the structure and representation of criminal actions

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by Jamie Lee

March 1998

acknowledgements

This thesis has been helped immeasurably by innumerable people, from whom I have received an education in all senses of the word.

Firstly, I would like to thank my supervisor David Canter for giving me regular doses of inspiration and perspiration in varying proportions. I would also like to thank Ian Donald and Tony Coxon for their significant contributions to this thesis. I am grateful to Donna Youngs and Stuart Kirby for the use of their data sets. Also thanks go to my colleagues throughout the department, including Lol, Duncan, Shadd, Nick and Steve, and the secretarial staff.

I'd have to write another thesis to explain my acknowledgements to Kate Fritzon, so here is a summary: thanks.

Finally, I would like to pay tribute to the support from my whole family. In particular, thank you to my mother and father for taking me from GOS to PhD.

abstract

The domain of criminal actions contains numerous biases and unreliabilities. Recently the faceted approach to research (eg Shye and Elizur, 1994) has been sucessfully applied to investigate the domain (eg Canter and Fritzon, 1998). However, key methodological issues have not been addressed. The thesis assessed structural hypothesis testing using non-metric representations on the domain. Three criminal actions data sets with various challenges were used to test the faceted approach to analyse the domain. Improved recovery of structure was found with faceted representations through understanding the impact of methodology. Key findings include: the significance of the association or correlation measure; the need for item design to reduce exclusivities; the importance of reliable representation of items in local geometric subspace; the importance of the modulating facet; the predisposition of factor analysis to misrepresentation; the link between regional structures and Guttman scales.

contents

CHAPTER

CHAPTER

CHAPTER

CHAPTER

CHAPTER

illustrations and formulae

TABLE

TABLE

TABLE

FIGURE

FIGURE

FIGURE

FIGURE

FORMULA

empirical studies

EMPIRICAL STUDY

EMPIRICAL STUDY

Chapter 1
the structure and nature of information on criminal action

> To look at the results of a "blind" analysis after the fact and then proclaim that they "make sense" is hardly an objective scientific procedure. (Guttman, 1985, p. 53)

The Structure of Criminal Actions

The structure of a domain denotes the constituents of that concept and any relationship between the constituents. A structural analysis attempts to establish the constituents of a concept and then to understand what - if any - relationship exists between the constituents. The constituents of a human action are behaviour and intention, which are independent but jointly necessary to denote an 'action'. Thus an individual can be said to have performed an action if that individual behaved so as to get closer or achieve an intended end. Intention therefore implies a conscious understanding of the outcome or likely outcome of a behaviour, so behaviour has meaning to the individual.

By this rationale, criminal action consists of behaviour and intention that are against the law in that given situation. For example, a man talks to and gives a lift to a woman hitchhiker, making her at ease but also vulnerable, and providing an opportunity to rape her. The behaviour of picking the woman up itself is not illegal, but the intended goal is. A criminal action is thus behaviour done with the intention of achieving some goal, and where the goal, the behaviour or both is illegal. Consequently, the structural analysis of action - and criminal action in particular - must operate at two broad levels: the behavioural and the intentional.

The two strategies distinguishable for these two levels of analysis are the quantitative approach at the behavioural level and the qualitative approach for intentional level. In the former, the emphasis is on behaviour displayed by the actor and observed objectively, while in the latter the emphasis is on discovering the intention and meaning of the events to the actor.

Quantitative and qualitative approaches differ in how the research conceives of social reality (Bryman, 1988). The positivistic quantitative analyses would only seek to measure observable and replicable behaviours, without reference to any meaning attached by the actors to their actions. The reality of the situation would be defined, constructed and construed by the researcher and would not include the subjective intention of the actor. Postpositivistic qualitative analyses on the other hand would attempt to understand the significance and the reality of the event to the individual, asking directly how the situation was perceived by that individual.

At the same time that the 'what' or the ontology of qualitative and quantitative research is different, Bryman (1988) also stated that the 'how' or the epistemology was different. Epistemology refers to the methodology employed to gather information, though Jupp (1989) cautioned that in practice the boundaries between quantitative positivistic research and qualitative postpostivistic research were blurred. Though the two are used to complement each other, research broadly focuses on the one or the other.

Michell (1990) suggested that the undertaking of the first quantitative studies are generally recognised as when 'real' psychology started. These studies assumed that the same methodological approach and empirical ideals from the natural sciences were also applicable to the social sciences - the 'methodological monist' notion (Bryman, 1988). The objectivity of the style of observations were geared towards to independent researchers finding identical results, giving it replicability. These quantitative approaches tend to involve large scale analyses using samples that can be assumed to be representative of the population and the formulation of mathematical laws.

Although it too has 'a long and distinguished history' dating back to the 1930s (Denzin and Lincoln, 1994), qualitative research is distinctly more contemporary. Miles and Huberman (1994) stated qualitative studies had a 'phenomenal' expansion in qualitative inquiry during the 1980s. Miles and Huberman suggested that the strengths of qualitative research included the ability to place the domain in its naturalistic setting; the richness and complexity of information gathered; the flexibility of methods; and their understanding of the meanings of the situation to the actors. Bryman (1988) suggested that designs to gather the two types of information on behaviour and intentions typically include those of the social survey and experimental designs for quantitative research, and unstructured interviewing for qualitative research.

For example, in areas such as attitudinal research, it is possible to make a choice between qualitative or quantitative strategies according to the applications of the research, the size of the sample, etc. For example, quantitative research could be large scale surveys of preference for political parties to predict the next government, as opposed to qualitative focus groups exploring how the public perceives different political manifestos. It is argued in this thesis that in the analysis of criminal action, one key criterion for research is the interaction of the actual structure of the domain and analytical representation of that structure. In other words, the structural analysis of behaviour and intention may be restricted by issues external to the ideal of the researcher. As Bryman pointed out: 'preferences for one or the other or some hybrid approach are based on technical issues' (Bryman, 1988, p. 5).

For the domain of criminal action, these 'technical issues' include the availability of information, the nature of the information and the possible research designs given the information. These issues mean that the information tends to be unreliable on account of random 'noise' and systematic bias.

The Phases of the Research Process and Secondary Information

The first major restriction on researching criminal actions is the source of the material from which information is gleaned. Data on crimes, criminals and victims must be gathered ethically, with due consideration of its sensitivity. For this reason alone, existing information sources should be used wherever possible, even though it may introduce error into the research. Valuable information on criminal actions may have been gathered for purposes other than for a particular research project, which is termed 'secondary data'. Examples of 'secondary data' sources given by Jupp (1989) include diaries, letters and newspapers. However, there actually exist different types of what should be termed *secondary information*, of which Jupp's 'secondary data' examples form only a subset. To demonstrate the differences between the types of secondary information, it is necessary to deconstruct the research process to see its constituent parts.

A powerful model describing the stages in researching a domain such as criminal actions was given by Coombs (1964), termed here the 'Coombsian Research Model' (CRM). The CRM describes processes common to both quantitative and qualitative research paradigms, although the stages are more distinct and recognisable in quantitative research. The CRM is given in Figure 1.1.

