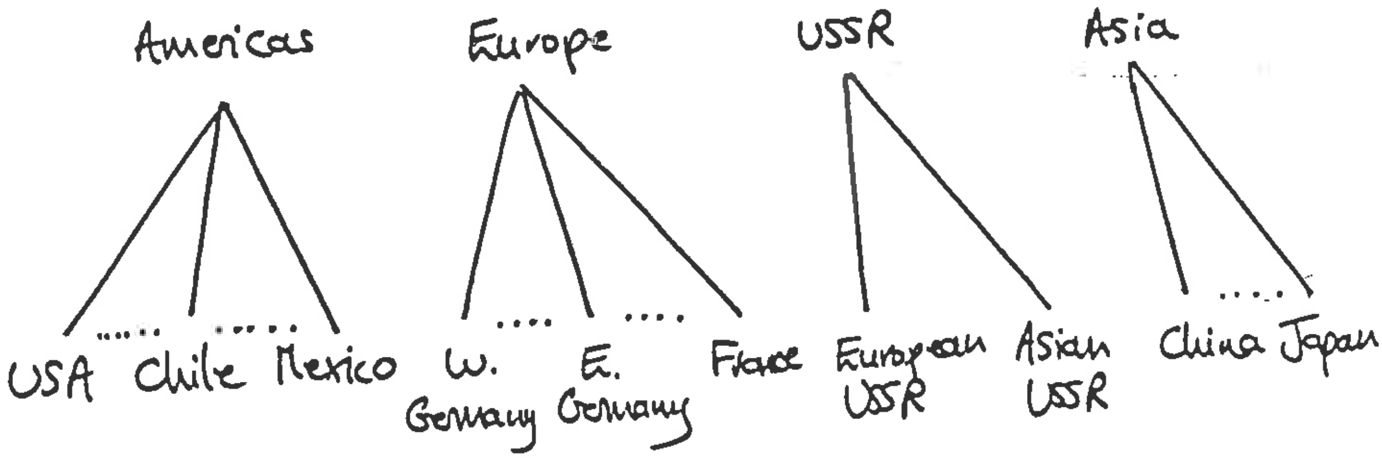
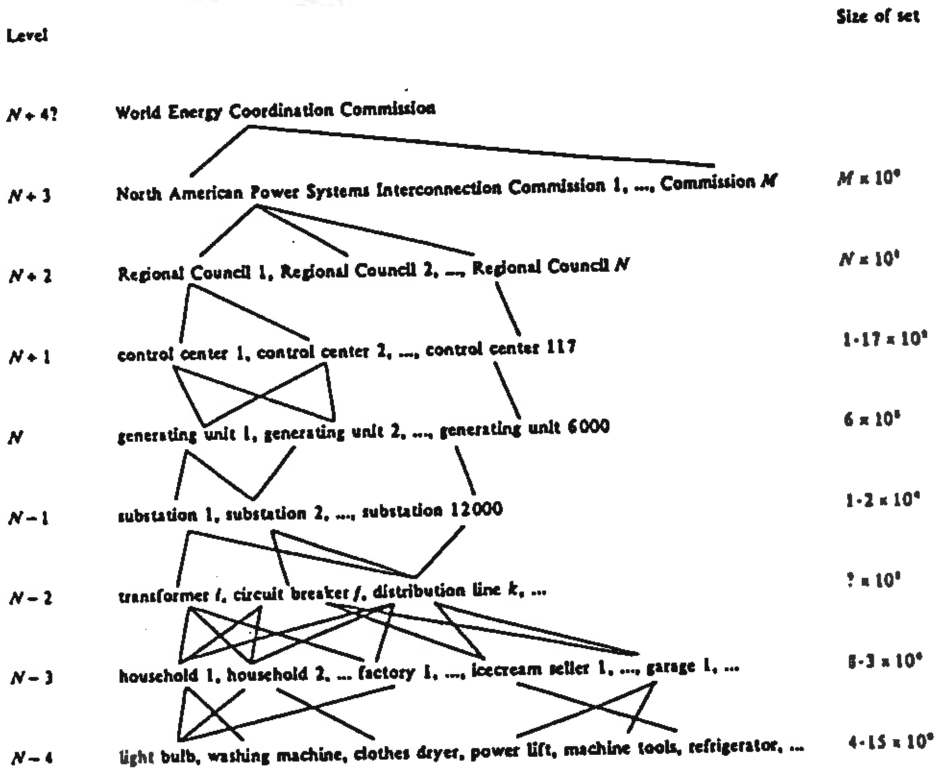
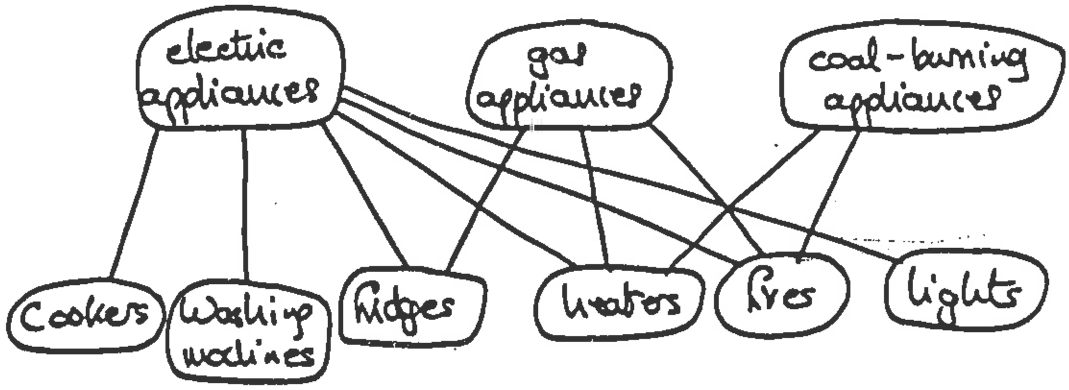


Figure 5 (cont'd) A further strict hierarchy (Gould 1981): the hierarchy of cover sets of physical objects in the electrical power restoration problem

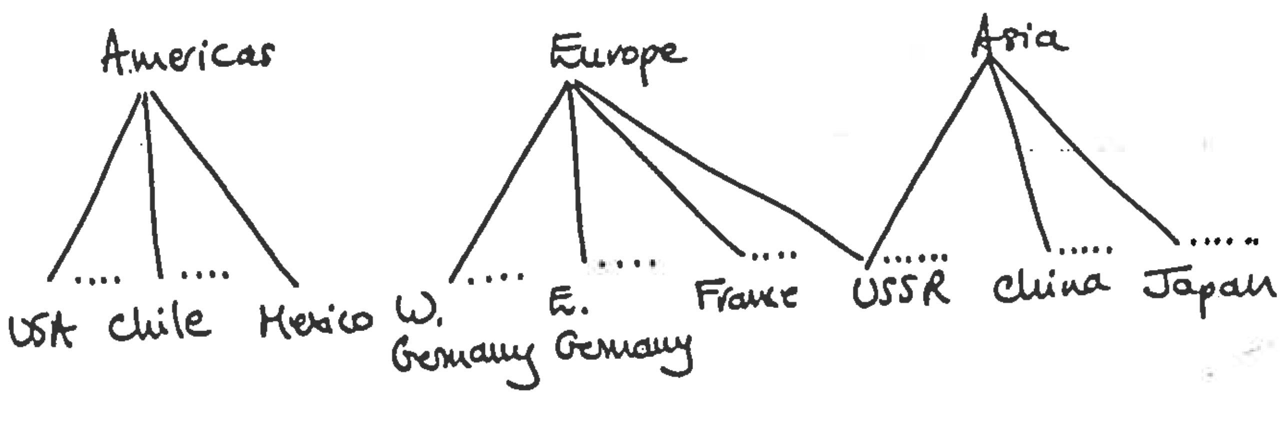


<u>Figure 6</u>. Two hierarchies which are incorrect in strict set-theoretic terms

6a



6b



In many contexts the identification of elements of well-defined sets** can be very demanding, though to convince others on this point may typically require examples more obscure and contentious than those in Figure 1, and ultimately examples of one's own making. It is due to the rigour demanded, and agreement over whether certain entities belong to certain sets, that set membership defines a criterion of hardness. By way of counter-illustration, a number of 'soft' statements are given in Figure 2. The (type (ii)) sets apparently being alluded to in these statements are {successful cities}, {socialist newspapers}, {acceptable risks}, {scientific inquiries}, {reasonably practical measures}. These, however, are bogus sets; without further qualification, they do not refer to unambiguous properties. (When is an urban area a city? What is meant by a *successful* city? What is meant by a socialist paper? and so on.)*

Gould (1980) makes the point that the non-trivial nature of defining set membership may encourage careful thought about which sets to identify and which not. In this connection he also pursues Atkin's (1974, 1981) argument (indicated in a different context by Stewart, 1975, p. 43) that a number of difficulties that are encountered in social sciences (fuzzy description, sterile debate, blind alleys, unrecognised semantic difficulties) may be traced to the fact that no attempt is made to identify (either implicitly or explicitly) which sets are really being referred to. This point will come over more forcefully under proposition II, but even at this stage it may be recognised that (paraphrasing Gould) the lack of attempted identification of set membership allows people to talk without making clear, either to themselves or to their audience, what they are talking about (though this does not apparently, stop them from talking). Chomsky's antagonism towards pundits posing as experts in fields where there is as yet no hard expertise would appear to draw on just such an observation (see Lyons,1977) It is also a general characteristic of common speech, and enables a greater degree of freedom of (sterile?) debate and difference of opinion tha n might be the case in a more rigorous set-up: something some politicians have an instinctive feel about?

---

*The answer is often determined in terms of sets at some finer level of resolution:see under proposition III

** under rule (ii)

The pitfalls of omitting to establish relevant well-defined sets are not confined to traditionally non-quantitative fields of study. On the contrary it is all too easy to assume that the well-defined sets that are relevant to any particular area of problem solving have been identified when, in fact, this is not so. There appear to have been a number of applications of the Q-analysis algorithm to data sets without, perhaps, paying due regard to this point.

Another situation in which set definition can prove telling is in the context of hierarchically related entities (see below). In general however, strict set membership criteria are likely to prove most exacting in traditionally non-quantitative fields of inquiry. The possibility of using such criteria to define meanings of words is not meant to suggest that such meanings should be fixed for all time; merely to recognise that language must have some stability in order to be comprehendable and therefore to capture the nature of the stability, or expose the degree of diversity (and hence lack of meaning) in given contexts.

Although identification of well-defined sets may demand a degree of rigour not previously contemplated, in itself it is a purely taxonomic operation. It is necessary to go further than this in pursuing the foundations of Q-analysis.

Proposition II. Any intrinsic interest in members of a given set arises through their relations with members of other sets.

A structuralist might take the task of relating a given entity to others to be a self-evident necessity, and possibly even too intuitively obvious a requirement to mention. However, it is here superimposed on the idea of set membership. Moreover much of the power of Q-analysis lies in its ability to make concrete much that has hitherto been sensed largely at the level of intuition.

Words have meaning because they may be related to other words or to experiences, ie. elements of given sets may be related to elements of other sets. (As an aside at this point it may be observed that individuals relate given words to their own experiences - which may not be the same as those of others - and hence derive different meanings*. Similarly, precise translation of words from one language

*Whether or not a significant number of readers can understand this paper will depend on whether the writer has been successful in using words with sufficiently well 'accepted meanings'. There has certainly been an attempt to do so.

to another is often impossible because the wider set of words to which they relate in each language are not the same. However, the basic operation in each case - comprehension and translation - may be represented as one of relating members of sets to other sets.) Events or phenomena similarly have meaning because they may be related to other events or phenomena. (A similar aside may be made here to note that 'meanings' are different for different cultures - whether the cultural distinction is related to professions, nationalities, sects or other social groups - because given events may be associated with a variety of different other events.) All explanation involves relating things to things, or, more rigorously, sets of things to sets of things. Substances and materials in the biosphere become significant as resources only in as much as they relate to man's utility (sets of needs and wants). Measurement involves relating (sets of) observations to (sets of) numbers. (As another aside we may note that the depiction of measurement in this way - invoking set membership and inter-set relations - may further suggest that the roots of Q-analysis are potentially more fundamental than those of traditional quantitative methods. In Figure 3 a small selection of the many inter-set relations that have been of traditional concern within geography are listed.

Some of these relations are so-called one-to-one relations (eg. measurement, speed ...). In such cases the representation in terms of sets may appear to be unnecessarily laborious. It does, however, illustrate the point that such one-to-one relations are a special case of something more general (more highly joined relations). An intermediate stage between one-to-one relations and varying degrees of overlap are so-called partitions (many-to-one relations ...). See Figure 4. In this case, the corresponding diagrammatic representation depicts the classic tree structure or dendogram.

It can be seen that in more realistic examples, the explicit identification of relations between sets can readily expose a natural web of linkages which may be said to form a structure. Exploration of the significance of such structure is the key concern of the algorithm and theories of Q-analysis. In view of the fact that this structure can often be rich and complex, it may seem surprising to realise that in the countless contexts which effectively involve dealing with sets and relations between them, the preference in social science has traditionally been to suppress this structure and to search for or enforce one-to-one relations, or, at best, partitions, rather than allowing the naturally richer relations. Thus there is a tendency towards the rule that if A relates (or, more crudely, belongs) to X it cannot also relate (or belong) to Y. Network analysis (and associated graph-based methods, see, for example Roberts, 1976) may be mentioned as notable exceptions, as networks often exhibit characteristically rich structures. However, these are essentially one dimensional structures, composed of a number (perhaps even several hundred in some cases) of one dimensional links between nodes. One of the most distinctive features of Q-analysis lies in its exposure of multi-dimensional linkages between entities. Thus a traditional network analysis of the relation in Figure 4b would be based directly on the (one dimensional) links already shown; a Q-analysis would be based instead on the derived structure which is indicated in Figure 4c. Note that the linkage between X and Y in Figure 4c is immediately seen to be two dimensional (rather a different concept than the two one kimensional links given in Figure 4b). In addition to its ability to expose multi-dimensional linkages (through revealin hitherto unrecognised structures) further distinctive features of Q-analysis lie in its searching inquiry into what sets are relevant and how they are hierarchically inter-related, and the particularly distinctive separation of structure and traffic (whereas in network analysis, traffic and structure are typically one; because links are causal relations or active channels of movement or interaction).