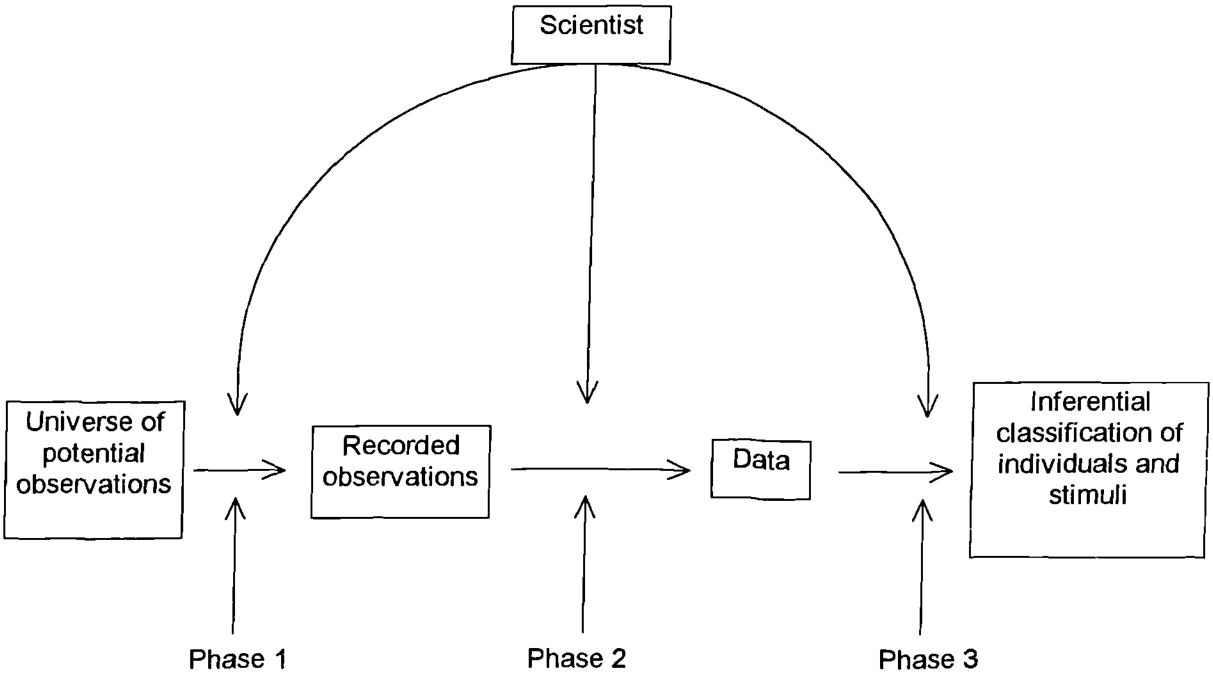


Figure 1.1. Coombsian Research Model (from Coombs, 1964)

As the research progresses from the real world of possible observations to inferential classifications in the CRM, there are distinct phases at which the researcher makes crucial decisions. The superset of the whole universe of possible observations would cover all the actions of the objects of interest. The first phase is when a sample of observations from the universe of all possible observations is chosen to be relevant, which become recorded.

Phase 2 involves the translation from recorded observations into data or:

> a classification of observations in the sense that individuals and stimuli are identified and labelled, and the observations are classified in terms of a relation of some kind between individuals and stimuli, or just between stimuli. (Coombs, 1964, p. 5)

The meaning of 'data', in the sense intended by Coombs, is therefore restricted to observations that have been abstracted by some low level classification. Phase 3 concerns the detection of order and relations amongst the data obtained by phase 2, and the conclusion of structure in the recorded observations.

Coombs asserted that the decisions at phase 2, such as the focusing on the relations between stimuli and individuals or just between stimuli, necessarily restrict those conclusions possible at phase 3. The conclusions and theories are built on the same observations but the data may be different. Extrapolating from this, we see that *real world actions* are recorded as *observations*, which are coded as *data* and then analysed into *classifications*. Between each of these four distinct entities are the three phases, where a filter is introduced by the researcher, limiting the possibilities.

The final phase of developing classifications is typically less marked in qualitative studies, where information tends to be left in a more raw form. Nevertheless, the material is still used to support an argument or disprove hypotheses.

It is necessary to expand the CRM to allow the exploration of the nature of secondary information in the domain of criminal actions. This is because Coombs neglected phase 1 of the CRM, stating that this first phase was 'perhaps the most important of all ... [but] beyond the scope of this theory' (Coombs, 1964, p. 5). In short, the decision at phase 1 is to ask the question 'Which responses count as valid'; phase 2 is 'What is a valid way to represent respondents and/or responses'; and phase 3 is 'What is the most valid way to analyse relations among respondents and/or responses'.

But prior to deciding at phase 1 of 'Which responses count as valid', a logically prior question must be asked. This question is 'Which stimuli should be used to elicit responses from respondents'. As Ackroyd and Hughes (1992) stated: 'data are always the result of a selection from what can possibly be said about some phenomenon' (p. 5). The operative word here is 'said', which should really be 'asked and answered'.

Therefore an additional phase for the CRM is necessary which is concerned with the universe of possible valid questions as opposed to the concern of phase 1 over the universe of possible observations. To maintain comparability with Coombs (1964), this is to be termed phase 0. Therefore phase 0 is a consideration of what constitutes a statistically and substantively valid stimulus from the universe of all possible stimuli. Phase 1 is therefore a consideration of what constitutes a valid response to those stimuli, as taken from the universe of all possible responses to that stimuli. After all, it is hardly worthwhile recording any response where there is no meaningful and reliable stimulus. This was intended for example by Borg and Shye (1995) who stated that a research 'item' was a stimuli question and the valid response to it (p. 20).

The expansions to the CRM to create the revised CRM are illustrated in Figure 1.2.

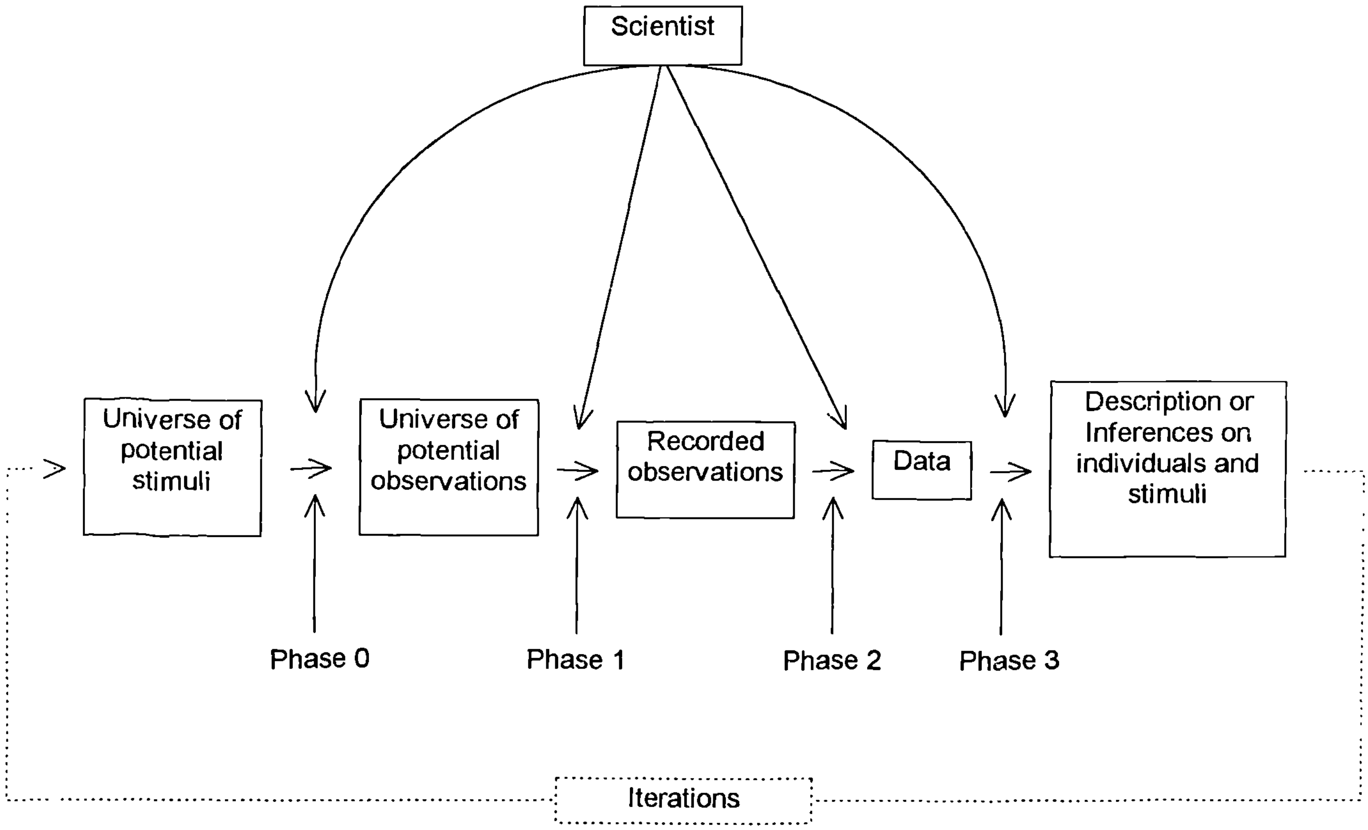


Figure 1.2. Revised Coombsian Research Model (adapted from Coombs, 1964)

First attempts at suggesting the structure of the domain are made at phases 0 and 1. Formal hypotheses of structure could be proposed at this point, which are then used to organise the data gathering procedure. Conclusions are then made using the representation of the domain as a guide to its structure, a representation which may be graphical, geometric or simply a single test result.

Figure 1.2 adds an important connection between the stages of the research process missing from the original CRM in Figure 1.1. This is that information can be 'looped' back so that classifications or inferences feed back into the construction of new items to gather more information. Furthermore, the same information can be used in a variety of ways. Information used in one research project can be used in another by starting at the lowest possible or desirable point in the CRM. For example, field notes on a phenomenon that have been analysed in one way can also be used in another way.

However, this model has so far assumed that the same researcher makes all the decisions at all four phases. When the decisions are made by different people the same person or for the same rationale then the research uses what is termed 'secondary data', or what will be termed 'secondary information' in the Coombsian sense. Thus field notes might have been taken by one person who made decisions at phase 1 about the relevant observations. These notes might be used by a different researcher to create data to test hypotheses about structure in classifications. The second researcher therefore uses the notes as *secondary recorded observations*, according to the CRM. *Secondary data* in the Coombsian sense would imply that the material has already been coded and stored in a form ready for analysis, such as a data matrix. Information in the form of existing classification schemes, perhaps as part of a meta-analysis or higher-order factor analysis, would be *secondary classifications*.

But what ties each of the forms of secondary information together is the fact that some person *external* to the research process has asked the fundamental question 'What are the relevant questions that must be asked about this phenomenon for my purposes?' In other words, a subset of the whole content universe has been sampled to create the secondary information. Therefore at whatever stage the secondary information is used for research, it must be realised that there is only a *partial content universe* under examination, i.e. the domain is not complete and has already been partially sampled. Even if the information were to be used for another purpose, it must be recognised that the information has an inherent bias in it on some external rationale. Research must acknowledge and work within this biased framework, unless the external rationale was to sample randomly from the content universe of possible items describing some phenomenon - which is unlikely. The 'partial' in the partial content universe has twin meanings of firstly that the content universe is not complete and secondly that the content universe is biased.

The relevance of the expanded CRM for understanding criminal actions is explored in the next section.

Secondary Recorded Observations on Criminal Actions

One example of secondary recorded observations which contains very detailed qualitative information on criminal actions is police and Crown Prosecution Service (CPS) case files. These files typically contain such material as crime scene photographs or videos, victim and witness statements, forensic pathology reports, etc. Additionally, in solved cases where someone has admitted to the offence or has been found guilty of it, offender information from interviews, statements and previous convictions is also found.

In these case files, phases 0 and 1 of the expanded CRM in Figure 1.2 have been completed by someone with a different rationale to the researcher re-examining the case files. That is to say, the investigating officer in the inquiry has decided what aspects of the crime, the description of the offender, the victim's statement and any witness statements are relevant to the investigation, namely phase 0. Useful responses to these lines of enquiry have been selected and recorded either contemporaneously, such as by interviewing the witness, or sequentially, by following up lines of enquiry, namely phase 1. The explicit rationale of this activity is to gather enough information to identify a suspect and then to produce evidence to charge that person. This can be shown using the example of an allegation of sexual assault.

In sexual assaults, the survivor is also typically the main and only witness who can give a full account of the crime. If the sexual assault survivor were to be encouraged to recall what using Cognitive Interview procedures (Fisher, Geiselman, Raymond, Jurkevich and Warhaftig, 1987) it is actually the assault survivor who determines the universe of stimuli and the responses. This is done by his or her internal processing of the event, asking questions such as 'What happened next?', 'What was the attacker saying?', etc. and then answering these questions as part of a narrative reconstruction. Where this is not the case, or for interviews with witnesses or suspects for other crimes, it is the investigating officer who selects the questions or stimuli surrounding the offence which are then responded to by the offender in a satisfactory way or not. In either case, the person is external to any research endeavour.

This process turns into an iterative negotiation over the reality of the offence, with or without the offender. The important feature of this process is that of the universe of possible things that can be known or asked about, only a few are recorded as observations. In other words, the content universe for that particular offence becomes a partial content universe. This sub-universe of actions will have 'patches' missing and inconsistencies.

The rationale for information recorded in crime files is for evidence. Inference of intention may be noted as motive, though the noting down of behaviour *qua* behaviour as a psychological phenomenon is done to a lesser extent. Additionally it may have unreliability on account of for example ambiguous *post mortem* results or poor eyewitness recall. Therefore several sources of systematic bias exist in the data.

For example, Canter and Heritage (1990) used statements taken from women raped by strangers to analyse offender behaviour as a first step towards developing a model of rape for 'offender profiling'. The rape statements were clearly secondary information (*viz.* secondary recorded observations) and were collected for a purpose other than research. These statements were taken by police officers interviewing the assault victims, who thereby determined the sub-universe of recorded observations and gathered suitable information to identify the offender and prosecute him. Therefore the recorded observations are details such as the movements of the rape survivor prior to the attack, the clothing and description of the offender, the direction in which the offender went, etc. In other words, the universe of stimuli, the universe of potential observations and the recording of these observations has been already decided trimmed by an external party. Therefore the decision model under the expanded CRM is actually done on a subset of the universe. Each attribute in the data matrix is a partial measure of the domain under examination; each case in the data matrix is included for some external reason.

The statements or similar material can be content analysed to achieve a carefully structured design. Content analysis adds reliability and robustness to qualitative material, and has been used in a variety of disciplines. Berelson (1952; cited by Weber, 1990) defined content analysis as a 'research technique for the objective, systematic and quantitative description of the manifest content of communication'. Krippendorff defined it as a 'research technique for making replicable and valid inferences from data to their context' Krippendorff 1980, p. 21).

Referring to the expanded CRM, content analysis has been used by many researchers through to phase 3, namely making inferential classification. A data set is used in chapter 7 in which classifications were created using a similar content analytical process in all the stages of the research. However, it is not necessary to always go that far, and was used by Canter and Heritage (1990) to encompass phases 0 and 1 only.

If material can be structured from first principles then content analysis is not required, such as in Fisherian experimental design. However, where the material cannot be structured in such a way then it becomes useful, and indeed Krippendorff (1980) cited this as one of the specific advantages of the method. Given the distinction of action as behaviour and intention, the content of the material that is being examined is a report of the behaviour, leaving inference of intention in the structural hypothesis of the theme, according to this thesis. Furthermore, the coding of the qualitative material into dichotomous presence or absence is all that can be expected at this level especially when unreliable sources of data such as on criminal actions.

The qualities of a good content analysis include ensuring that the categories collectively exhaust the domain and that the levels in the categories are mutually exclusive and exhaustive (e.g. Weber, 1990, p. 23). But as with most methods, error present at the start of a content analysis is carried forward and Krippendorff (1980) pointed out error may even be magnified. Where a research design does not follow the code imposed by properties such as exclusivity or exhaustiveness, error is introduced. With the use of content analysis as an approximation to faceted design, the question naturally arises as to whether or not the error is magnified at the level of inferential classification.

Attrition and Self-Selection in Criminal Action Information

The second major restriction in researching criminal action that methodologies must take into account is that all information is retrospective and selected. The information is retrospective for the simple reason that an action only becomes criminal after someone recognises that an act was against the law and recorded it as such. People defined as criminal are done so on the basis of their having committed or admitted to criminal acts. It almost goes without saying therefore that the sample of criminal acts and criminals is self-selected, which already runs counter to the quantitative ideal of random selection and assignment to cases. Without a random sample, it is not possible to estimate with complete certainty population parameters and make inferences about the population based on the sample. The representativeness of any information on criminal actions is extremely limited.

Using again the example of sexual assault against women, O'Connell-Davidson and Layder (1994) pointed out that there are several key criteria that tend to characterise those cases which are prosecuted. These include an assault by a stranger, the use of violence during the assault and the prompt reporting of the assault. To these can be added further investigative considerations to ensure a conviction, including the recovery of DNA, the offender being local and being known to the police. This means that the proportion of actual sexual assaults to those that are successfully prosecuted is due to attrition in the criminal justice system. A data set featuring these problems is presented in chapter 8.

The difficulty of using official police statistics on other crimes is also widely noted (e.g. Maguire, 1995; Jupp, 1989; Lea and Young, 1984). A variety of reasons for not reporting crimes create the 'dark figure' of crime, with reasons including, such as unwillingness to report offences or the low expectation of the police (Sparks, Genn and Dodd, 1977; Hough and Mayhew, 1983).

Other sources of information that overcome this self-selection and attrition include victim surveys, where people are asked about their experience of crimes over a set period. Within these there is a qualitative/quantitative distinction. Broadly speaking, the more quantitative the approach is then the less sensitive it is, in both statistical and substantive meanings of the word sensitive. Large scale door-to-door surveys of crime such as the British Crime Surveys (e.g. Mirrlees-Black, Mayhew and Percy, 1996) may not be sensitive to respondents and thus crimes against the person - especially sexual offences - will be under-reported. However, more sensitive interviewing using for example less structured questions and matching on gender and ethnicity has led to greater willingness to report sexual offences in victim surveys, such as in the Islington Crime Surveys (e.g. Jones, MacLean and Young, 1986).