The general preference for one-to-one relations and partitions amongst many social scientists may be due to at least three factors. First to the fact that the traditional mathematical methods which they use are based predominantly on functions (one-to-one relations) rather than more general relations (see especially Gould (1980) on this point). This whole traditional school of mathematics is ill-equipped to handle the latter. (Traditionalists wishing to be updated can see for example Stewart, 1975.) Secondly the narrower one-to-one and partitional schemes may appear better defined, simply by imposing the 'rule' referred to above that if A belongs to X it cannot also belong to Y. Thus any overlap is suppressed seemingly as if it is an embarrasing irrelevance or unmanageable feature rather than a true characteristic that needs to be represented.

Thirdly it may arise out of sheer ignorance: it may not be recognised that overlap exists. It may be noted that each of these explanations potentially exhibits some degree of corruptions of the 'real world'. (i) the mapping of real-world phenomena onto functions rather than relations is bound to restrict the way patterns and processes can be depicted. Although some compression of reality is inevitably involved in modelling, it is far from obvious that this should always involve a priori exclusion of relations and subserviency only to traditional mathematical forms. The 'fitting' of data to straight lines through regression techniques is an extreme case in point; see also Gould (1980). (ii) the same author, following an anonymous sociologist rather disdainfully refers to the second 'explanation' as a manifestation of unduly thorough potty training, sacrificing the messiness of overlap to the neatness and convenience of a single categorisation: a natural (but shortsighted) human trait for least effort. Thus (Gould, 1981a,b) rocks have been unnaturally partitioned into unconnected categories, jobs into distinct types, and disciplines into separated trenches in each case to the detriment of an advancement of knowledge (since natural overlap has been suppressed. The data is apparently not allowed to speak for itself! (iii) if the second was deliberately self-imposed ignorance, the third, must be externally imposed ignorance or sheer naivity; in this case it is not suspected that there is any richness in the structure to speak of. A miner may see a decision about the development of a new coalfield purely in terms of jobs (rather than, say, wider environmental concerns as well: an ecologist may, on the other hand, concentrate only on the latter): a pragmatist may view any theoretical analysis as unconnected to the real world: though others may see it as being of considerable relevance: remarks intended to be innocuous may be taken as insults because the person to whom they are addressed locates them in different partitions than the person who utters them - sterile or offensive exchanges may be the result i.e. they are interpreted differently by the speaker and receiver. There may be many more such illustrations. One thing they suggest is that the analyst may benefit from exposing what is known of inter-set relations, and view the resulting structure as a whole, rather than from the perspective of a single element. In the field of chess Atkin (1974) argues that the grand-master has mastered the art of doing this, whereas the lesser player has not. In another context Rushdie's hero (1981) is frustrated by being denied the opportunity of doing this - a theme portrayed in different ways in many other novels.

The first two propositions appear to capture and expose a curious paradox that exists within much social science inquiry. On the one hand ignoring the rigours demanded by basic rules of set membership may allow a profusion of flabby rhetoric and soft generalisations with relatively little in the way of precise definition or agreed meaning. On the other, ignorance of overlap (for whatever reason) fosters a pre-occupation with seemingly rigorous discrete categorisation and the strict but unjustifiably confident ordering of 'data' into pre-fashioned moulds, an ordering that a more fundamental view may reveal to be false. Chomsky's bogus experts and undeservedly 'successful' politicians seem to be among those who have perfected the art of tapping this paradox. Reichmann (1978) notes a similar paradox in statistics, plagued by soft interpretation of originally 'hard' data and statistical measures, compounded by people claiming to understand statistics when they do not (though they probably would not make the same claim of other languages). 'Real' statisticians do not help by superimposing very precise meanings on words already in everyday use. Fischoff et al. (1981) expose so-called risk experts in a corresponding way; trained scientists who might read very carefully beyond the available data (and hence hard analysis) in their own areas of specialisation (heavy gas flow, say), at times seem to feel no restraint in making (soft) statements about societal acceptability of their technologies. The foundations of Q-analysis may herald an approach that can combine precision (via its adherence to the definition of hardness) with generality (via its universally applicable ways of seeking structure through interset relations) and thus provide one way of breaking through the paradox that has been such a torment to those who are not blind to its existence.

Proposition III. Sets may be hierarchically arranged: members of sets at a given level may be defined in terms of sets at a lower level.

Figure 5 gives some examples of hierarchically composed sets. The very act of identifying a set as a collection of entities (members) gives rise to a natural hierarchy in which the name of the set is necessarily at a higher (coarser) level of resolution than its members. Moreover, at this higher level, the set name is itself likely to be a member of another set. Thus Atkin (1978) notes that "if X is a set at level N then the subsets of X (groups of members of X) qualify as elements of a set X' at level (N + 1)". In an analogous way particular groups of elements of sets at level N - 1 may define elements of sets at level N*. In the light of these concepts it can be suggested that the 'hardness' of the specification of type (ii) sets at any given level N (cf. propostion I) rests either on some suitable specification at level N - 1 (and hence, by induction, possibly also at levels N - 2, N - 3, N - 4, ...), or on indisputable evidence (those greek words for which word and meaning were one, or objectively observable entities). Thus to be rigorously defined, hierarchies must be defined from the bottom upwards. Moreover, in order to define hierarchy unambiguously Atkin has argued that an element at a given level (level N, say) must relate wholly to at least one and perhaps several elements at a higher level (level N + 1). In other words it must always be clear that the higher level can cover the level immediately below it. If this is not the case, the hierarchy is ambiguous (incorrect). In Figure 6 an example of a (strictly) incorrect hierarchy is given.**

The notion of hierarchy is, of course, much more widely applicable than the strict cover set hierarchy discussed in the previous paragraph and illustrated in Figure 5. The idea of hierarchy is very important in all social science. Indeed it is usually important to be aware (though it is not always explained) of how any particular entity fits into a broader hierarchy, and how aggregation and movement between

---

* The technical term for all possible combinations of elements from a given set is the power set. Thus the elements defined at level N are elements of the power set from level N-1. See Atkin (1981a) for an attractive account of these hierarchical notions.

** We cannot answer either 'yes' or 'no' to the question "Is the USSR part of Europe?"

levels within the hierarcy can occur. This is often not clarified to the extent that it might be in many reported studies. The third proposition is intended in the first instance to convey the particular type of hierarchy that is embodied in the notion of well-defined set membership and hence particularly close to the first proposition. In Figure 6 where a more general hierarchy is given; its correct form is of the classic 'tree' structure reflecting partitioned sets (cf. discussion under proposition II), though elsewhere other hierarchical 'shapes' are also possible. Note that the word 'hierarchy' is not intended to suggest that sets at higher levels are inherently 'better' than others lower down; only that they may be a cover (and are always less detailed than) the sets beneath them.

One of the purposes of establishing well-defined hierarchies in the context of practical applications is that whatever level in the hierarchy some 'problem area' is identified, the necessary dimensions of it that need to be addressed in analysing the problem are likely to be found at some lower level (eg. towns at level N; social-economic and physical characteristics at level N - 1, eg. risks at level N; risk dimensions at level N - 1). In itself, this does not entirely solve the problem of how many characteristics at level N - 1 need to be identified. (A similar problem arises with cluster and factor analyses; results necessarily depend as much on the entities and descriptors than happen to have been identified as on the 'real world' that is purportedly being represented (see Everitt 1980) Hierarchical principles thus help in distinguishing aspects of relevance; moreover, if hierarchies are not self evident, the validity of analyses which proceed as if they are may be questioned.**

These three propositions in the author's view provide the modest but powerful building blocks of Q-analysis. Having introduced a working definition of hard data (under proposition I), and argued that it is not sets per se but the relations between them that are important (under propositions II and III), an important corollary of these propositions is in order before proceeding to 'operationalise' them.

---

*Although the original hierarchy given in figure 6 does not meet the conditions for a strict cover-set hierarchy c.f. propositions I and III it may nevertheless represent a useful inter-set relation c.f.proposition II, for some purposes.....?

** The usefulness and exacting nature of the strict set membership hierarch y of Q-analysis as a framework for data categorisation is clearly brought out in some of the most detailed applications of the Q-analysis framework to date (Atkin 1976-7 Gould and Johnson 1978). It may be appropriate for others to invoke the framework for data categorisation in their own fields of inquiry, whether on not the further aspects of the approach are pursued. In the writer's view however, the most telling contributions of Q-analysis as an intellectual tool do not lie purely in its powers of static taxonomy.

Corollary

The only 'hard' questions that may be asked or deductions that may be made are those that refer to sets or relations between sets that have already been defined under the constraint of the above propositions*.

The importance of this corollary can be illustrated firstly (directly in the spirit of Atkin (1974)) by considering the following description of an electricity distribution system of a region in which there is a public electricity utility and several other activities (industries, commercial activities and households, say): all activities use electricity; some have their own generators. "The public electricity utility supplies electricity to all activities in the region who do not supply their own".

Taken at face value this may seem a quite reasonable description of where each sector gets it electricity from. However, those familiar with the type of paradoxes that Russell (1946) was to expose (and therefore perhaps sensitive to the need to avoid such statements) will observe that the description cannot even stand up to the simple question: "Where does the public utility get its electricity from?". We cannot answer that it supplies its own because activities who supply their own are not supplied by the electricity utility. We cannot answer that it does not supply its own, for then it would get its electricity from the public utility.

The problem can be traced to the fact that both the initial statement and the question asked were (in present terms) 'soft'. A question about where activities get their electricity from calls for the identification of a set of activities who supply their own, and another set who gets supplied by the public utility. The public utility would probably be in both sets. The answer to the question would then be self-evident. Thus hardening up the initial statement avoids later embarrassment. This way of hardening up invokes only propositions I and II. It is wise to be more rigorous and also invoke proposition III, noting that there was a natural hierarchical

*A novel extension to this line of reasoning can be found in Atkin's (1981) recent notion of a surprise being the answer to a question that has not yet been put.

relationship between the public utility and the other activities. The very fact that the public utility supplies electricity to activities other than itself earns it a place at a different hierarchical level with respect to current terms of reference than the other activities. Being a public utility it has a distinctive relationship with activities to which it supplies electricity. It was a confusion of hierarchical level in the use of the word 'activity' and not only the lack of identification of sets that caused particularly acute embarrassment: answers could be contradicted and not merely remain unverified. Paradoxes can be generated by deliberately exploiting such a confusion of hierarchical levels and perhaps social scientists unconsciously generate corresponding confusion by inadvertently doing so. Atkin is very keen to warn against this (see also Chapman, 1981; Russell, 1946). Where strict natural hierarchies exist they must be respected (cf. proposition III). An entity may be both an element of a set and a cover for elements in a set at a lower hierarchical level. If making a statement about it, the capacity in which it is being addressed (the set to which it is assumed to belong) must first be ascertained.

There are many other examples of warnings of questions asked and deductions made that do not relate to sets or relations that have been defined. Atkin (1974, 1981) and Gould (1981) cite a number of cases where the progress of knowledge seems to have been severely hampered due to a pre-occupation with such 'soft' deductions as: 'motion is the realisation of a body's potential' and 'Phlogisten is the essence of combustion', rather than hard descriptions (of motion in terms of distance and time and of combustion in terms of oxygen and heat) invoking sets and relations between sets (citing the cases of the advance of Galileo over Aristotle and of Lavoisier over his predecessors!). Atkin warns of further pre-occupation in present day social science with such statements as 'education is the realisation of the mind's potential'.

Attempts to defuse other minefields of bogus deductions have been made by the authors of the many excellent books on the theme 'damn lies and statistics' and here we can reinterpret the bones of contention in the light of the above three propositions. For example (after Reichmann):

Statement: 'Accident statistics reveal that there are many more injuries to intoxicated pedestrians than there are to intoxicated drivers.'

Deductions: It is safer to be a driver
If you become inebriated, all you need to do is to borrow someone's car and you thereby increase your chances of survival.

It may be observed that the deductions do not relate to the sets referred to in the original statement (the statement referring to drivers and pedestrians who have had accidents, and the deductions to all drivers and pedestrians) quite apart from the type of injury incurred. The fault is entirely a matter of defective phraseology involved in the attempted compression of ideas which, perhaps, are not compressible, and by implication ambiguity over which sets and relations are being referred to.

It should not be surprising that books expounding these corruptions are amusing. Surely the stuff of many jokes is to present just enough information for the 'audience' to conjure an image of one entity related to a particular set of things or circumstances only for the comedian to present a punchline that relates to a completely different set (or different hierarchical level). To be told that "In London a man is knocked over every minute, on average", may at first sight conjure up an image of unbearable suffering. However, it may be turned into a (albeit rather tame) joke with the comedian's comment that it is about time he got out of the way. There may be lessons here: how much social science involves fascination with inappropriate sets of relationships but is as yet undertaken with unfailing seriousness? (see Also Atkin, 1981, on laughter and tears).

In discussing the above propositions and corollory as the foundations of Q-analysis, it is not intended to suggest that they are unique to the approach. Indeed they have probably all individually been given more or less attention, in some way, especially given the recent popularity of philosophical approaches in the neo-quantitative age in geography (spurred on by the main Research Councils). What is refreshing in the author's view is the way they are explicity shown to be interrelated and, in subsequent sections, explicitly used within an operational framework and invoked at every stage of its use. This is what raises Q-analysis to being a framework for thought (a fundamental approach to data, philosophy and scientific inquirymore than amethod) and potentially more intellectually stimulating than "user ready" methods.

3. The standard Q-analysis algorithm

The working guidelines that are suggested by the above propositions are that, as far as possible, data and information should be depicted in terms of sets and relations between them. Nothing should be imposed on the data that is not already there (by forcing data into sets to which it does not belong) and care must be taken to define which sets and relations are of interest to any study. Due regard for hierarchies should be implicit in all this. Thus as much organisation as can be justified must be imposed on the data, but no more.

The standard Q-analysis algorithm is a mechanism for exposing certain aspects of a latent structure that exists given any two sets that are related to each other.