A drawback with the more qualitative unstructured interviews in these surveys though is that the information is not useful for investigative purposes, since the surveys were mainly for local authorities for policy-making purposes. The questions asked for instance about the reaction and sympathy of the police authorities, the effectiveness of local services such as Victim Support, etc. Thus the sample sizes were smaller even if the information was highly valid. Furthermore, the accuracy of accounts of crime by eyewitnesses and survivors decreases with time. This means that unstructured interviews may have this additional source of error if the crimes happened a long time before the interview. Although the most accurate statements will be those statements taken soon after the crime - usually those given to the police - these will suffer from the biases outlined previously. A data set with these problems is introduced in chapter 12.

These surveys that ask people for their experience of having committed crimes rely on the honesty of the respondents - more so than victim surveys. Because of social desirability, there may be systematic bias where respondents change the details of the offence, conceal their commission of offences or even exaggerate the number or severity of crimes committed.

Finally, the linkage of series of offences to a single offender may not be spotted by the police. This may mean that many offences in a sample may be artefactually more highly related together, assuming offender consistency, or alternatively offences may be less highly related together due to offender inconsistency. It is possible of course for some offenders to be consistent and others to be inconsistent, resulting in bias from both directions.

Analysing Intention in Criminal Actions

In criminal actions, behaviour may be bizarre or the crimes may seem 'motiveless'. This raises the issue of whether criminal actions can only be analysed in terms of objective records of behaviour, or inferences about the underlying intention. The meaning of the behaviour of a bizarre crime may be restricted only to the actor. Using the terms of Pinizzotto and Finkel (1990), given only the behavioural traces left by an individual such as a murder crime scene - the 'WHAT' of the offence - is it possible to discover and use the 'WHY' of the crime to find the 'WHO' or the offender. Or indeed is it even necessary to find the 'WHY'?

One approach to analysing criminal actions associated with the FBI, termed the motivational model (Alison and Lee, in press), holds that the motivation to commit serious sexual assaults 'has its origins in fantasy' (Burgess, Hartman, Ressler, Douglas and McCormack, 1986, p. 252). Indeed, it was claimed that the 'central role of daydreaming and fantasy ... is critical to what motivated [the offenders] to kill' (Burgess *et al.*, 1986, p. 256). Motives for otherwise 'motiveless crimes' can be inferred by observation of the behaviour or understanding of the behavioural traces. For example, Prentky, Burgess, Rokous, A. Lee, Hartman, Ressler, and Douglas (1989) claimed that 'acts such as peeping or exhibitionism serve to cultivate new secret experiences, which not only activate fantasy but provide the incentive (or motive) for playing out the fantasy' (Prentky, *et al.*, 1989, p. 890). In other words, sexual variation and violent fantasy precede murder. What is meant by 'fantasy' by the FBI is that it is an intention that has *yet* to be put into action is properly termed a plan or conspiracy.

A simple refutation of the FBI view that 'fantasy is equivalent to a plan' could be the number of women having 'rape fantasies' or similar violent fantasies (e.g. Friday, 1975), none of whom intend to become rape victims. Similarly the fact that some but not all paedophiles become child sexual abusers by 'acting out' their fantasies of sex with pre-pubescent children (e.g. Howitt, 1995) demonstrates the difference between sexual fantasy (i.e. without intention) and a plan (i.e. with intention). An alternative empirical example is during a sexual assault when the attacker asks his victim if she is enjoying herself. Such request for verbal participation in the assault occurred 15% of the time in a sample of rapes by serial attackers

(Canter and Heritage, 1990). In this instance, the intention of the rapist who requires verbal participation may be an attempt at gaining intimacy through exploitation of a differential power relationship (e.g. Canter, 1994), or it could denote the rapist asking a rhetorical question to reinforce his dominance over the victim.

The important aspect of intention in the context of the analysis of criminal action is in the nature of the measurement of behaviour and intention. Behaviour may be measured quantitatively without reference to the actor. By contrast, intention may only be measured qualitatively *with* reference to the actor. The true arbiter of intention is the individual alone. It is suggested that any inference about the intention of a criminal action must be modest, and cannot reliably be gauged given the nature of criminal actions information, as outlined above. This is particularly salient when the actor is unknown, as is the case with classifications according to FBI criteria (Lee, in press). But given many behaviours empirically shown to be repeatedly co-occurring then some inference of theme is possible. This would require a multivariate quantitative approach.

Summary of Chapter 1

This chapter illustrated some of the problems of research into criminal actions. Criminal action was split into components of behaviour and intention. It was suggested that analysis of criminal action may operate on two levels, which are associated with quantitative and qualitative approaches respectively. Limitations on research include the secondary nature of much otherwise valuable information and the skewed sampling of all possible crimes, criminals and victims on which information is held. Another problem illustrated with the revised Coombsian Research Model was how decisions made by people external to the research process create a partial content universe. The restrictions on a more qualitative understanding of criminal intentions were highlighted using the motivational model. Methodology must be suited to the task, which may require formal consideration of the technical issues and their impact on substantive models.

Chapter 2
representations of action

Representation and Models of Measurement and Classification

The previous chapter suggested that the inference of intention from secondary sources of criminal actions was unreliable, but that the quantitative summation of behaviour could allow inference to intention. It was further suggested that this was due to the nature of information. Nevertheless, many categorical typologies exist with which to classify serious offenders such as rapists in a clinical setting (e.g. Groth, 1979). This type of classification model has been applied to live police investigations - where the offender is unknown and not in the clinic (e.g. Ressler, Douglas, Burgess and Burgess, 1992). Such typologies assume that a great deal of complexity can be reduced into discrete mutually exclusive types, as contrasted with measurement models which attempt to use continuous dimensions or themes to represent the complexity parsimoniously.

The representation of actions concerns the modelling of observations made on a structure. This can be done in a variety of ways, one of which follows from the distinction introduced in chapter 1 of the quantitative and qualitative approaches to researching structure. Quantitative research seeks representations in the form of numerical measurement models while qualitative research seeks representations in the form of categorical classification models. The fundamental underlying difference is that measurement requires comparability on some latent or manifest dimension, whereas classification is based on equivalence or non-equivalence (Ghiselli, Campbell and Zedeck, 1981).

Jacoby (1991) indicated that the process of measurement was one of theory testing, where support for the model - though not conclusive proof - is found when the evidence suggests the model is a valid abstraction of the real world. It then holds sway, at least until a better fitting or more general model is proposed. However, falsification of the measurement system occurs 'whenever the specified properties of the real number system do not correspond to the empirical property under investigation' (Jacoby, 1991, p. 6). As Roskam put it, '[m]easurement and scaling is neither a condition nor an objective of scientific research, but rather a carrier of theory' (Roskam, 1981, p. 207). Indeed, Guttman stated that where possible he avoided using the term 'measurement' for fear that associations conjured up by it may be 'barriers to progress in theory construction and in research' (Guttman, 1971, p. 330).

Coombs, Dawes and Tversky (1970) stated that solving the problems of measurement gave the 'ability to isolate critical properties for an experimental investigation and to reveal structure that underlies a given numerical representation' (Coombs *et al.*, 1970, p. 30). The first problem was the 'representation problem' which meant that for the model to be meaningful, there must exist a strong correspondence between the relations amongst the empirical observations and the relations amongst the numerical system - in other words, whether the system can be measured. The uniqueness problem was where the 'status' of the scale must be discovered by examination of what could count as permissible operations, with the emphasis on the descriptive rather than proscriptive. This would achieve 'representational measurement' (Dawes, 1972). In other words, this offered the possibility of analysis which could provide a two-way flow of information from the property being measured to the scale measuring that property. Such measurement models would be structure-seeking - but not structure imposing - and would offer inferential prediction rather than descriptive reporting.

Dawes (1972) also described 'index measurement' models, where there is a one-way flow of information from the property being measured to an index purporting to measure that property. However, there can be no converse flow from the score on the index back to the prediction of behaviour, meaning there can only be description but not inference. Dawes (1972) uses the example of a single rating scale, though summated rating scales offers just as little and may in fact be more misleading. Index measurements such as these occur 'whenever there is a specifiable rule that leads to assignment of measurement scale values' (Dawes, 1972, p. 15). This assignment rule would be essentially atheoretical, and could even exist as a checklist of heuristic value but no more.

For example, a wide range of behaviours may be classified as arson; that is to say, an index measurement would be 'arson'. However, this index firesetting behaviour could in fact be described in legal terms ranging from vandalism to murder or attempted murder. Furthermore, it has been suggested that arsonists may be differentiated meaningfully on the actions associated with their crimes (Canter and Fritzon, 1998). Specifically, this refers to the source of action and the locus of actualisation, such that actualise feelings onto the world is expressive, while actualising external influence onto the world is adaptive (Shye, 1985b). Consequently, types of arson associated with these two modes might be burning of a building of significance to an individual as opposed to groups of youths burning bins on a spree. To return the meaning and intention of the firesetting would not be possible if both were classified simply as arson, as in an index measurement model.

Stevens and the Scales of Measurement

These issues in the measurement and modelling process have however been divorced from research in psychology. Cliff (1993) explained that in psychology the separation of methodological issues away from the substantive was an unintended result of Stevens' 'scales of measurement' (Stevens, 1946). These had been proposed by Stevens as a way of systematising measurement in psychology, which according to Michell (1990) had been in a 'dismal situation' in the first half of this century. This situation had arisen from the conclusion by a committee from the British Association for the Advancement of Science that psychological measurement could not achieve the mathematical basis, which would elevate it to the status of a 'science'.

This was a shock to psychology since each of the areas of the emerging psychology were bound together as the new 'science of behaviour' by what could be termed a strong positivistic 'quantophilia' - a love of measurement (Estes, 1993; Lingoes (1977a). Psychology had firmly heeded Galton's call in 1879 that it be grounded in quantitative measurement to allow it to 'assume the status and dignity of a science'. Similarly, the founder of the factor analysis, Spearman, would suggest shortly after that the experimental method could be made complete in the scientific sense only by the 'further and crucial method [of] measurement' (cited by Michell, 1990). Psychologists had therefore received Stevens' scales of measurement so eagerly since the history of psychology had been characterised by a desire for and gradual move towards quantification.

Stevens argued 'the real issue is the meaning of measurement' (Stevens, 1946, p. 677), and that what caused the issue to be confused was the fact that there were different types of measurement, each with a different meaning. Stevens' new operationalist explanation proposed that the meaning of measurement was to be found in how it was used in the different circumstances (Michell, 1990). If measurement was operationally defined as 'the assignment of numerals to objects or events according to rules' (Stevens, 1946, p. 677, paraphrased from Campbell, 1940) then there were four different ways in which it was done, and each was a scale of measurement.

Stevens (1946, 1951) described how at a basic scale, objects may be assigned to categories if and only if they conform to the membership requirements of that category. Each object put into the same category is similar in some respect. This forms the most basic scale of measurement, the nominal scale. With the ordinal scale, the rank ordering of objects along some dimension may be measured such that relative position on that dimension is meaningful. When the intervals on such a dimension are known and are equal then the measurement is an interval scale. If the zero point of an interval scale is anchored to a meaningful point, then the absolute ratios of objects on the dimension make a ratio scale. Each of the scales was a measurement model which assumed certain mathematical relations in the numerical scale would hold in the objects empirically examined.

As scales progress from nominal to ratio, the scale criteria become progressively stricter by adding more conditions onto those of the previous scale. This lead to a hierarchy within the four scales, with an implicit goal being to work towards a ratio scale, which is 'most commonly encountered in physics' (Stevens, 1946, p. 679). Not suprisingly perhaps, Stevens' own 'sone scale' of auditory perception was at the ratio level. Indeed, scale snobbishness can be found elsewhere. For example it was also evident in the meta-analysis of Poland, who remarked on the 'trend in the use of more powerful data' (Poland, 1983, p. 283) as 'criminal justice studies' became added to the 'empirical sciences' on account of its use of quantitative data and statistical inference.

For many the all-encompassing scales of measurement, answered once and for all the key issues concerning measurement in psychology and would continue to shape psychological methodology well into the future. More importantly, it allowed the pursuit of the 'methodological monist' notion (Bryman, 1988) that the same methodology was applicable to both natural and social sciences, albeit at some restricted levels. Consequently, Stevens' definition, description and utilisation of 'scales of measurement' were an instant success, and continue to be used widely in psychology as synonymous with measurement theory (e.g. Shavelson, 1988, p. 18; Jacoby, 1991, p. 5; Minium, King and Bear, 1993, p. 19). After a domain had been investigated, and data gathered on it, all that was required was to consult which scale it belonged to. Other issues in method and measurement were relegated to an afterthought to decide on which statistic to use.

However, the use of these scales also created a widespread use of index measurement models at the expense of representational measurement models. This was noted early by Coombs, Raiffa and Thrall (1954), who stated that:

> A measurement scale, such as an ordinal, interval, or ratio scale, is a model and needs only to be internally consistent. As soon as behaviour or data are "measured" by being mapped into one of these scales, then the model becomes a theory and may be right or wrong. (Coombs, Raiffa and Thrall, 1954, p. 137)

Therefore a scale of measurement can only be accepted or rejected on logical grounds, in terms of internal criteria. When these are applied to real-world data, then it must be accepted or rejected on the basis of external substantive criteria. This is the 'value-added' part of theory construction and testing over pure measurement, which was missing in Stevens. The difficulties associated with inappropriate and inadequate measurement models can often be easily ignored using scales of measurement. By contrast, in Empirical Studies 7.1, 9.4 and 13.1 the importance of appropriate and empirically testable models is all important.

The Implications of Scales of Measurement

The fact that a set of empirical observations might conform to one of the four scales of measurement model in itself was not problematic. But if researchers took the importance of the scale levels as merely a proscription on which statistics were allowed to be used with which scales, then the structural properties of the scales functioned only as external constraints. For example, ordinal scales, containing only information on the relative position of scores, should only use the median and semi-interquartile range as measures of central tendency and dispersion. Even these proscriptions are abused, though.

Consider the 'ubiquitous rating scale' (Dawes, 1972), where the attitude of an individual to a statement is measured by asking if they agree, feel neutral or disagree with the statement. It cannot be ratio since there is no meaningful zero point. The zero refers to either no opinion, or equality of positive or negative attitudes for the statement. It cannot be interval, since that would require the assumption that the difference between each and every point in the scale was equal. For this to hold, the difference between 'agree strongly' and 'agree' would have to be the same as for example 'agree' and 'feel neutral' on a five point scale, which would require empirical proof. But there is more information about magnitude than is found in the nominal scale, since each point on the rating scale is related to intensity of the same attitude.

Strictly, then, the rating scale is ordinal. This would imply that the full range of variance-based statistical procedures - requiring an interval scale - should be denied to the ubiquitous rating scale. In fact, the use of variance-based statistical procedures should have been low since Stevens himself stated that 'most of the scales used widely and effectively by psychologists are ordinal scales' (Stevens, 1946, p. 679). A further implication of the rating scale being stiictly interval is that it then becomes desirable and essential for the scale of attitude to assume the average person has no attitude at all, making the responses symmetric (Guttman, 1976).

None of this has stopped researchers using means and variance-based statistics to analyse scales such as rating scales, however. Different rationales have been found to justify such use. These range from fudging - where good quality rating scales are considered somehow between ordinal and interval scale but worthy of interval scale analysis - to expediency, where both ordinal and interval analyses should be used and no questions should be asked provided the conclusions are the same (Minium *et al.*, 1993, p. 77). Therefore Stevens' popular solution to the crisis in quantitative psychology led to researchers neglecting what exactly was being assumed by a particular style of analysis and scale of measurement. Thus in an answer to a 'Tricky Statistics' questions, Booth (1995, p. 197) stated that parametric analysis of ordinal rating scales was acceptable. This was immediately followed by the apologetics of an editorial comment stating that 'it does not follow [from this] that *all* ordinal data can be analysed and interpreted meaningfully using parametric tests' (BPS, 1995, p. 197, emphasis in original), adding trenchantly that researchers should use substantive criteria rather than follow maxims learnt by rote from undergraduate psychology!

But by focusing on the mathematical axioms that made up the process of measurement, there was and still is a diminished emphasis on what exactly was the purpose of measurement. Data analysis was performed within the confines of scales of measurement. The sole practical use for the scales of measurement made by researchers was to decide on 'appropriate' statistics given a particular measurement model. Researchers need only pay lip-service to issues of data and of measurement, if at all. The only consideration of these issues would be in terms of the statistical restrictions they impose, rather than substantive questions of replicability, validity or utility.