In this section of the paper, the standard Q-analysis algorithm will be applied in the context of retailing activity. We will assume that there is a set {X} of seven contiguous zones and a set {A} of twelve retailing activities which are each located in one or more of these zones. The first proposition (pertaining to set identification has already been invoked in the preceding description of the system (the sets {X} and {A}). In more realistic applications it may be a sizeable job in itself to define a zoning system and decide which activities are of interest and at what level of resolution they are to be studied. (For the same problem in a different context see Johnson and Chapman (1979) and Gould and Johnson (1978), Atkin (1977), Atkin (1975-77) and Gould (1981)). The second involves the construction of a so-called incidence matrix $\underline{\Lambda}$ with zones as row labels and activities as column labels, and a 0 or 1 in each cell λ_{ij} according to whether activity j is present or absent in (ie. whether or not related to) zone i. For initial purposes, no distinction will be made to account far how many retail outlets for a given type of retailing activity are to be found in a given zone. This restriction will be relaxed later.

Suppose that the incidence matrix $\underline{\Lambda}$ is given by:

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
X1			1	1						1	1	
X2	1	1										
X3		1		1			1		1			1
X4				1	1							1
X5				1						1		
X6								1	1	1		
X7				1		1	1	1	1			

The third proposition ought strictly to have been invoked in defining the chosen level of resolution of activities and zones. Hierarchies in which they could each be embedded can be indicated as follows:

				Coarser
level N+1	Economic activities	level M+1	cities	↑
level N	Retailing activities	level M	zones	
level N-1	Types of goods	level M-1	streets	↓
				Finer

There will be many other possible hierarchies, with complex relations between levels. These will be invoked more fully later.

We can next derive a so-called shared face (or q-nearness) matrix which depicts how similar (in terms of the type of retailing activity they have in common) pairs of zones are to each other. Mathematically, the shared face matrix is given by*

$$\underline{S} = \underline{\Lambda}\,\underline{\Lambda}^T - \underline{\Omega}$$

where $\underline{\Omega}$ is a matrix of 1's. More simply, the value for each cell S_{ik} is found by counting how many activities zone i has in common with zone k, and subtracting 1. The reason for subtracting 1 is due to the pre-occupation of Q-analysis with the dimension of the spaces in which latent structures can exist (more on this below). This is given as follows:

*T denotes matrix transpose.

	X1	X2	X3	X4	X5	X6	X7
X1	3	-1	0	0	1	0	0
X2	-1	1	0	-1	-1	-1	-1
X3	0	0	4	1	0	0	2
X4	0	-1	1	2	0	-1	0
X5	1	-1	0	0	1	0	0
X6	0	-1	0	-1	0	2	1
X7	0	-1	2	0	0	1	4

The diagonal of the matrix gives the dimension of the simplices. Since zone 3 has 5 activities, $\lambda_{33} = 4$. Note that the matrix is symmetric about its diagonal. Examining the extent of pairwise similarity is close to the roots of many clustering and discrimination techniques (see for example Duran and Odell 1974, Table 1.2). A more subtle representation of these similarities is to examine which zones have a certain number of retailing activities, grouping zones together whenever they have specific activities in common with each other. In other words, a list will be compiled representing for descending values of some integer q (of course!) all those zones which contain at least q+1 activities (for q=0 to 6). In compiling the listing, at each level q, any zones that share at least q+1 activities with other zones are grouped together. The results of this listing are given as follows:

q = 6	None
q = 5	None
q = 4	{X3} {X7}
q = 3	{X1} {X3} {X7}
q = 2	{X1} {X4} {X6} {X3 X7}
q = 1	{X1 X5} {X2} {X3 X4 X6 X7}
q = 0	{X1 X2 X3 X4 X5 X6 X7}

The zones with the greatest number of activities are listed first; in this case those with 5 (at q = 4), zones 3 and 7 (refer to shared face matrix). Those with at least 4 (at q = 3) are listed next (zones 1, 3 and 7), and then those with at least 3 (zones 1, 3, 4, 6 and 7). We see at this point that not only do zones 3 and 7 each have at least 3 activities, but also that they have 3 in common with each other. They are therefore grouped together. The zones with at least 2 activities are listed next (all zones).

noting here that zones 1 and 5 have two activities in common with each other and zones 3, 4, 6 and 7 all pairwise share at least two activities with each other though not necessarily the same two in each case. There is said to be a chain of connection between this set of zones at this q level. Zones 1 and 5 are therefore grouped together, and 3, 4, 6 and 7 are grouped together. For completeness at the lowest level, q = 0, the zones with one (or more) activities are listed. Since each zone had at least two activities they all necessarily have one, and moreover, since there is no activity that is found in only one zone (or, equivalently all pairs of zones share at least one activity) they may all be grouped together (see boxed information at the foot of Figure 7).

The listing given above may be obtained from the shared face matrix by reading from the diagonal, either upwards and to the left or to the right and downwards, looking on each 'round' for successive integer values of q in descending order. For large examples, a computer algorithm may be used to produce the corresponding listing.

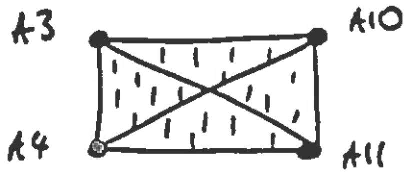
As well as considering the standard Q-analysis algorithm and results in basic algebraic form, as above, it is also useful to consider pictorial representations of the same things. This provides a better feel for the structure of multi-dimensional spaces which Atkin (see for example Atkin 1981) has been keen to expose in developing Q-analysis. Thus a geometrical view of the above procedure will now be given.

Firstly, each individual zone (each row of the original incidence matrix) can be represented pictorially as so-called simplices (see Figure7a). Each zone is represented by the activities it contains, the vertices of the simplices. Thus row labels give the names of the simplices, column labels define their vertices. The simplex for each zone is drawn by allowing a vertex for each activity it contains, and drawing a line between all pairs of vertices. This last step is important because without the line A3-A11, for instance, it might be wrongly inferred that A3 was 'closer' to A10 than it is to A11. It is also useful to represent vertices in bold face to distinguish them from the 'points' where the diagonal lines cross. These simple rules allow simplices of any dimension to be drawn on paper. However, they are only partially successful representations of simplices because, for instance, X1 should really be a tetrahedron

Figure 7. Geometrical representation of aspects of the Q-analysis algorithm

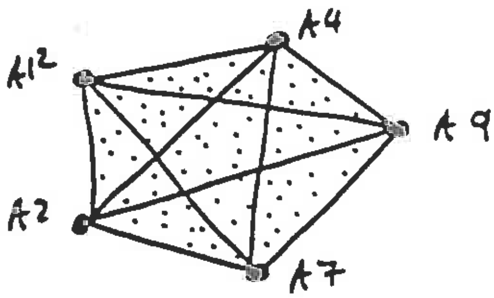
Figure 7a. The simplices

X_1 {A3 A4 A10 A11}

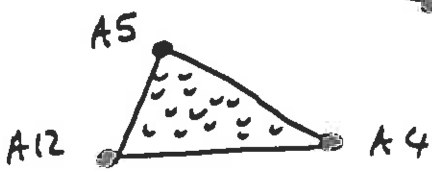


X_2 {A1 A2}

X_3 {A2 A4 A7 A9 A12}

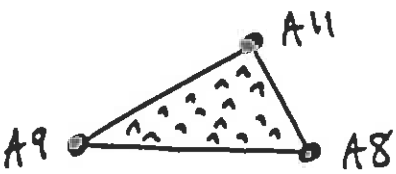


X_4 {A4 A5 A12}



X_5 {A4 A11}

X_6 {A8 A9 A11}



X_7 {A4 A6 A7 A8 A9}

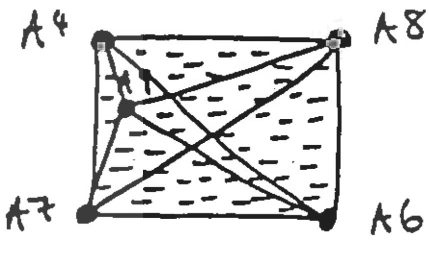


Figure 7b. The simplicial complex

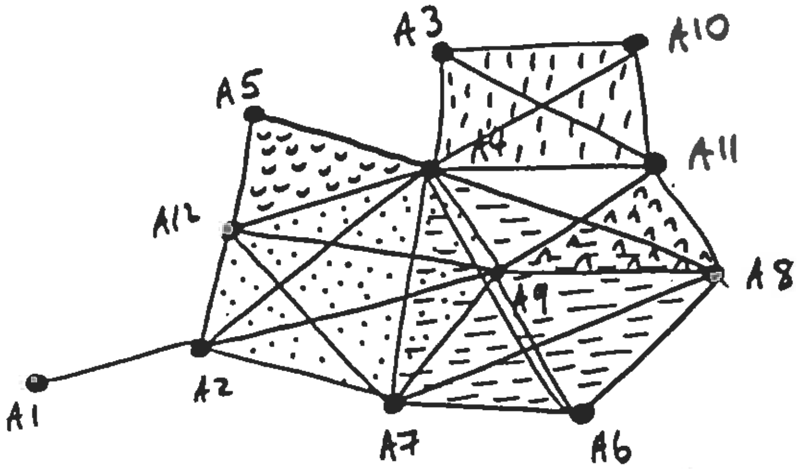


Figure 7c. The structure vector

$$\{ \overset{4}{2} \quad 3 \quad 4 \quad 3 \quad \overset{0}{1} \}$$

Figure 7d. $\hat{q}$ $\check{q}$ and eccentricities

	$\hat{q}$	$\check{q}$	$e = \dfrac{\hat{q} - \check{q}}{\check{q} + 1}$
X_1	3	1	1
X_2	1	0	1
X_3	4	2	0.666
X_4	2	1	0.5
X_5	1	1	0
X_6	2	1	0.5
X_7	4	2	0.666

Figure 7e. Contour representation

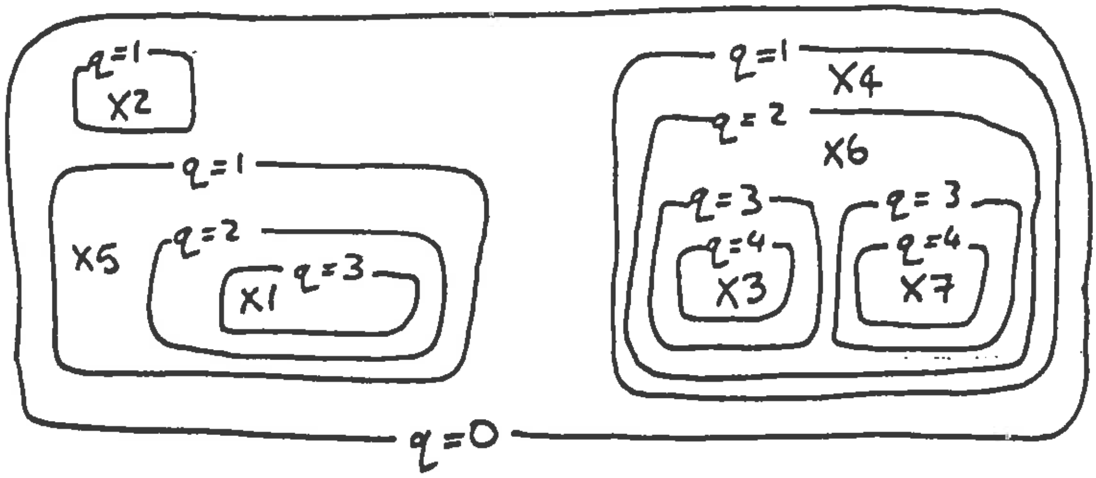


Figure 7f. Tree structure (dendogram)

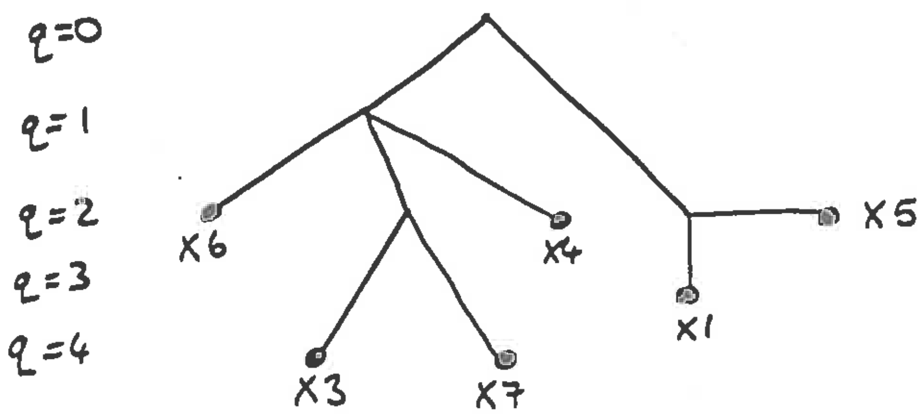


Figure 7g. Graphical representation of structure vector

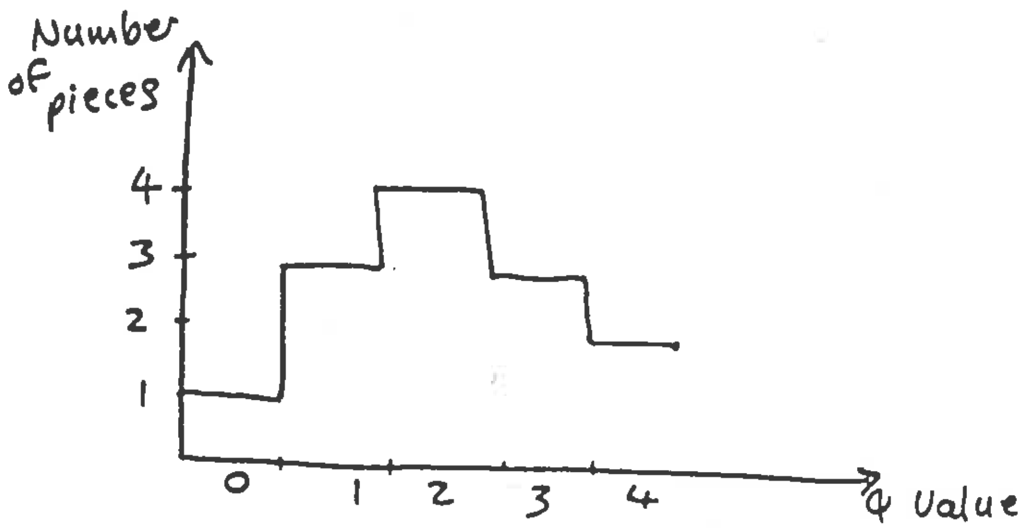


Figure 8. The conjugate analysis

Figure 8a. The shared face matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
A1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
A2	0	1	-1	0	-1	-1	0	-1	0	-1	-1	0
A3	-1	-1	0	0	-1	-1	-1	-1	-1	0	0	-1
A4	-1	0	0	4	0	0	1	0	1	0	1	1
A5	-1	-1	-1	0	0	-1	-1	-1	-1	-1	0	-1
A6	-1	-1	-1	0	-1	0	0	0	0	-1	-1	-1
A7	-1	PO	-1	-1	-1	0	-1	0	1	-1	-1	0
A8	-1	-1	-1	0	-1	0	0	1	1	-1	0	-1
A9	-1	0	-1	1	-1	0	1	1	2	-1	0	0
A10	-1	-1	0	0	-1	1	-1	-1	-1	0	0	-1
A11	-1	-1	0	1	0	-1	-1	0	0	0	2	-1
A12	-1	0	-1	1	-1	-1	0	-1	0	-1	-1	1