Alternative Views of Measurement: Coombs and Carroll and Arabie

A far more useful contribution to the representation of the structure was given by C. H. Coombs, which in turn has an impact on conceptualising criminal actions. Coombs (1964) not only offered a wider understanding of the research process introduced in chapter 1 as the 'Coombsian Research Model', but he also explained how data were not a static concept but differentiable into a classification of distinct types in a theory of data. This theory of data incorporated and expanded on much of Stevens' scales of measurement, and will be used to understand how the same information can be made into different representations with which to test structure of criminal actions.

Coombs' theory was derived from the axiom that '[d]ata may be viewed as relations between points in a space' (Coombs, 1964, p. 1). This radical, geometric interpretation of data requires a consideration of the representation of data in that space. Previously hidden concerns were revealed, such as the measurement and meaning of proximity between points in space (see chapter 8) and the kind of possible orders in that space (see chapters 14 and 15). The emphasis consequently is shifted towards understanding the nature of the data. Data theory is said to 'examine how real world observations are transformed into something to be analyzed' (Jacoby, 1991 p. 4), including how structure can be represented.

The geometric representation of data puts into space the behaviour of individuals responding to stimuli, the stimuli themselves, or both. The points are related in some way. From this a classification of types of data may be built. Coombs classification of data was derived from the answers to three questions:

1. whether the relation is between points or pairs of points,
2. whether the space contains a single set of elements (e.g. only stimuli), or two sets of elements, (e.g. stimuli and subjects) and
3. whether the relation is either proximity or order (dominance).

The first dichotomy was diminished in importance later by Coombs, Dawes and Tversky (1970), so that the second and third categories were cross-tabulated to distinguish between four types of data. Each of these types was named and given a 'quadrant number', illustrated in Figure 2.1.

	Dominance matrix	**Proximity matrix**
Two sets: the off-diagonal submatrix	Quadrant II Single stimulus	Quadrant I Preferential choice
One set: the intact matrix	Quadrant III Stimulus comparison	Quadrant IV Similarities

Figure 2.1. Quadrant of data types from Coombs, Dawes and Tversky (1970)

The classification of data as a type was shown by Coombs (1964) to have an impact on how the data can be collected, represented and analysed by the researcher. Coombs, Dawes and Tversky (1970) provided examples of these distinctions, such that dominance relations of order may be that one tone is louder than another while proximities are where two tones may be judged similar or dissimilar. The key difference between the two is that dominance is judged on an explicit dimension, namely volume in the dominance example. In the proximities example, the dimension or dimensions on which the individual judges are left implicit. In other words, the intention of the individual's judgement is left open, making the judgement more qualitative than in the dominance relation.

The second aspect of this typology, the number of sets for the data, was expanded on later and given a fuller terminology by Carroll and Arabie (1980). Jacoby (1991) described this as an 'alternative' theory of data to that of Coombs (1964), but the fundamental emphasis in Carroll and Arabie (1980) was to outline the nature of the data input matrix, rather than act as a challenge to Coombs (1964) - as implied by Jacoby. To this end, the two theories should be seen more as complementary than competing.

According to Carroll and Arabie (1980), the standard form of data matrix (used as a default in most statistical packages such as SPSS, SYSTAT and MINITAB) is rectangular in shape. In such a matrix, columns are variables ('objects', 'stimuli', 'attributes', etc.) and rows are cases ('subjects', 'responses', 'respondents', etc.). For criminal actions data, this may be the different behaviours (variables) during a sample of offences (cases). The matrix would be described as *two-way*, meaning that it have two distinct dimensions, i.e. rows and columns. Since each way is from a different source of data in this matrix - namely from stimuli and responses - it is described as having *two modes*. For Coombs (1964) this would correspond to two sets. This means having two distinct sources of information, distinct entities which in this case are population and variable information. Entries into the cells of this matrix are the recorded observations, which may be numerical or non-numerical, depending on the nature of the research. It is possible for matrices to be three-way three-mode or even more (Kruskal and Wish, 1978), where the third mode and way could be for example an index of time. Therefore replications of an experiment constitute an additional dimension. The only limitation in terms of ways and modes is that the number of modes cannot exceed the number of ways.

An alternative shape for a data matrix is *triangular*, where the same objects are in both rows and columns. The cell entries are the association or correlation between the objects, giving a measure of the similarity between objects. This style of matrix is therefore described as having *one mode*, since information from only one source is included, but still having two ways since there are two dimensions. Analysis techniques which work with such one-mode data often start from a two-mode rectangular matrix and then convert that information into a one-mode triangular matrix by correlating or associating variables. In other words, it creates either a n by n matrix or an N by N matrix, according to whether variables or cases are associated. The one-mode matrix is then a correlation or association matrix, which is then analysed by the technique. This additional level to the analysis is highly important, involving issues such as the nature of the data, the choice of appropriate correlation coefficient or similarity measure and the justification of any statistical assumptions.

The theory of data was 'offered as an analysis of the foundations of psychological measurement' (Coombs, 1964, p. vii). Like Stevens, Coombs placed emphasis on the nature of the data during the research endeavour. However unlike Stevens, Coombs developed a classification based on what constituted the type of data rather than what constituted the type of measurement. Interpretations of the representation of data are guided by an *a priori* substantive consideration of what the data refer to in the real world. Interpretation is not dominated by *a posteriori* and extrinsic statistical restrictions from the code of permissible operations. The principles of data collection are tied into the structural representations of those data. This was a move towards prioritising substance over statistics in research.

Coombs noted that a 'measurement or scaling model is actually a theory about behaviour' and went on to state the dictum 'we buy information with assumptions' (Coombs, 1964, p. 5). So by using a scaling model, it must be assumed that the very observations are in fact scaleable. As more assumptions are made, more information is gained. Necessarily this means that better information is traded-off with more restrictive, elaborate analyses.

There may come a point at which the cost of using heavily restrictive analysis makes the information too expensive. In variance-based statistics, for example, it is possible to find the exact source of difference in the multiple comparison of means. Similarly in multiple ANOVA, the use of many *post hoc* tests such as the Scheffé or Tukey gives extra information, but at the cost that each test may be a false positive. Similarly, many correlations carried out among a set of variables may result in false positives. In such cases, it may be necessary for the analysis to be 'relaxed' and re-evaluated in the light of the quality of the data. For this thesis, another significant lesson from the theory of data is the 'plea for the use of weak measurement models' on account of his 'uneasiness about the assumptions we adopt in whatever model we use' (Coombs, 1964, p. 284). The nature of criminal actions information introduced in chapter 1 would seem to indicate already the need for weaker measurement models.

However, Coombs' theory of data was largely restricted to identification of the formal structural properties of data rather than having an impact on the substantive and theoretical applications for the data. The usefulness of data and the measurement models are derived from the more qualitative nature of the data in terms of reliability and noise. Facet Theory is shown in the next chapter to be concerned both with formal structural properties of data and the measurement models that may be realistically found in applied settings. It offers a range of methodologies with weaker representations that do not lose any power to test structure in criminal actions.

Summary of Chapter 2

Representation was introduced as being the modelling of observations made on a structure. Representations of quantitative measurement models of criminal actions were preferred to qualitative classification models. 'Representational models' can act as structure testing devices. Stevens' scales of measurement were inadequate as representational models, and diverted attention away from normative examination of what should happen to descriptive *post hoc* examination of what does happen. Coombs' data theory is helpful for understanding representation, but a more formal correspondence for testing structure is required.

Chapter 3
the faceted approach to structure and representation

Towards a Geometry of Behaviour and the Inference to Action

It follows from Coombs (1964) ideas that if the data created from observations of structure are considered as points in space then the most natural and intrinsic representation of data should also be geometric. The configuration of the points in the space would be a reproduction of the constituent attributes which created the structure of the real-world concept. To achieve this mapping between structure and representation, not only the empirical observations made on the structure be valid and reliable but the representation must also be adequate to model the structure. In the ideal case, these conditions are met and the structural hypotheses can be modelled mathematically and represented as distances between the data points in space. Therefore the mathematical models of behaviour give insight into the meaning of the domain and allow substantive inference to be drawn from the research.

The need for this geometry of behaviour accommodates the desire for a multivariate analysis of human actions. While the tradition of experimental psychology has been proud of its use of independent variables and control groups, no such luxury can always be afforded to applied measurement of more complex human actions. Cattell (1988) stated that some early researchers recognised this and consequently tailored their own methodologies, such as Spearman in the study of human ability. The applicability of Spearman's factor analytical methodology to applied research in other areas was realised with the advent of computers to perform the large number of calculations required. While moving psychology out of the laboratory, it still also allowed objectivity and replicability to regulate its results. But by the same token, such popularity led to weakening of the theoretical precepts, since the procedures was so accessible but required little or no esoteric knowledge of the theory. This of course was coupled with the sheltering of measurement issues away from the theory-building process and behind Stevens' scales, as shown in chapter 2.

While it was Spearman (1904) who performed the first large quantitative analyses of intact Coombsian Quadrant IV 'similarities' matrices in geometric representations, Torgerson (1952) later suggested that the 'closeness' in a proximity matrix should be translated into distances in explicitly graphical solutions. This ran contrary to the actual plotting of vectors in space which is usually pushed behind the scenes in most factor analyses, a critique returned to in chapter 10.

While the proposal of a geometric representation of data seemed simple, there was no simple solution to it. There was no 'perfect' analytical solution to the problem of translating associations in a correlation matrix into distances in a geometric space. What was needed was a best-fitting solution, and for this iterative computation was necessary. The period between Torgerson's publication of his ideas and Coombs' data theory saw rapid development in the computerisation of iterative programs to achieve this. For Quadrant IV, Guttman (1965) and Kruskal (1964a) each developed their own similar but distinct methods, as is described in chapter 5. The key idea promulgated by both, though, was that there could be a reduction of the dimensionality of a concept into its less complex constituents, which could then be represented geometrically. The process therefore posited is a dimensional reduction to parsimony - the smallest space with the lowest number of dimensions.

Guttman was also involved in the development of techniques to analyse Quadrant II data with the Guttman scale (originally Cornell scale; Guttman, 1944). This is important for chapter 14, where a different structure is tested from the same data by conceiving it as different Coombsian data types. With Guttman scaling it is possible to classify individuals according to intensity on a given dimension, measured in discrete chunks by several variables, which in turn are ordered themselves. Guttman scaling is placed in Quadrant II because it is two sets of data connected by dominance relations, namely the issue is whether the individual dominates the stimulus or the stimulus dominates the individual. In the former, the individual passes the test or endorses the item but in the latter the individual fails the test or cannot endorse the item (from Coombs, 1964, p. 227). The ability to explore such Quadrant II structures to higher dimensionality, is possible through Multiple Scalogram Analysis and Partial Order Scalogram Analysis, which take further Coombs' assertion that '*behavior is* *intrinsically partially ordered*' (Coombs, 1964, p. 285, emphasis in original). These procedures are used in chapter 15.

However, Coombs *et al.* (1970) suggested that the creation of programs to analyse data using geometric representations formed only part of research into multi-dimensional scaling (MDS). Equally important, it was suggested, was research into MDS models with special consideration to more broad measurement issues. Guttman in particular was concerned with both geometric representations of data and formal explorations of structure, and his development of Facet Theory linked hypotheses of structure to the methods of representation outlined above.

Structural Hypotheses and The Faceted Approach

Canter (1983) suggested that Guttman proposed Facet Theory as a method of creating structural hypotheses due to his concern with the 'selection of items for test construction and with the weaknesses in factor-analytic procedures as well as with the lack of clarity of existing approaches to the definition of research problems' (Canter, 1983, p. 37). The faceted stance on structure and representation was stated forcefully by Guttman in his presidential address to the Psychometric Society (Guttman, 1971), in which Guttman provided a succinct summary of his thoughts on issues of measurement.

In the address, Guttman accepted that measurement should be understood in terms of the ways in which it is used - an operationalist slant similar to Stevens (1946). However, there was emphasis on the view that measurement should not be purely an adjunct of statistical theory. Measurement was concerned with 'the construction of structural hypotheses ... [and] the structure of regressions among variables' (Guttman, 1971, p. 332) and should be sensitive to the nature of the domain under scientific examination.

The link between structure and representation of that structure was brought to the fore by Guttman. For example, Guttman's definition of a 'scale' as a one-dimensional structure (Guttman, 1971, p. 343) can be readily contrasted with Stevens', which stated that a scale is a 'rule for the assignment of numerals (numbers) to aspects of objects or events' (Stevens, 1951, p. 23). Guttman's structural definition of the scale implied testability, measurement and representation. Codes of permission and principles of measurement were not part of this view. Though used in similar ways, these views on 'scales of measurement' differed on definitional terms.

Guttman elaborated on the awkwardness inherent in discussion of interval and ratio scales such as the difficulty of mathematically comparing scores on an interval scale. Two times 60 is 120, but two times an IQ of 60 is not 120 since these scores are part of an interval IQ scale. However Stevens' definition is merely an *a posteriori* descriptive possibility, where the scale of measurement is fitted *post hoc*. In fact, Guttman later added to this point that '*Nominal, interval and ratio scales are not scales*' (Guttman, 1977, p. 105, emphasis in original).

The difference between interval and ratio levels of measurement is simply the understanding about the zero point; ratio level uses units compared to a zero, interval uses units compared to each other . Furthermore, Guttman went on, coefficients that are used to measure ratio level scales were intended in fact for interval level, under the rules of measurement. Thus the Pearson product moment coefficient uses the sample mean to calculate the sum of squares rather than zero, as it should do if the measure were genuinely ratio. And the physical sciences could not strictly speaking achieve ratio scales of measurement anyway, Guttman went on, since the distance measured between two points is relative to other points in space, and time elapsed is never made relative to the absolute zero of time.

Stevens had tried to argue that the meaning of measurement in his scales was crucial (Stevens, 1946, p.677). However, Guttman suggested that Stevens' scales were still blighted by a 'mere matter of communication ... not a deep philosophical problem concerning a principle of measurement' (Guttman, 1971, p. 340) and rejected the '*a priori* mathematical and statistical considerations and prescriptions - especially codes of permission' (Guttman, 1971, p. 346). This should be replaced by substantively-guided thinking in measurement and attention to the universe of observations. Guttman finally concluded the exposition of measurement as structural theory by welcoming the growing acceptance of 'qualitative observations with the same respect as numerical observations' (Guttman, 1971, p. 346), as exemplified by the use of the Guttman scale illustrated in chapter 11 and empirically tested in chapter 14.

Therefore, while embracing a measurement theory based in science - the quantophilia described earlier - Guttman stressed the need for substantive criteria to guide the process. In this way, measurement means making empirical observations to test for the presence or absence of hypothesised structures. Those hypothesised structures have been proposed on the basis of existing work and are readily amenable to analysis. The statistics and mathematics with which to model the empirical observations must not limit the structure-seeking process, or at least to minimise it. Conclusions drawn from an empirically obtained structure or lack of it should be phrased so as to lead directly to replication or variation, giving the essence of the faceted approach.

The faceted approach to research begins at phase 0 of the expanded CRM.

Item Selection and Facet Theory

The universe of potential stimuli for inclusion in a test is not a random or unstructured domain. Yet Guttman asserted that the decision over which stimuli to select at phase 0 of the expanded Coombsian CRM, introduced in chapter 1, was rarely done with some *explicit* substantive rationale in mind. By contrast, most research is conducted using an *implicit* culling rationale for phase 0. The importance of decisions made at phases 0 and 1 is that they constitute the construction of test items with which to measure respondents or subjects. An 'item' contains the valid response to a valid stimulus, which is then recorded as an observation. To use implicit culling rules for items on an assumption of mutual understanding is unscientific, and is avoided by Facet Theory.

Aside from the faceted approach, Borg and Shye (1995) identified two different approaches to the creation of items, namely the exploratory and the item analysis approaches. The exploratory approach takes batteries of items and then usually factor analyses then to define the content of the groups items found. Meaning is attached to the classification *a posteriori*, and consequently the 'general problem of the exploratory approach is that it mixes substantive and data analytic procedures' (Borg and Shye, 1995, p. 81), by allowing the factor analysis to create structure.

The item analysis approach is also essentially exploratory, operating on an explicit statistical rationale which has an implicit substantive basis. The explicit statistical rationale is that only items with a non-negative correlation with other items are included in the analysis. Thus individual items found to be statistically independent of the bulk are deleted as not being partial measures of the construct, leaving the remaining items to be summed into a scale which is the implicit substantive rationale. If several variables correlate systematically but negatively with the mass, then these are deleted as not being partial measures of the concept. But it is possible that the mass of variables is not a better measure, even if numerically it was greater. In such an instance, the handful of variables to be deleted could in fact represent the true nature of the content universe, or could represent a distinct subset of it at a higher dimensionality.