Figure 8b. The Q-analysis results

q = 4 {A4}

q = 3 {A4}

q = 2 {A4} {A9} {A11}

q = 1 {A2} {A4 A7 A8 A9 A11 A12}

q = 0 {A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12}

<u>Figure 8c</u>. The structure vector

$$\{\ \overset{4}{1}\quad 1\quad \overset{2}{3}\quad 2\quad \overset{0}{1}\ \}$$

<u>Figure 8d</u>. $\hat{q}$ $\check{q}$ and

	$\hat{q}$	$\check{q}$	$\dfrac{\hat{q} - \check{q}}{\check{q} + 1}$
A1	0	0	0
A2	1	0	1
A3	0	~0	0
A4	4	1	1.5
A5	0	0	0
A6	0	0	0
A7	1	1	0
A8	1	1	0
A9	2	1	0.5
A10	0	0	0
A11	2	1	0.5
A12	1	1	0

Figure 9. Some applications of Q-analysis as a clustering device

Classification of data on the multi-dimensional nature of risk perception (Macgill 1982b)

Classification of data on periodic market towns in India (Johnson and Wanmali 1981)

Classification of data on towns according to socio-economic characteristics (Beaumont and Beaumont 1982)

Classification of data on T.V. programmes according to subject matter (Gould and Johnson 1978)

Classification of data on characteristics of urban social areas (Gatrell 1981)

Classification of data on structure in bungalow plans (Flemming 1981)

Classification of data on interests of urban systems analysts (Spooner and Batty 1981)

Search for structure in art (Atkin 1974, ch. 4)

Structure in a Tudor Village (Atkin 1974, ch. 6)

Structure in land use (Atkin 1975-1977)

Structure in a University committee system (Atkin 1977) in 3 dimensional space, X4 a figure in 4 dimensional space, and so on. (Note that the dimension of the relevant space is always one less than the number of vertices on the simplex.) Representations such as those in Figure 7a serve for many purposes, however.

Note that different zones have common vertices (A12 can be found both in zone 4 and zone 6; similarly for others). We may, in fact, compress the seven pictorial representations by joining simplices wherever they have vertices that are shared with others. This produces a so-called simplicial complex (see Figure 7b). It is evident that zones not only share vertices, but also edges (through sharing a line joining two vertices) and faces (through sharing a plane that joins three vertices) and so on in higher dimensions. The shared face or q-nearness matrix simply gives the strength of connections between zones. It will become increasingly apparent that Q-analysis lays emphasis on the dimension of things, and the shared face matrix accordingly captures the dimension of the link rather than its number of vertices (the former is always one less than the latter). Thus since zones 3 and 4 share two activities, there is a one-dimensional link between then (and $\lambda_{34} = 1$).

By identifying the connectivities between all simplices at different dimensional levels (it is important to see that this is what the Q-analysis does)*, the results of the Q-analysis algorithm identify tunnels and spaces within the pictorial structure. Gould (1981b) uses the analogy of looking at a system through spectacles of different dimensional strengths. At a given q-level, say $q = 4$, we should see only simplices and faces of simplices that are of dimension 4, and where there are 4-dimensional links between simplices, we should see a space or a tunnel of that dimension linking the simplices together. Similarly a different strength of spectacles, at $q = 2$, say, would enable us to see only 2-dimensional simplices, some possibly inter-connected via 2-dimensional 'tunnels', others quite disconnected from others. Thus one way of viewing the written outputs of the q-analysis algorithm is as a depiction of the number of pieces the full geometry falls into at each dimensional level, along with the composition of each piece.

---

* Thus, the standard Q-analysis algorithm, in the context of the present example, clusters zones according to the number of retail activities they have in common with each other. For any two zones X_i and X_k that are grouped together at a given dimensional level they will either (a) have a given set of retail activities in common, or (b) share a certain set of activities with some intermediate zone X_j (i.e. $X_i \rightarrow X_j \rightarrow X_k$), or (c) there will be a longer so-called chain of connection between X_i and X_k (e.g. $X_i \rightarrow X_j \rightarrow X_l \rightarrow X_m \rightarrow X_k$). In all cases the number of activities required to be in each 'set' is defined by the appropriate dimensional level. Thus the Q-analysis algorithm produces a listing of connected entities at each dimensional level.

The notion of structure may initially be grasped most readily from the geometrical (pictorial) representations. However, it is important to recognise that the algebra of the algorithm, along with other measures that will be given below provide alternative ways of describing the same structure. This recognition is important for more complex cases where it may not be possible to draw the geometry of a complete simplicial complex.

The basic Q-analysis algorithm has now been given both computationally and geometrically. Before interpreting the results and building further on the foundations it offers, some further measures and indicators that capture particular properties will be given.

A so-called structure vector (Figure 7c) may be used to summarise the number of clusters (or components) at each level, ie. the number of legitimate groupings (or groups with chains of q-nearness) for each values of q (ie. different levels of intensity or retailing activity). Thus there are two components at the 5 activity levels, 3 at the 4 activity level, and so on down to 1 at the one activity level. It should be recognised that if it had happened that at this lowest level there were 2 components, the complex sketched in Figure 7b would be in two pieces. The fewer the number of components at each level, the more highly bonded the original system of activities and zones. The components themselves can be recognised as being sets (since there is a well-defined procedure for identifying them). As subsets of the original sets they qualify as elements of a 'higher level' set.

The concept of eccentricity is used to capture the degree to which a given simplex is embedded within the rest of a complex. For each simplex (in this case zone) in turn, and letting $\hat{q}$ (top q) be its own dimension (the number of activities it contains minus one) and $\check{q}$ (bottom q) be the level at which the simplex is connected with any other, the eccentricity of the simplex is given by

$$\text{eccentricity} = \frac{\hat{q} - \check{q}}{\check{q} + 1}$$

$\hat{q}$, $\check{q}$ and the eccentricities in the current example are listed in Figure 7d. In theory, the magnitude of eccentricities can range from 0 to ∞. A simplex that is completely disconnected from the rest of a complex (has nothing in common with it) will have very high eccentricity (∞). At the other extreme, a simplex that has no vertices (characteristics) that are distinctive or unique to itself will have low (zero) eccentricity. The latter will merely be a face of some other simplex. From Figure 7b it is seen that zone X2 is the most eccentric (indeed, it 'sticks out' of the complex in figure7b), and X5 the least so, having zero eccentricity (it turns out to be a 'face' of X1). There are no zones with infinite eccentricity; this ties in with the fact that the complex is in one piece.

The shared face matrix, Q-listing and structure vector give a summary of the global structure. The eccentricities say something about local structure.

Alternative ways of presenting the results of the Q-analysis algorithm are via contour representation (see Figure 7e) or in a tree structure (Figure 7f). The structure vector can alternatively be depicted graphically (Figure 7g).

The operations involved in the above algorithm are simple and mechanical (if rather tedious, if the sets are large, but in this case a computerised version of the algorithm would be used and the graphical representations would typically be omitted) once the sets have been defined and the relations between them identified. Note that a similar analysis could be performed based on the columns to complement the one above which was based on the rows of the original incidence matrix. This would be a so-called conjugate analysis. This is summarised in Figure 8, but omitting the graphical representations. Whereas the original Q-analysis described the pattern of retailing activity within zones, the conjugate analysis describes the distribution of activities across zones. In formal notation, the original analysis would depict the relation $K_X(A;\lambda)$, and the conjugate, the relation $K_A(X;\lambda^{-1})$.

The whole of the exposition of the Q-analysis algorithm has so far been based on data in the form of a matrix of zeros and ones. This in turn was intended to capture the interrelations between two specified sets. It is reasonable to enquire whether this is not unduly restrictive, particularly as in some cases, more numerate information about the relationship between two sets will be readily available (eg. the number of retailing outlets of a given type in different zones, or total floorspace or turnover, or whatever).

Q-analysis can handle more numerate relations and does so via a procedure known as slicing. Given a data matrix depicting a relation between sets, slicing would involve deciding some threshold value for each cell (sometimes the same value is chosen either for all cells, or for all terms in one column or for all terms in one row) and setting equal to one those cells whose original value was greater than or equal to the threshold value, and setting equal to zero all other cells. Suitable threshold values would be determined by the analyst (in an analysis of primitive markets, Johnson and Wanmali (1981) related the threshold value to the maximum distance that could be walked in a day). It is often appropriate to run a family of Q-analyses corresponding to a family of different slicing values, because it is not always possible to define <u>a priori</u> what the most suitable slicing level is. This results in a family of Q-analyses for a particular data set.* Lest the consequent profusion of results should appear unduly disconcerting it can be noted that complex data is not necessarily most productively analysed by simple means, and, more importantly, Q-analysis provides explicit guidelines on how to interpret the results, and this in turn can shed light on the appropriateness of particular slicing levels.

4. <u>Utilisation of results I: static interpretation of structure</u>

It is one thing to perform an analysis such as that above; quite another to gain practical benefit from doing so. The formal results of Q-analyses are often presented as listings of the general form of page 19 which, for high dimensional systems, can run to several pages. But how can they be interpreted, and what do they tell us that we did not already know, or could not get some other way?

Firstly, the analysis reveals a natural clustering in the original data through the grouping of elements (zones) into components or clusters. Thus it contributes to the long search for indicators and summary measures, that are essential in order to be able to manage, communicate and discriminate within large data sets. In order to be successful in this capacity, it is necessary to appreciate the significance of the Q-analysis components. The clusters that are shown there have been referred to above as legitimate groupings. What does this in turn imply?

*see for example Gatrell(1981) and Macgill (1982b), At the 5 activity level ($q=4$) zones 3 and 7 are identified, but discriminated from each other (because they do not share 5 activities). At $q = 3$ zone 1 is also identified, but 1, 3 and 7 are still all discriminated from each other (because they do not share 4 activities). At $q = 2$, zones 1, 4, 6, 3 and 7 are all identified. At this level there is now no discrimination between 3 and 7. The other zones at this level, 1, 4 and 6, are discriminated from each other and from 3 and 7. At $q = 1$ all zones are identified, and there is no discrimination between 1 and 5 or between 3, 4, 6, 7, but there is discrimination between any other possible groupings. At $q = 0$ all zones areagain identified and there is no discrimination between any of them.

In using the results of the Q;analysis algorithm for data representation it is necessary to consider all q-levels, and not just any given one, in order to capture the richness of the structure of the data. To consider only one levèl would be no better than the simple partitioning schemes that were criticised under proposition II above.

Results of clustering analyses are not only useful in their own right, but also, of course, in comparative work (eg. intertemporal or interregional comparisons of difference and change).* In the present context we may compare the retail structure of two cities. The original Q-analysis summarises the similarities between zones in terms of the variety of retail outlets presented. The conjugate pin-points whether, for instance, a given retail activity is to be found in many zones in one city and only a few zones in another. It also pin-points whether retail activities are found in similar combinations in different cities, and the relative scope for multi-purpose trips. The basic Q-analytic terms of reference outlined above provide a useful basis for such comparison because they provide <u>hard</u> descriptions of such characteristics.

---

* The existence of meaningful and well defined indicators in such contexts is of paramount importance.

The other properties identified by Q-analysis augment such summaries and comparisons. The shared face matrix picks out the degree of pairwise similarity between zones (page 19) and activities (Figure 8a)*. 'Top q' highlights the range of goods provided (in the original analysis) or the extent of provision of some good (in the conjugate analysis). The eccentricities highlight the distinctiveness of zones (those with low eccentricity being not at all distinctive, those with high eccentricity being very distinctive in terms of retail provision) and of activities (those with low values being found always in spatial association with other activities, those with high values being found in relative isolation of others). A spatial plot of the zonal eccentricities may be a particularly suitable way of bringing out zonal distinctiveness. An urban redevelopment scheme which removed retailing activity from a zone with zero or very low eccentricity (zone 5) would have a very different effect than one with high eccentricity (zone 4), because in the former case the type of retail outlet concerned can be found elsewhere. The resulting effect on shopping habits and trading patterns cannot, however be described by the tools of Q-analysis presented so far but instead involve the characterisation of traffic (see below).

The structure vectors summarise properties of the data in a more aggregate way than the listing on page 19 and Figure 8b, and it would typically be interpreted alongside the fuller q-listing. Thus there are five zones that provide three types of retailing activity, but the three in the structure vector represents the fact that only two of these provide the same three types.

The powers of Q-analysis as a clustering device make interesting comparison with traditional clustering and factor-analytic techniques that are perhaps currently more familiar to geographers. The general context is where there are a number of entities or individuals, and a number attributes or descriptive features that can be associated with them to various degrees. A

*The key role of A4 has not been adequately picked up by the summary measures. It is the main cause of the large component at q = 1 in the conjugate, because it is connected to a large number of other activities at that level (eccentricity does not pick this up). Perhaps the standard outputs of the Q-analysis need to be augmented with a listing showing the number of other elements each element is connected to at each q-level.

numerical data matrix depicts the weight with which each feature or attribute applies to each entity. For any reasonably sized data set, formal analytical techniques and summary measures are required in order to comprehend and find some way around such multi-dimensional data, and there is by now a wealth of so-called factor and cluster analysis methods available (Everitt (1980), Duran and Odell (1974), Klecka et al (1975), Goddard (1976). However, whereas many traditional methods first 'work on' the original data in what many would argue to be an unnatural way (by calculating correlation coefficients, factor scores, rotations and so on, and then referring to these rather than the original data), the Q-analysis uses the data in its virgin form and changes only its appearance and not its content. Gatrell (1981) is particularly scathing about the violence done to data by traditional methods. Thus the clustering powers of Q-analysis seem to be a challenge to others to rework their factor analytic (and other discriminant) analyses to see what differences and consequentially alternative insights into the structure of data in results are obtained. Gatrell (1981), Chapman (1981) and Beaumont and Beaumont (1982) have done so with significantly different and what they would argue to be more plausible and more discriminating results, and no artificially imposed structure. This has inspired them to add to a campaign against traditional methods that may already be found in the literature (Williams, 1971). Applications of the Q-analysis algorithm as a clustering device are listed in Figure 9. Thus the standard outputs of the Q-analysis algorithm as with other clustering and discrimination devices, can act as a catalyst for guiding the user back through his or her original data, hopefully seeing meaning where previously there were just numbers (or, in the case of qualitative information which could be grouped into sets, just lists). Potential users should not be too tied to the basic algofithm alone for such work; other measures and indicatore (the shared face matrix, eccentricities, for example) or even further indicators that have not yet been developed or recognised as "mainstream Q-analysis" may be more meaningful in some cases. Some familiarity with all possible methods is required before would be user can judge which is most suitable. Given the wealth of methods, this is no easy task (see Everitt; Duran and Odell). The Q-algorithm may not be as new as generally assumed (see Macgill 1982a). Moreover Johnson (1981) argues that a number of refinements to the basic method of Q-analysis outlined above are in order if clustering is the main aim of an analysis (see also Macgill 1982c). This involves inspecting the numerical values from the original data matrix to see whether there are grounds for refining the clusters produced from a standard Q-analysis. This in turn is based on the observation that wherever two simplices in a simplicial complex share common vertices, each such vertex carries two different weights (one weight for each simplex). Wherever such weights are relatively similar there would seem to be no grounds for further discriminating the clusters produced from the basic Q-analysis. However, if such weights are relatively different, there would seem to be grounds for further discrimination. An example illustrating this method originally presented by Johnson (1981) is reworked in simpler terms by Macgill (1982c). A more complex case is given in Macgill (1982b).

Despite its apparent virtues for clustering, Q-analysis is not primarily a clustering device. It is clear from the literature that Atkins' development of Q-analysis was always bent towards a more ambitious goal than providing yet another clustering method; thus clustering is an incidental, albeit quite powerful, off-shoot. To appreciate this we must refer back to what was given above as part of the underlying philosophy and theory of the approach, namely to use the Q-analysis algorithm as a means of revealing a latent (hitherto unseen) structure which constrains the way activity (traffic - monetary expenditure, physiological stress, disease, time) can move, or exert influence. The general theory is therefore that there are forces and other patterns and processes of activity that are supported or produced by the structure, depending on the configuration of the structure, and exert possible reciprocal influence on the structure. Such phenomena may have been intuitively sensed, Q-analysis seeks explicitly to reveal them.

5. Utilisation of results II: traffic

The next idea that needs to be conveyed is that the structure that may be defined via the above propositions and algorithm may provide a backcloth or stage for events, action and processes; so-called 'traffic' in the terminology of Q-analysis. Paraphrasing Atkin (1981): The stage is not the three dimensional structure on which plays are acted out, but the multi-dimensional structure that must be inhabited to argue politics, wage wars ... . Three dimensional space is an important special case, it is one that Einstein found over-restrictive in developing his understanding of the world. He introduced a fourth dimension, time, and theorised about tiny points called Sun, Earth, Mars ..., moving around in a modest four dimensional space. Atkin is pleading for social scientists to realise that they must typically inhabit a higher dimensional space. * The dimension of the 'world' that must be inhabited is the dimension of the structure that may be identified (via the above principles and operations) in any area of study, fully to address the issues that are fundamental to that study.

Atkin has stated that ideas of traffic are not always well-defined. Although there are some obvious examples of traffic, there has also been much confusion and misunderstanding over this concept (the writer has not escaped unscathed). At the risk of getting it wrong, it will be suggested here that we can identify at least four different types of traffic. First so-called free traffic that can be readily supported on simplices and in tunnels between them (expenditure patterns, conversation, disease, for instance). Second, traffic that is manifested as a force on or between simplices (pain, laughter, tension, pressure of time). This type of traffic is a 'resultant' of the structure and of changes in 'free' traffic. Third, traffic which is manifested as weights on the vertices of individual simplices; the weights are typically not the same for a given vertex, since this vertex may be shared by more than one simplex (so-called discrimination traffic). This may suggest that the structure could be broken up. Fourth, traffic whose dimension does not match the dimension of the simplex it should relate to (probability, frustration, incomprehension).

Thus any given structure may be related to a variety of different types of traffic. The identification of the sort of traffic of interest in any particular context rests largely with the analyst (and his audience), as does the interpretation of which properties may be of interest or use. A number of properties and interpretations of traffic that have been profitably exploited in various contexts will be discussed below in separate (though necessarily linked) subsections, to indicate the range of possibilities; not all will be meaningful let alone relevant in all contexts. Crucial in all cases is that traffic is less static than backcloth and the nature of the structure will profoundly influence the sort of traffic that can exist, and the way in which

---

* Many others have made the plea for multi-dimensional analysis of social space, for example Harvey (1973) Ch.1. Atkin goes considerably further than most in developing appropriate tools for the purpose. The starting point is to recognise that there are two multi-dimensional structures (spaces), namely the simplicial complex and its conjugate, associated with any two sets that are related to each other.

it can change and move about. Atkin (1965, 1971) has argued this to be the case in physics*, and examples cited below seek to mirror this in the social sciences. The importance of knowing the configuration of the spaces, the strength of the connectivities and the number of pieces (via the above algorithm and indicators) is important because these define the only spaces and channels where traffic can exist and relate to. Thus these become more than signposts back to the original data in terms of interpreting the data in a static sense. We start with relatively simple ideas, and build up to something deeper. The goal throughout is to present hard descriptions of properties and phenomena. Having achieved this, it may be possible to move from the descriptive to the prescriptive. For example, weak points in the structure may be identified; focal points similarly. Thus in terms of prescription it may be desirable to ease (information, vehicular congestion, etc.) or hinder (eg. disease) the flow of traffic, or redistribute it over the structure. This is done by examining where the structure should be broken or added to, or whether there are closed loops in the structure around which 'traffic' endlessly circulates. These are all notions that can be expressed in plain English in the context of particular applications, but need Q-analysis as a foundation as guidance for what can be said.

(i) The basic idea of 'free' traffic

The simplices that may be identified from each row of the data matrix, (or each column in the case of the conjugate) determine locations (in multi-dimensional space) where certain types of so-called traffic can occur. The vertices determine the nature of the traffic. Possible traffic for the retailing example worked through above, amongst other things, could include consumer patronage or consumer expenditure in various zones. A zone providing a particular range of retailing activities will support consumers (traffic) that wish to purchase that range (or a subset of the range). More generally, a structure (or part of a structure) of a given dimension allows traffic with the same dimension to exist on it. It is called 'free' traffic here because there is no conflict between the dimension of the traffic and the space in which it exists. Traffic of dimension n is called n-traffic

---

* cited in Johnson (1981b)

or traffic of grade n. Possible traffic for the conjugate analysis includes retail turnover for a particular activity, consumer expenditure in a particular activity, capital investment in or tax on a particular type of retail activity. It can be seen that in some cases, traffic on the original structure can be redefined in terms of traffic on the conjugate and vice versa (this would certainly be possible for consumer expenditure, for instance), and it is sometimes useful to do this, and examine the way in which traffic on a single simplex in the original analysis (expenditure in zone 3, for instance) is spread out over several in the conjugate (A2, A4, A7, A9, A12 i.e. the activities present in that zone).

Since traffic is related to vertices, simplices can only experience it if they have the appropriate vertices. Zone 1 will clearly not directly benefit from expenditure on A2, nor will shoppers in zone 4 benefit from promotion of A3 (though this is not to say that there will not be indirect effects, see below). These are contrived examples. A more significant illustration in a different context may be the case of a committee making an investment decision overlooking a particular opportunity (or research council misjudging the value of particular research proposals) because it did not relate to its members own interests or concerns.

The latter situation may be represented by defining simplices whose vertices represent the concerns of individual members. Making an investment decision could be visualised as traffic on such simplices, but the decision outcome will necessarily be restricted to the concerns of the committee members. As another example, a scholar who is not aware of a particular analytical approach (ie. this approach does not figure as a vertex on the simplex representing his area of knowledge) will not be able to utilise this approach (his traffic-the application of knowledge will exclude that approach).

A different perspective on the same idea may emerge by considering the concept of a feasible region in mathematical programming. The idea of 'no feasible solution' in mathematical programming terms would appear to correspond in broad Q-analysis terms with the sought for 'solution' relating to a vertex or vertices that have not been identified. (either because it does not exist or because it has been overlooked).

In other words, the backloth is missing or deficient.

Examples of 'free' traffic are probably even more diverse than the variety of contexts in which latent structures can be identified (since there may be many different kinds of such traffic for any particular structure. It may be tangible, intangible, measurable or qualitative. Further examples are given below. Time is a particularly imaginative case (see Atkin, 1978 a,b, 1982). Chapman (1981) asks how to distinguish between backcloth and traffic. After all, measurement of traffic on vertices could itself be representable in an incidence matrix which in turn could generate a backcloth (thus apparently traffic can also be backcloth). Atkin (1979) suggests that the choice of features which go to make up the structure must be a relative one. Thus the structure is (relatively) static compared to the traffic it may support.

(ii) Local movement of "free" traffic

Traffic not only occurs on simplices and faces of simplices, but may be transmitted across faces and therefore between adjacent simplices. If two zones both provide a given combination of retail activities, then, ceteris parabus, shoppers will be indifferent between them, and there could be a free movement of consumer patronage between them. In the example worked though above, shoppers buying only from activities 4, 7 and 9 would be expected to be indifferent between zones 3 and 7.

In one of the different contexts alluded to above namely a structure relating academics to their topics of research, one type of traffic could be individuals'application of knowledge (traffic on the simplices which represent the academic interests of each individual). Another type of traffic could be scholarly communication between individuals. This can only occur across vertices (on subjects) they share (or, more generally, across the faces defined by the vertices they have in common), and the traffic will be of dimension of their shared vertices.

The coincidence ofthe dimension of free traffic with the dimension of simplices or faces between simplices has further implications. Notably, any simplex with infinite eccentricity (ie. completely disconnected at the $q = 0$ level) cannot have traffic transmitted to it from elsewhere because there are no links on which it could be carried.

These general ideas about traffic movement can be related to the formal structures revealed by the Q-analysis. This is because traffic is highly dependent (for its existence and nature) on the structure on which it exists. The shared face matrix identifies the extent of pairwise links (and hence bridges or tunnels between simplices across which traffic can move)*and the strength of the links (and hence the grade of possible traffic). The Q-analysis listing summarises the connectivity of the structure in a more comprehensive way than the shared face matrix. It shows how many pieces the structure falls into at each Q-level, and the simplicesinvolved, and these pieces bound the movement of traffic of a given grade. Thus 2-traffic can only be transmitted between simplices that are 2-connected, ie. move within components at q = 2 i.e. 3 dimensional tunnels. High dimensional traffic (ideas, ...) can only move across highly joined structures (ie. within components that exist at higher levels - in other words, within higher dimensional tunnels).

(iii) Global movement of "free" traffic

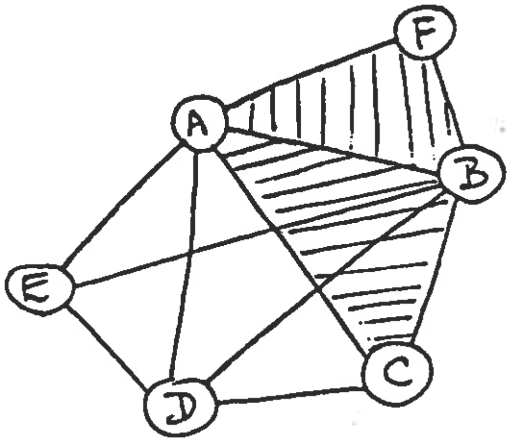
The idea of traffic being transmitted between adjacent simplices leads to the idea of traffic being transmitted potentially through the whole structure, as long as the structure is appropriately connected. Again, this is not particularly well illustrated in the retailing example (not all q-analysis properties are relevant for all examples any more than consumers are likely to want to buy all goods in a given shop). Suppose, however, that zone 2 were pedestrianised and subsequently attracted more customers. This would (assuming the same total expenditure on each activity) cause reduced expenditure in zone 3 (assuming shoppers patronised activity 2 in zone 2 rather than zone 3. Whether more or fewer shoppers would be attracted to the other activities in zone 3, as a result of this zone being less busy can only be guessed at. Whatever the effect, it in turn, is likely to affect patronage in other zones (in terms of changed patronage in sharedactivities). Thus an initial change in zone 2 can affect zones 1, 4, 5, 6 and 7, even though these are not directly connected to each

* For any given simplex, the set of all other simplices to which it is directly linked is sometimes called its 'star' and for simplex i is defined from row i (or column i) of the shared face matrix. The star can be visualised as a multidimensional space within the latent structure.

other. This of course, is a simple example of the classic systemic effect; a possible way of representing indirect/feedback or interrelationships (more on this below). A simplex disconnected from the main body of the complex would not be affected by changes in the main complex.

Three general qualitative features of global movement of traffic will be noted here.

First the nature of the traffic is likely to change as it passes from one simplex to another. This is because different pairs of simplices are connected by different sets of vertices; thus traffic is channelled in different ways (ie. must relate to different vertices). This may be illustrated in a different context by considering a lecturer who (simplex I) draws upon knowledge in five areas - ABCDE (see diagram).



Simplex I	{ABCDE}
Simplex II	{ABC}
Simplex III	{ABF}

Suppose his students (simplex II) only assimilate ABC, in this case the traffic (knowledge) they receive will be diluted (it starts as 5 traffic, but is received as 3-traffic). Moreover, an absentee (simplex III) asking a colleague to recap the lecture for him, but offering only vertices AB and F, will receive an even more diluted 'version' of the original traffic. Chapman (1981) suggests that corresponding changes are to be expected in the way that inflation and recession, language and fashion, disease and catastrophe, pain and pleasure may flow across connected structures. (Note that activities/units/entities that are adjacent to each other in a simplicial complex representation, are not necessarily spatially adjacent to each other in the physical world.)

The notion that traffic may be changed or transformed as it moves through a structure (simplicial, complex) requires an awareness of the sense in which a given type of traffic (communication, time, disease, ie. traffic known by a given name) is a distinct entity. Not only does traffic change dimension (and therefore character) but 2-traffic on one part of a structure is likely to be rather different from 2-traffic on another part of a structure (because different sets of vertices will be involved). Thus traffic and 'free' traffic are generic terms used for various notions of movement on a simplex and unless given further qualification do not refer to unique entities in any given context (see also under (iv) Grading below).

Second, since traffic can only be transmitted between simplices which are adjacent to each other then from any given starting point traffic is likely to have to cross other simplices in a particular order. Thus to get to D from A, traffic would (diagram i) have to traverse B and then C. (Note that the fact that simplices are adjacent in a complex does not necessarily mean that traffic <u>will</u> move between them.) In this case there is no direct contact between A and D, and traffic cannot reach E at all. In other cases structural configurations may arise in which traffic circulates and eventually arrives back at its starting point, having undergone various transformations, and perhaps even goes round the structure again. Such traffic has been called 'noise', and a simple example of a structure inviting noise is indicated here (diagram ii)

Diagram (i)

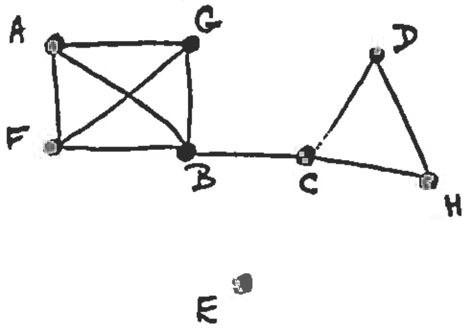
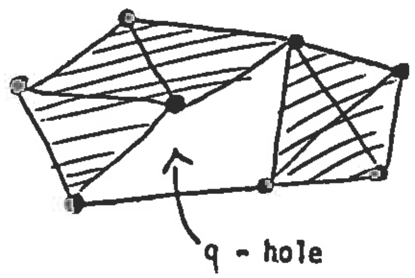


Diagram (ii)



The hole in the middle of the structure is analogous to a physical object in the physical world which would have to be walked around rather than passed through. The physical world offers only 3 dimensions (the spatial dimensions), typically not the three most important in the social world. The idea of a q-hole captures the notion that we find different kinds of objects in the multidimensional world of human affairs. These may be visualised as closed loops of various dimension that might exist in the geometrical (simplicial complex) representation of a relation being studied. Atkin (1974) apparently found a multitude of such q-holes in the committee structure of the university of Essex which encouraged continuous circulation of committee business and mitigated against decisions ever being arrived at. Elsewhere (Atkin, 1977) has suggested that q-holes may be focal points for attracting (or generating) energy: an idea worth further development?

There would appear to be some ambiguity over whether it is possible to identify the existence of Q-holes from some Q-analysis algorithm. Atkin (1977), p. 85 suggests the existence of an algorithm that can find them. However, Gould (1981) and more formally Johnson (1981c) Earl and Johnson (1981) and Griffiths and Evans (1981) more recently suggest that more careful analysis is required. The sticking point seems to be the illusiveness of being able to specify the number of simplex edges needed to make a hollow loop in relation to the dimension of the simplices involved (and hence the dimension of circulating traffic).

Thirdly, it will be remembered that the so-called structure vector summarises the global connectivity structure of the complex. The obstruction vector, formed by subtracting one from each element in the structure vector, thus accordingly summarises the degree of obstruction to the movement of traffic (the number of gaps between the q-connected components). For the retailing activity example above (Figure 8), it would be {1 2 3 2 0} and indicate the scope for multi-purpose trips.

This indicates no obstruction at the 0 level and only a modest amount at intermediate levels. A structure vector consisting entirely of 1's (and the obstruction vector entirely of zeros) indicates a free flow of traffic. This does not necessarily mean that all simplices are connected at all levels, because all may not be identified. Caution is therefore required in interpreting the structure and obstruction vectors independently of the rest of the Q-analysis. Johnson (1981b) starts to get round this by calling the above the <u>first</u> structure vector, and a further vector, giving the number of named simplices at each q level a <u>second</u> structure vector.

Prescriptive suggestions may be made given an understanding of the structure generated in a Q-analysis. It may, for example, be desirable to break the structure up in order to inhibit the transmission of traffic (disease, inflation?). Alternatively it may be desirable to decrease the obstruction in order to encourage traffic (pleasure, energy ...?). The former would involve examining the shared face matrix (or the suggestion on p.26) in order to identify elements that are crucial in holding the structure together. The latter involves adding extra vertices to elements hitherto poorly connected with the structure (again found from an examination of the shared face matrix).

(iv) The grading of traffic

In his development of Q-analysis Atkin has been more specific in drawing attention to the grade (or dimension) of traffic. A number of examples suggesting that different types of traffic can have different dimension (or grade) have already been noted in passing - different 'levels' of conversation on a complex whose vertices are 'interests'and 'people'; different 'levels of academic communication' on a complex whose vertices are academic topics and simplices are scholars: shoppers requiring different numbers of goods. Other examples have been investigated more closely by Atkin. In an analysis of a Shakespearean drama, for instance, a latent structure was derived from the relationship between scenes and characters. Traffic was depicted as the audience's appreciation of the play. The most superficial level of appreciation (0-traffic) comprised following the action as a simple sequence of events. At successively higher levels of appreciation (higher dimensional traffic, t-traffic) the fabric of the web of interlinkages of plots and characters is more fully appreciated by the audience. A similar view has been taken by Atkin (1974) of such high level games as chess or snooker where the master of the game appreciates each move as a 'gestalt' experience. In such cases the structure is derived from locations on the board and pieces. 0-traffic would be an understanding of the game would be to see it as a sequence of moves that were made independently, one after another. T-traffic (high grade traffic) would recognise the advantages of positional play and the intimate interconnectivities between pieces, possible moves and sections of the board.

We might even extend these ideas of grade of traffic to one's own individual understanding of Q-analysis. A superficial understanding could be depicted as low grade traffic (0 or 1 dimensional); only one or two facets of the approach can be followed at any one time. With increasing understanding of the approach, several features can be appreciated simultaneously, so that the 'traffic of understanding' would be of a much higher grade. The pursuit of any other kind of knowledge in all manner of fields of inquiry - whether as researcher, teacher or student - can probably be depicted in a similar way. At a more mind-boggling level still, the meaning of life! In terms of understanding given subjects, people claiming to understand them when they do not could be said to have a traffic of understanding of too low a dimension for the dimension of the backcloth structure on which it is supposedly based. More pragmatic illustrations of the same idea could arise in considering what makes for good managerial practice (in private industry, health care institutions, war, government, ...).

All'activity' may be represented as traffic on some backcloth: thus traffic and backloth each arise in infinite variety. The strict set-membership guidelines, well-defined inter-hierarchical relations and ideas about the gradation of traffic turn this universal observation into a potentially powerful conceptual model. The utility of the model will vary with the context, and in any particular context if is necessary to be flexible about the choice of both backcloth and traffic: the former can arise from the interrelation of <u>any</u> two sets that may be defined in any given context; any given backloth in turn can support a variety of different types of traffic.

Given that some ideas about global circulation of traffic appear to have such close affinity with characteristics of systems analytical methods (holism, feedback, systemic effects), it is useful to inquire into the nature of the new perspectivethat Q-analysis may provide. One particularly telling and distinctive feature would seem to relate to the grade (dimension)of circulating traffic. The links in the classic network or graph-theoretic representations in systems analysis correspond to channels suitable for carrying (only) 0-dimensional traffic (0-traffic). The connectivities identified in the higher level geometry invoked by Q-analysis are able to differentiate between different dimensions of traffic. Thus, it may not be unreasonable to suggest that where the 'free' traffic analysed in a Q-analysis is only 0-dimensional,a suitable representation could be obtained by invoking traditional network or graph theoretic concepts instead. However, wherever higher dimensional traffic occurs, the richer framework of Q-analysis is needed.

Even in the case where traffic in 'only' 0- dimensionsl , the characteristic distinction between backcloth and traffic is a useful conceptual tool that does no arise in network analysis. Also, the more searching consideration of set membership and hierarchy mentioned earlier. Moreover, whereas traditional methods are confined to dealing only with 0-dimensional interactions, Q-analysis will find any such interactions as a special case of something more general, if they exist.

It is not sufficient merely to claim interest in a given type of traffic of a given grade. It is also necessary to indicate the simplex on which it occurs and the vertices on which it is defined. A contribution to 2-traffic on zone 6 (Figure 7a) may be defined on any of the pairs (9,11), (8,9), (8,11). There are many more possible pairs in zones 2 or 7. The writer has found some ambiguity in the literature in determining the exact vertices for a given grade of traffic. It would appear however that for some purposes (plot of play), the grade of certain traffic is of more immediate interests than the vertices on which it is defined. It is for analyst to clarify whether the grade or the type (vertices) is of greater initial significance (ultimately of course, both aspects should be clarified). Note that in this connection that traffic of a given grade can only circulate within components of the appropriate dimension (as found from the Q-analysis algorithm) and cannot 'reach' simplices that do not appear in such components. It may further be inferred that traffic of a given grade may potentially be transmitted through long tunnels in the structure (when Q components are large). However, applications to date have apparently only encountered relatively short chains of transmission for traffic. Long chains can, however, arise when traffic changes grade as it moves through the structure. In a more advanced paper, Johnson (1982) examines further the conditions under which transmission of traffic of a given grade can affect transmission of traffic of a different grade.

(v) Notation for traffic

In many cases, it is necessary to be more specific than in the above qualitative discussion in representing traffic, namely to seek to quantify it. This might involve counting the numbers of people at shopping centres or ascertaining their expenditure, noting the length (duration) of television programmes or some other way of capturing the 'size' or 'intensity' of the traffic. The symbol Π is usually used to represent a given type of traffic on a given structure. A superscript may be attached to it to identify different grades of this traffic, and hence the dimension of the simplex (or face of the simplex) on which it can occur (but see below for traffic whose grade does not match the dimension of the simplex). For a simplex of dimension n, there can be potentially n + 1 grades of traffic (corresponding to q levels 0, 1, 2, ..., n), which can be represented by a so-called pattern

$$\Pi = \Pi^0 + \Pi^1 + \Pi^2 + \ldots + \Pi^n$$

where we read + as 'and' rather than 'plus' (the operation + has a strict topological meaning, but 'and' will do for present purposes). Π^i thus represents traffic of grade i (or i-traffic), and Π is called a graded pattern. The numerical value of each Π^t gives the 'intensity' or 'strength' of the traffic. It may be a single number for each grade (if there is only one location in the simplex of interest, or if locations are summed), or a set of numbers for each (one for each vertex or face on which traffic of that grade occurs. Johnson (1981b) argues that any pattern of numbers Π that can be represented in the above form and mapped onto a simplicial complex can represent traffic, of some kind. So far in the present paper we have considered only 'free' traffic, further types will be considered below and it will be possible to represent these further types in the same form of pattern Π as given above. Thus Johnson (1981b) defines traffic as any graded pattern on a complex.

(vi) <u>Forces - static backcloth</u>

Atkin's Q-analysis propounds the theory that changes in the intensity of traffic that may be observed on given simplices may be interpreted in terms of so-called t-forces. The most obvious analogy is to compare traffic with velocity and force with acceleration. Atkin goes much further than this. Thus, just as the idea of a multi-dimensional structure is to replace the classical notion of three dimensional space as the stage on which human affairs are enacted, so the idea of a t-force is (a multi-dimensional) analogy to the (uni-dimensional) idea of a gravitational force in classical physics. Gould (1981) following Atkin (1974) notes that in Einstein's theory of relativity, gravity is not considered as a force (which no -one has defined except in a completely tautological way), but rather in terms of a geometric structure of the universe that allows certain things to happen and forbids others.

Thus as Einstein gave a revolutionary interpretation of force in the physics of his day, so now Atkin is trying to carry this revolution forward to the social science of today by arguing that human affairs take place on multi-dimensional structures and multi-dimensional structural forces will thus affect the intensity of traffic that occurs on them; such changes in traffic may be the result of forces). Thinking of changes in traffic in terms of forces (if we are sympathetic to the transition to a multi-dimensional world, Atkin 1981) may be a basis for generating generating new insights into systems of interest. It may or may not be of obvious practical benefit (in terms, say, of generating immediate new insights, or finding 'solutions'), but it would seem to be an exciting and imaginative basis for the generation of new hypotheses or theories. It is in this sense that Q-analysis might be seen to provide a framework for thought, and not only a method of analysis. The physical analogy, moreover is more productively regarded as a catalyst for thought than a basis for belief.

As a simple example, a shopping centre providing three types of goods (and thus representable by a 2-simplex) and taking £x in one year and £y a year later could be said to experience a t-force (in fact a 2-force, since it occurs on a 2-simplex) of £(y - x). A force is thus associated with change or 'difference' in line with the physical meaning; in some cases it may be a symptom of stress, pain, pleasure, ... stress may occur between two people having different intensities of understanding of a given issue - the issue being represented by some simplex, their respective understanding being traffic on that simplex. Where traffic can be measured, and not just spoken of in a qualitative way, terms in the following expressions may be evaluated numerically.

Given an initial pattern of traffic

$$\Pi_1 = \Pi_1^0 + \Pi_1^1 + \Pi_1^2$$

which changes to

$$\Pi_2 = \Pi_2^0 + \Pi_2^1 + \Pi_2^2$$

then the change in pattern would be given by

$$\delta\Pi = (\Pi_2^0 - \Pi_1^0) + (\Pi_2^1 - \Pi_1^1) + (\Pi_2^2 - \Pi_2^2)$$

thus the numerical values of

$$(\Pi_2^0 - \Pi_1^0),\ (\Pi_2^1 - \Pi_1^1) \text{ and } (\Pi_2^2 - \Pi_1^2)$$

are 0-forces, 1-forces and 2-forces, respectively. Note that the pattern of $\delta\Pi$ (although now referring to traffic as a force, and not as 'free' traffic) is of the same qualitative form as Π above, and thus further illustrates Johnson's (1981b) definition of traffic as a pattern. Atkin (1974) has likened changes in patterns on a static backcloth to a 'framework under stress' (the geometry of the framework is unchanged). It is not clear* however, whether a change in some pattern is a sufficient or a necessary condition (or both) for such change.

(vii) Forces on a changing backcloth

In many cases it is not sufficient to consider the possibility of changes in traffic on a static backcloth, but also changes in the backcloth itself. Beaumont and Gatrell (19) note that changes of the former kind may be seen to correspond to so-called 'slow' dynamics (and considered in the previous section), changes of the latter kind (to be considered in this section) to be so-called 'fast' dynamics. Since the backcloth is built from relations between sets, changes in the backcloth will arise from changes in the original relations. To seek the causes of such change is beyond the present scope which is concerned essentially with describing change rigorously and unambiguously (though Atkin, 1981 hints that the cause is likely to lie at higher hierarchical levels). Changes in the original relations will correspond to adding vertices to or loosing vertices from the simplices that originally made up the complex. It should not be difficult to appreciate that the addition or removal of vertices may profoundly affect traffic. For if a simplex of dimension 3, originally carrying traffic of grade 3 looses a vertex, its traffic must either move elsewhere (to another appropriate simplex of dimension 3), or change to traffic of grade 2. If traffic can readily accommodate such change, there will be a so-called free change ; however, this will often not be the case, in which case stress or some other type of force may arise (the feeling of exhileration when the final link in the understanding of a given problem is mastered? - traffic of understanding changes from grade n-1 to grade n).

*to the writer.

In the case of the above retailing example, change in the backcloth could depict the withdrawal of a particular type of retailing activity from a zone. Thus subsequent shoppers visiting this location would be forced to move on a leaner backcloth. The use of the word 'force' in this context can be seen as no accident: the nature of the traffic is determined by the underlying backcloth; when this changes, the traffic is forced to change. (The formal mathematics does no more than to describe the intuitive meaning of the word.)

More specifically given zones 1 and 3 with the following activities:

$$X_1 = \{A_3A_4A_{10}A_{11}\} \text{ and } X_3 = \{A_2A_4A_7A_9A_{12}\}$$

and activities 4 and 7 left each of these locations then zone X_1 would be said to experience a 3-force of repulsion - $K_1A_3A_4A_{10}A_{11}$ and a 2 force of attraction $K_2A_3A_{10}A_{11}$ and zone X_3 would be said to experience a 4-force of repulsion - $K_3A_2A_4A_7A_9A_{12}$ and a 2-force of attraction - $K_4A_2A_9A_{12}$(where $k_1k_2k_3k_4$ are, say, counts of shoppers). What we infer from this is of shoppers lowering their sights (they may or may not increase in numbers); they may or may not be more beneficial, eg. spend more money; they may or may not be the same individuals (if not, then they may go elsewhere thus slowing an induced movement of traffic through the complex). Thus if a simplex loses a vertex traffic must either move elsewhere or lower its sights. Atkin (1981) amusingly calls this the toothpaste syndrome: the loss of a vertex would correspond to squeezing the tube; toothpaste either remains where it is (under stress) or moves out (if the top is off). To give another example, if employment opportunities in a given town were reduced overnight from nine different job-types to six different job-types, people seeking employment would experience a 8-force (of repulsion) (and a reciprocal 4-force (of attraction)). The intensities of the individual forces could depend on the number of people affected. Forces associated with changes in $\hat{q}$ (top q) are often of particular interest.

Just as traffic can move through a structure (complex) so too can forces. In some cases, it may be desirable to disconnect the structure in order to inhibit the movement of forces(in the same way as it may be useful to inhibit movement of traffic). Other cases arise of traffic of too high a dimension for its structure: a housewife frustrated with too few opportunities being offered by her domestic environment; a shopper looking for 4 goods at a centre that provides only 3 (ie. t-force between actual and desired). Moreover if we can be more imaginative and envisage time as traffic on p-dimensional structures of events, then if the grade of time is higher than that of clock time, time may fly or drag, i.e. further stresses or forces may arise.

In some of his more formal and technical papers, Atkin (19) seems to have tried to suggest what the structure of a given backcloth needs to be like if it is to withstand forces. This has involved an examination of so-called coface operators, the technicalities of which are beyond the scope of the present paper (and the grasp of the present author). This particular aspect of his work however, does not seem to have been completely successful; see Griffiths and Evans (1982) for a reformulation.

(viii) Slicing and its significance for traffic

As well as changes in a structural backcloth that can arise from actual changes in 'traffic' in the real world system being depicted, different slicing parameters can also change the structural backcloth. These latter changes are, of course, of the analysts working and therefore, in a sense 'artificial'. Depending on the threshold levels for slicing that are chosen, different vertices and simplices are sliced out of the complex, and patterns of traffic must consequently adjust to the different configuration of the structure. Since changes in patterns can be interpreted as forces, slicing would appear to be an artificial way of generating forces (Johnson (1981b) represents these in terms of strain pairs). Thus, slicing changes the analysts 'view' of latent structure, and therefore of the forces that are 'seen'.

(ix) Discrimination traffic

In reviewing the static results of the Q-analysis algorithm in section 4 of this paper, it was suggested that Johnson (1981a) has formulated a certainrefinement of the Q-analysis algorithm for use in the context of clustering and discrimination work. The nature of the method was discussed in very brief terms above, for a fuller account, see Johnson (1981a) and Macgill (1982c). The consequence of applying this method is that the basic clusters that are found from the original Q-analysis algorithm are further dissected in various ways. This dissection in turn is determined by the value of numerical weights on individual vertices. Since these weights can be depicted in the general form of a graded pattern (see π under subheading (v) above), they meet the criteria given by Johnson (1981b) for being traffic. Since this pattern is used as a basis for discriminating simplices from each other (ie. further dissection of clusters). This he calls discrimination traffic, Johnson (1981a).

(x) Hierarchy

The idea of hierarchy was introduced as one of the fundamental building blocks of Q-analysis, and it is appropriate here to investigate its significance with regard to traffic. Atkin (1979) draws attention to the fact that 4-traffic at level N may be 0-traffic at level N+1. Thus, for instance, retail expenditure in zone 3 (cf. retailing example above) can arise as 4-traffic (5 minus 1). However viewing this retailing system as a single activity at a higher level of resolution (as one of a number of broader economic activities), retail expenditure in zone 3 would be defined on a single vertex, and therefore be 0-traffic. Thus the notion of hierarchy is bound up not only with the initial identification of structure, but also (and necessarily) with interpretation of traffic.

The natural dynamical hierarchy that arises when a given system is described over time has been spelled out by Atkin, as follows:

Time period	Dynamical level	
$T_1 > 0$	T_1 level	π constant in T_1, variable in T_2
$T_2 > T_1$	T_2 level	π " " T_2, " " T_3
$\vdots$		
$T_n > T_{n-1}$	T_n level	π constant in T_n, variable in T_{n+1}
T_∞	static level	constant (all time)

In the case of structures that support an endless circulation of traffic, Atkin suggests that entities or agencies with the power to control or intercept or terminate such circulation typically arise at some higher hierarchical level.

(xi) Cases of ambiguity over the backcloth structure for given traffic-probability

The idea of t-forces on n-dimensional structures as a social science counterpart to the idea of a force in physics may already see quite imaginative to the conservative world of the geographer. A further idea is given here. We have so far referred to traffic which occurred (spontaneously, or under force, or as a numerical measurement) on simplices and which in principle could be related unambiguously to particular subsets of vertices of those simplices. Further and yet more imaginative perspectives arise from identifying traffic on a backcloth for which there is ambiguity over the vertices that are (should be) relevant to the type of traffic that is trying to exist. Atkin (1978, 1981, chapter 6) argues that statements couched in terms of probability (for example of achieving a six with a single throw of a dice) are of this type. Thus he suggests that the outcome of a given throw may be depicted as traffic of dimension 0 (a single observation) on a backcloth of dimension 5 (six possible outcomes). He argues (1981 p. 173)

> "The description of a set of events by way of assigning them probabilities is a characteristic indication that the 'seeing agent' (or 'process of observation') is traffic of too low a dimension on the actual events ... thus when we find ourselves using the language of probabilities ... we have been unconsciously trapped into a method observation which constitutes low-dimensional (inadequate) traffic on some high-dimensional structure."

He argues further to suggest that as long as problems are dealt with via probabilities, a natural barrier will exist to inhibit full understanding, as probability will never address the full dimension of the structure involved.

> "When social scientists naively adopt statistical (probabilistic) methods to analyse their 'data', in the mistaken belief that they are thereby being 'scientific' in their methodology, then they are unwittingly admitting that the actual events which constitute their universe of observations are of such a high dimensionality as to be beyond the reach of their low-dimensional traffic of observation. But how can they hope thereby to attain an understanding of the actual structure of the universe? Indeed, the methodology based on the use of probabilistic notions is a barrier to the very concept of that structure."*
> (Atkin, 1981, p. 174)

*Conversely, we may say that the use of probability provides a useful way round the immense "observation" problem. Atkin's interpretation of probability is nonetheless a useful clarification of what is being done.

Yet worse may occur when successive 'events' are not fully connected (that is, when the latent structural backcloth is in pieces or awkwardly joined); as then observation traffic may find it impossible to move from one simplex to the next: it can only start again. In expounding this new theory of probability, he therefore invites some searching questions to be asked about the appropriateness of some of the most basic tools of analysis that are popularly used. Moreover, in recently formulating a theory of surprises (seemingly most intangible phenomena), Atkin (1981b) has shown that the language of Q-analysis cannot only be used as a basis for questioning the validity of past approaches, but also of offering some concrete alternative.

6. The language of Q-analysis

In the above discussion of Q-analysis, an effort has been made to be deliberately sparing of formal notation. This has been done in order to convey the flavour of the approach without drowning it in incomprehensible mathematical vocabulary and grammar. However, in order to pursue the approach further, some facility with the type of mathematics on which it draws is needed. Chapman (1981) or Beaumont and Gattrell (1982) are useful for basic introductions. Thereafter, see the work of Atkin. For the enthusiast, note that Griffiths and Evans (1982) have reformulated some of Atkin's original expositions on q-holes and face operators, the latter apparently having some shortcomings (though not of a kind that would seem to undermine the approach as a whole).

Atkin's development of Q-analysis appears to have derived much stimulus from the observation that the recognised modelling and statistical approaches hitherto used in geography (and the social sciences more generally) have not drawn on the full range of applicable mathematical concepts. In particular, the development within mathematics over the last hundred years or so which has recognised the pre-eminence of set theory over more traditional (Grecian!) tools of mathematical analysis has remained relatively untapped in the social sciences. Q-analysis provides a specific development of relatively recent ideas from set theory for the social sciences, which presents a new perspective for insight and analysis, and framework for thought, and in so doing, tools for tackling questions of a type that have previously been neglected.

It should be apparent from the material in preceding sections that Q-analysis is not an approach which applies some black-box methodology to preconditioned data in order to produce some conspicuous results. Rather, it has been developed as a language - a medium of expression and description - that aims to bring a degree of clarity (hardness) to the spectacles through which each of us (as observer or analyst) can view the world, and organisation to our minds in thinking about matters. Beaumont and Gatrell (1982) after Olsson (1980) reiterate the recognised philosophical point of view by noting that there may be an intimate connection between language and the things it can describe and say. As well as arguing that the use of mathematical language is necessary for handling complexity and the only source for languages that are capable of inquiry (as opposed to expression) Gould (1981a) writes particularly enthusiastically about Q-analysis as a language which can be free thinking from conventional and self-imposed straightjacket of mathematical approaches based on functions rather than more general relations.

All languages have their own syntax, grammar and phonology, and Q-analysis is no exception. With Pollock (1908) we may note the riddle of new languages before they are conquered, but their power afterwards. Many of the insights offered by Q-analysis are expressible in plain English. Ultimately, however, if these are seen to be useful some degree of fluency in the original would seem to be called for. As with foreign languages in everyday speech, much may be lost in any translation. Moreover, the power of a language depends on an understanding of it both by those who wish to use it and by their audience.

The language of Q-analysis has its roots in the branch of mathematics known as topology, a branch that has witnessed a meteoric rise over the last few decades. In developing Q-analysis Atkin has drawn on some basic topological notions - in particular certain aspects of the geometry of multi-dimensional objects and spaces, and with others, notably Johnson,added a number of refinements and extensions. The refinements and additions seem to have been necessary because standard pure topology does not provide enough precision for describing aspects of latent structure in social systems which Atkin believes might exist. (These beliefs come out more formally in his theories and philosophical papers). The language is still developing.* Griffiths and Evans (1982) is the most explicit paper available which relates Q-analysis to topology, though is not written for beginners. Armstrong (1981) provides an introduction to topology that might be within the grasp of some social scientists. (It may be of incidental interest to note here that the letter Q which is used within the title "Q-analysis" seems to come from standard topological notation; the letter is commonly reserved for denoting the dimension of a simplex in multi-dimensional space). The most significant long term contribution of Q-analysis may lie at least as much in its provision of an inroad for social scientists into a type of mathematics hitherto untapped (the qualitative mathematics of algebraic topology) as in specific applications of the type of individual concepts summarised above. In a more extreme argument it has recently been suggested to the writer that an alternative to using those aspects of Q-analysis that have so far been developed as a basis for exploiting topology, would be to explore topology more in its own right as a basis for generating insights into social systems.

7. Conclusions

It is the author's view that Atkin's development of Q-analysis constitutes a rare and valuable contribution to social science. This paper was begun several months ago without too many reservations about exploring a high-level language that appeared to offer a very general and yet precise basis for geographical inquiry. The explorations have not been disappointing. Chapman (1981) has called Q-analysis a bridge between the nagging paradoxes of generality and precision. Atkin (1978) has described it more simply as a hard language for the soft sciences.

---

* and demonstrates the rare phenomenon of social scientists developing their own mathematics rather than imposing onto their systems of study the mathematical forms and structures that have been developed by others.

As already noted, Atkin has argued in a number of his writings that some of the most significant breakthroughs in the physical sciences have arisen only after the phenomena being studied have been depicted (described) in what amount to hard (set-theoretic) terms. His thesis is that inquiry into social systems has often fallen short of invoking hard description in its representation of those aspects of the "real world" that are relevant in particular studies. Such description in turn can reveal a structure in multi-dimensional (social) space.* These appear to be revolutionary concepts (and may herald their own modest paradigm-shift c.f. Johnston 1979) and may even suggest that Atkin might be seeking to achieve in social science what Galileo and then Einstein achieved in physics (insofar as the very different environments permit reasonable comparison to be made at all). The very ambitiousness of this brave new beginning (a new culture or a new "gestalt" experience for the mathematically minded geographer?) makes it somewhat difficult to assimilate. An attempt will be made to appraise certain aspects in these conclusions.

The view that many existing theories and hypotheses that have been devised with the laudable aim of making social studies "scientific" are inevitably short-lived (because they are not based on hard description) raises the question of what the adjective "scientific" is taken to mean. Pirsig (1978) suggests a relaxed but, in the present author's view, quite telling definition of scientific, namely that it is to pursue the idea of not being fooled into assuming we know more than we actually know. Thus the adjective refers to the nature of an inquiry (meaning that it must aim to be hard) and not to its field, though there are admittedly more examples of hard (scientific) inquiry in the "physical sciences" than in the "social sciences". In this view of things then, hard description must come prior to hypothesis testing, theory building, verification and falsification, and the latter cannot expect to be able usefully to proceed without hard description.

---

* To recap, briefly, any two sets that are related to each other have an underlying structure (two simplicial complexes - one being the conjugate of the other) and this is turn is a backcloth for 'traffic'. The first few stages involved in invoking the Q-analysis framework are firstly to decide which sets are relevant in any particular context (often not a trivial exercise); secondly to recognise the richness that exists in the resulting web of inter-set relations (typically much richer than the one-to-one relations that are traditionally used), noting the absence or persistence or persistence of particular relations at different slicing levels; thirdly to recognise the resulting richness and diversity in the dimension and intensity of possible traffic. This then provides a powerful conceptual framework within which to pursue further enquiry.

The set-theoretic description that provides the foundation of Q-analysis generates its own type of representation of activity and in turn general theory for explaining the system being represented. This representation and theory of multi-dimensional space comprises individual representations and theories about particular aspects, (expressed in terms of of traffic). Relating these notions to the ongoing though often frustrated search for theory within social systems we see that not only is hard description presented as a necessary foundation for the longevity of theories, but as a useful pre-requisite for their development in the first place (due to its capacity to relieve the imagination from pondering over what is already known, and, within what is known, to separate the static (backloth) from the non-static (traffic).

Once formulated, the theories are themselves expressible according to set membership rules. Thus Q-analysis can be viewed as a comprehensive but essentially taxonomic meta methodology for categorising all manner of features of the real world into particular classes (elements of structure, types of traffic and the relation between structure and traffic; all three are expressible in set membership terms). The superficial monotony of taxonomy should not disguise power as a means of representing understanding (a similar sentiment would appear to underlie the quotation cited at the end of this essay).

Despite the insularity of Q-analysis from other mathematically based methods (the bibliographies of some Q-analysis papers cite little more than other Q-analysis papers!) and apparent competition with them, it is not most usefully seen to have been developed to oust existing methods; rather to complement them. One of the challenges to the contemporary scholar is to seek true understanding of the analytical approaches that are suited to his or her field of inquiry, and then to know how to choose between them. Although this is a phenomenal challenge, it is hopefully not too tall an order. For instance, it should take future generations of scholars less time to digest what current and past generations have found to be difficult new approaches, because of the emergence of various pedagogic guides and exemplifications. It should also be easier for future scholars to implement the methods, given the power and flexibility of recently developed computer hardware and software. One result of this is that data for a given system of interest can be more readily thought about and analysed in a variety of different ways. A further aspect in the case of Q-analysis is that new generations of scholars are more familiar (than their elders) with the so-called 'new maths' on which Q-analysis draws. In making these remarks, however, it is not suggested that analytical methods should be any more than a servant for pursuing stimulating scholarly inquiry and for addressing and, hopefully, alleviating substantive problems.

Indications that an initial insularity of Q-analysis from other approaches is gradually being reduced are seen in recent papers relating the approach to, for example, cluster analysis (Johnson (1981), Macgill (1982a, 1982c)) and graph theory (Earl and Johnson (1981). Further linkages can be sought in the future, for instance with discrete mathematical models based on difference and differential equations (relating, for example, to the methods explained in Wilson (1981)). Singularity theory, for example, may seek prediction of when (say) vertices of a particular latent structure may disappear: ideas about traffic developed in Q-analysis might then be invoked as a framework for deeper inquiryinto consequent effects. Moreover, whereas singularity theory when applied to fields of social inquiry may be susceptible to criticism from its lack of plausible laws (and hence difficulty in formulating a potential function, Casti (1981)) the theory building that Atkin has shown possible in his development of Q-analysis may lead to plausible and testable theories (falsifiable, in the Popperian sense?) and hence plausible laws.

The traffic (patterns, flows, forces and stresses) that arises in the multi-dimensional view of human affairs that the approach of Q-analysis generates may seem to suggest that quite a high level of analogy between the sciences of human and physical systems is being sought. It was noted above that any such analogy should be used only as a catalyst for further propositions, theories or laws which must ultimately be able to stand in their own right.* Much further exemplification in the context of human systems of the type of idea discussed in broad terms above under the heading of traffic would seem to be required before the utility of such analogy can be explicitly

* The analogy is more subtle than that hitherto invoked in the social sciences; the sense of the analogy is that 'flows' and 'forces' and other concepts which have obvious physical analogs arise out of the configuration of the structure that underlies the social systems being studied. A further basis for unification, perhaps?

demonstrated. One eventual finding might be the discovery of regularity in the configuration of backcloth 'types' and traffic 'types'. In the meantime it can be used at worst as a powerful conceptual tool.

The exploitation of particular aspects of the methodology of Q-analysis depends on the representation of appropriate aspects of the "real world" in set theory terms in the first place. The crucial test is that it must be possible to answer either "yes" or "no" to the key question of whether any particular entity belongs to any particular set. Initial ambiguity over the answer to such a question can often be removed by referring to sets at some finer level of resolution (see under proposition III). However, there are still likely to be instances where the answer "yes and no" can be given to the question of whether a certain entity belongs to a certain set. It should not be too surprising if these correspond to intrinsic questions of value. With apologies to scholars of language, we may stumble on the realisation that the inability to date to represent these in hard set theoretic terms defines a sense in which English is more versatile than mathematics. The latter can hold down the known; the former can substitute for the unknown. The answer to the key set membership question thus defines a boundary of the scientific (the time-honoured, but shifting, distinction between the classic and the romantic or the sciences and the arts?). Moreover, in defining a boundary in this way, energies may be more productively directed (eg. by not being unduly hampered by unnecessary semantic vagueness). It would be a great claim indeed if the distinctive exploitation of set-membership criteria in Q-analysis can make a significant inroad into the tremendous and tantalising interconnections between language, theory, truth and value.

Judgement of the extent to which this is a reasonable claim can await further work. In the meantime it is exciting to encounter the avant garde provided by Q-analysis; a holistic approach to data and inquiry. While not regarding Q-analysis as a universally applicable approach that can suit all tastes (if only people would learn about it), it should not be too surprising if some aspects of the approach lead to new insights in a wide range of individual fields of inquiry. If the need for "scientific" inquiry to be "hard" is accepted, moreover, the so-called quantitative revolution can with hindsight be seen to have heralded a pseudo-scientific age. The post-quantitative age exploiting the power of qualitative mathematics (the bad climate generated by blanket criticism of any mathematically based approach notwithstanding) may invoke a more practically adequate notion of the scientific. It is the author's belief that various aspects of Q-analysis provide some sound guidance for knowing what we should be searching for in claiming social science to involve a scientific inquiry into the social world, the deep locked mysteries of art and humanity notwithstanding.

"The first thing to be observed about descriptions given in terms of lists of items (sets?) is so obvious that you have to hold it down or it will drown out every other observation. This is: It is just duller than ditchwater ... But if you can hold down that most obvious observation, some other things can be noticed that do not at first appear. First, such information may be impossible to understand unless you already know something about the real world system it purports to represent. Second the objects exist independently of any observer. Third, a minimum of value judgements have been expressed. Fourth (and perhaps most powerfully), there is a knife moving about here. A very deadly one; an intellectual scalpel so swift and so sharp that you may not see it moving. You get the illusion that all those parts are just there and are being named as they exist. But they can be named quite differently and organised quite differently depending on how the knife moves .... It is important to see the knife for what it is and not to be fooled into thinking objects of inquiry are the way they are just because the knife happened to cut them up that way." Pirsig (1978) pp. 71-72.

References

Atkin, R.H. (1965) Abstract Physics. *Nuovo Cimento 38*, 496-517.

Atkin, R.H. (1971) Cohomology of observations in T. Bastin (ed.) *Quantum theory and beyond*, Cambridge University Press, pp. 191-211.

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

Atkin, R.H. (1975-1977) The Methodology of Q-analysis Research Reports V, VI, VII, VIII, IX, X. University of Essex, Colchester.

Atkin, R.H. (1977) *Combinatorial connectivities in social systems*, Birkhauser.

Atkin, R.H. (1978) A hard language for the soft sciences. *Futures 10*, 492-499.

Atkin, R.H. (1978a) Time as a pattern on a multi-dimensional backcloth. *Journal of Social and Biological Structures 1*, 182-295.

Atkin, R.H. (1978b) Dynamics of patterns on a hard data backcloth. R.L. Martin, N.J. Thrift and R.J. Bennett (eds.) *Towards the dynamic analysis of spatial systems*, Pion, pp. 19-36.

Atkin, R.H. (1981a) *Multidimensional man*, Penguin.

Atkin, R.H. (1981b) A theory of surprises. *Environment and Planning B, 8*, 359-366.

Beaumont, J.R. and Beaumont, C.D. (1982) A comparative study of the multivariate structure of towns. *Environment and Planning B, 9*, 63-78.

Beaumont, J.R. and Gatrell, A.C. (1982) An introduction to Q-analysis. C.A.T.M.O.G. 34, Geoabstracts, Norwich.

Casti, J. (1981) Topology and crisis. *Options, 3*, p. 4-6.

Chapman, G. (1981) Q-analysis. in: *Quantitative Geography : a British View*, eds. N. Wrigley and R.J. Bennett, Routledge and Kegan Paul.

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

Earl, C.F. and Johnson, J.H. (1981) Graph theory and Q-analysis. *Environment and Planning B, 8*, 367-392.

Everitt, B. (1980) *Cluster analysis*, Heineman.

Fischoff, B., Lichtenstein, S. and Slaic, P. (1981) Lay foibles and expert foibles in judgements about risks. in: T. O'Riordan and R.K. Turner (eds.) *Progress in Resource Planning and Environmental Management* III, Wiley.

Flemming, V. (1981) Structure in bungalow plans. *Environment and Planning B, 8*, 393-344.

Gatrell, A.C. (1981) On the structure of urban social areas : explorations using Q-analysis. *Transactions of the Institute of British Geographers, N.S.6*, 228-245.

Goddard, J. and Kirby, A. (1976) An introduction to factor analysis. C.A.T.M.O.G. 7, Geoabstracts, Norwich.

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.

Gould, P. (1981) Electric Power Failure : reflections on compatible descriptions of human and physical systems. *Environment and Planning B, 8*, 405-418.

Gould, P. (1981a) A structural language of relations. in: *Future trends in Geomathematics*, (eds.) R. Craig and M. Labovitz, Pion, London.

Gould, P. (1981b) Letting the data speak for themselves. *Annals of the Association of American Geographers 71*, 166-176.

Gould, P. and Johnson, J.H. (1978) The structure of television programming : some experiments in the application of Q-analysis. University Park, Pa : International Television Flows project. Cited in P. Gould (1980) op.cit.

Griffiths, H.B. and Evans, P.J. (1981) Atkin's Q-analysis : a critique and reformulation. SSRC end of Grant Report (mimeo).

Harvey, D. (1973) *Social justice and the city*, Arnold.

Johnson, J.H. (1981b) Some structures and notation of Q-analysis. *Environment and Planning B, 8*, 73-86.

Johnson, J.H. (1982) Q-transmission in simplicial complexes. *International Journal of Man-Machine Studies*.

Johnson, J.H. (1981) Q-discrimination analysis. *Environment and Planning B, 8*, 419-434.

Johnson, J.H. and Chapman, G. (1979) Television on Coding Manual. Cambridge : International Television Flows project (European Region). Cited in P. Gould (1980) op.cit.

Johnson, J.H. and Wanmali, S. (1981) A Q-analysis of periodic market systems. *Geographical analysis, 13*, 262-275.

Johnston, R.J. (1979) *Geography and Geographers : anglo-American geography since 1945*. Edward Arnold.

Klecka, W., Nie, N.H. and Hull, C.H. (1975) *The SPSS primer*. McGraw-Hill, London.

Lyons, J. (1977) *Chomsky*, Fontana/Collins.

Macgill, S.M. (1982c) A consideration of Johnson's Q-discrimination analysis. *Environment and Planning B, 9* (3) (forthcoming).

Macgill, S.M. (1982b) Exploring the similarities of different risks. Working Paper 344, School of Geography, University of Leeds.

Macgill, S.M. (1982a) Cluster analysis and Q-analysis (forthcoming).

Olsson, G. (1980) *Birds in Egg*, Pion.

Pinkava, V. (1981) Classification in medical diagnostics : on some limitations of Q-analysis. *International Journal of Man-Machine Studies 15*, 221-237.

Pirsig, R.M. (1978) *Z en and the art of motorcycle maintenance*. Corgi.

Pollock, F. (1981) *Clifford's Lectures and Essays, Vol. 1.*

Reichman, W.J. (1978) *Use and abuse of statistics*. Pelican.

Roberts, F.S. (1976) *Discrete mathematical models*. Prentice Hall, Englewood Cliffs.

Rushdie, S. (1981) *Midnight's children*.

Russell, B. (1946) *History of Western Philosophy*. Allen and Unwin.

Spooner, R. and Batty, M. (1981) Networks of urban systems analysts. *Environment and Planning B, 8*, 449-475.

Stewart, I. (1975) *Concepts of Modern Mathematics*. Penguin.

Williams,K. (1971) Do you really want to be a factor analyst? *Area 3* 228-9

Wilson, A.G. (1981) *Catastrophe theory and bifurcation : applications to urban and regional systems*. Croom Helm.