Guttman, was highly critical of constructing scales by 'item analysis', accusing its practitioners of being unscientific. Guttman pointed out that scales found using item analysis assume that a specified dimensionality will exist, and finds combinations of variables that prove this by 'gerrymandering the data' (Loevinger, cited by Coombs, 1964, p. 231). It is structure-imposing, as opposed to structure-seeking. Guttman claimed that the 'dimensionality of data is an empirical phenomenon, and not to be determined by *fiat*' (Guttman, 1971, p. 343). This was a reference to Cicourel's measurement by *fiat*, where 'measures are simply asserted, and little, if anything, is done to demonstrate a correspondence between measures and their putative concepts' (Bryman, 1988, p. 29). Guttman stated that research into structures must use measurement models which are falsifiable, as recognised widely by authors cited in chapter 2 such as Coombs (1964), Roskam (1981) and Jacoby (1991). Such an instance of an unfalsifiable methodology is given in chapter 12 and contrasted with the faceted approach in chapter 13.

Guttman stated that 'Scalability is not to be desired or constructed ... [and] is generally not a null hypothesis' (Guttman, 1977, pp. 100-101). Instead, it was an empirical supposition to be tested, and for example if unidimensionality was not empirically supported then an alternative hypothesis of multidimensionality could be supported instead. Deleting items that did not fit a unidimensionality hypothesis was 'like throwing away evidence that the world is round' (Guttman, 1977, p. 100).

It was suggested that 'one must conceptualise - in substantive terms - what is being studied before one proceeds to design tests, gather data, and go through elaborate statistical analyses', as Spearman and Thurstone had originally done (Guttman, 1967, p. 438). However, Spearman and Thurstone did tend to 'rely on the statistical analysis to formalize the theory for them' (Guttman, 1967, p. 438). The later proponents and users of Thurstonian factor analysis also tended to do the same. Guttman claimed it was the 'purifying' of items in search of internal consistency and scale construction - rather than empirical testing of possible scales - that led him to the notion of defining a 'universe of content' (Guttman, 1960, p. 4). It also led him to consider separately the problems of internal consistency and external prediction.

Mapping Sentences, Facets and Item Construction

The ability to make clear testable structural hypotheses in the faceted approach can be achieved through the use of a 'Mapping Sentence' (MS). A MS is not a hypothesis in itself, but gives the framework from which to generate hypotheses of structure (Donald, 1995). The MS serves as a semi-technical phrase that sets out the concept under investigation by outlining 1) the people being investigated, 2) the different aspects of the concept and 3) the responses for the measurement of the aspects. These three components of the MS are the 'facets', and are termed background, domain and range facets respectively. In other words, the MS is a definitional statement of how a given population may be mapped onto a set of acceptable answers (range) when examined on certain aspects of a concept (domain).

The faceted approach uses these definitional systems to guide exploration of the logical possibilities that may or may not be observed empirically. Useful definitional systems then inform theory. The faceted approach is consequently oriented towards theory construction and 'constructing the original observations which are to be subjected to data analysis' (Guttman and Guttman, 1976, p. 470), in much the same way that Fisher intended his ideas on factorial design to structure controlled experiments. The usage of the term 'facet' rather than 'factor' in Facet Theory was to avoid confusion with the factors in factor analysis (Guttman, 1954a).

To illustrate an MS, consider the following example. People who have committed arson (population) are males or females (background facet: gender) of different ages (background facet: age) who may or may not (range facet: existence) have left traces of instrumental or demonstrative actions (domain facet: mode) focused against person or property (domain facet: target). These facets contain 'elements', a continuation of set theory notation. Thus the range facet has elements 'present' and 'absent'; the domain facet of mode has elements 'persons' and 'property'.

The MS for a population of arsonists would is given in Figure 3.1.

$$\text{Person (x) who is a} \overset{\text{AGE}}{\begin{bmatrix}\text{young}\\ \text{old}\end{bmatrix}} \overset{\text{GENDER}}{\begin{bmatrix}\text{male}\\ \text{female}\end{bmatrix}} \text{ sets fires against } \overset{\text{TARGET}}{\begin{bmatrix}\text{person}\\ \text{property}\end{bmatrix}}$$

$$\text{leaving a} \overset{\text{MODE}}{\begin{bmatrix}\text{demonstrative}\\ \text{instrumental}\end{bmatrix}} \text{ action at the crime scene} \longrightarrow \overset{\text{RANGE}}{\begin{bmatrix}\text{yes}\\ \text{no}\end{bmatrix}}$$

where (x) is from the population of convicted arsonists.

Figure 3.1. Possible Mapping Sentence for population of arsonists

In this example, the population is self-selected in that the members of the set have committed and been convicted for arson, and care must be taken so that a sample for empirical purposes represented this admittedly skewed population. A well-designed facet should have been adequately theorised by the researcher, and have elements that are mutually exclusive and are collectively exhausted by the facet. The domain facets *in toto* should collectively exhaust the domain of interest (Donald, 1995, p. 123). Such a domain facet can be termed *fully formed.*

The domain facet and its elements act as 'a classification of stimuli' (Shye and Elizur, 1994, p. 5) and form the body of the MS. A research study will typically contain more than one area of focus, even though it will be examining a single domain. The ways of looking at the areas are the domain facets. The construction of items for a questionnaire to examine the concept is based around the Cartesian product of each domain facet. Therefore each element in a domain facet is combined with each other element in all the other facets. This property gives rise to the technical definition of a facet as 'A set playing the role of a component set of a Cartesian set' (Shye, 1978c, p. 412), although this particular definition is substantively barren.

Each member of the Cartesian set is termed a structuple. From these structuples, questionnaire items can be phrased so as to examine the response to the meaning of that particular combination of elements. For example, if a MS contained fully formed domain facets A, with three elements and B, with four elements. The Cartesian product of the facets is AB and the contains the set of structuples a_1b_1, a_2b_1, ..., a_3b_4 namely 3 x 4 = 12 structuples. Structuples are useful for designing items for questionnaires using each structuple ideally at least once, though Shye (1978a) this is not absolutely necessary where impractical. This allows a parsimonious number of items to be used, unlike the more vague recommendations of Ghiselli, *et al.* (1980) for example 'as we increase the number of measurements that enter into the determination of a score, there should be an increase in the reliability of the measurement' (Ghiselli, *et al.*, 1980, p. 231).

Broadly, the faceted approach satisfies both phases 0 and 1 of the expanded CRM in Figure 1.2. This was alluded to in Shye and Elizur (1994, p. 3), where it was pointed out that the 'continuous space imagery' of the faceted approach means that it samples from a continuous universe of stimuli - phase 0 of the expanded CRM. The definition of the characteristics of the item universe is formalised with domain and range facets, which serve as structural hypotheses for empirical examination. In this sense, a fully faceted design is a considered statement of what items to test which respondents and how to measure a response. The formalised expression of this statement leads naturally to fully testable hypotheses of structure, and guides the researcher through the other phases of the expanded CRM.

Testing Structural Hypotheses in Geometric Representation by Contiguity

To investigate the empirical structure of the sample given this definitional system, it is necessary to calculate the similarity between each element of each facet. This is done by deriving data from these observations in a correlation matrix to create Quadrant IV similarities data (Coombs, 1964). The matrix of correlations or associations is an hypothesised summary of the structure of the content universe. Therefore if this matrix were to be translated accurately into a geometric space, the representation would be a reproduction of the structure. Furthermore, if the universe were structured as hypothesised from the MS, the geometric representation would reproduce this. The space could be partitioned into regions which reflected the structural hypotheses in the MS, i.e. the facets. This process is termed regional interpretation and the partitions are regional hypotheses.

The regional hypotheses in the representation will be ordered according to contiguity. Contiguity was defined by Foa (1958) as being the property of structuples having elements or 'structs' in common. Many structs in common implies a higher conceptual closeness for those structuples. From this was derived the Contiguity Principle (CP), which 'suggests the following hypothesis: *The larger the number of contiguous facets between two variables, the higher their intercorrelations*' (Foa, 1958, p. 233, emphasis in original). Foa stated that 'conceptual contiguity is a necessary condition for statistical dependence' (Foa, 1958, p. 230), and later that 'variables which are more similar in their facet structuple will also be more related empirically' (Foa, 1965, p. 264). Such a contiguity relation was tested by Foa (1958) with workplace attitudes to bosses and workers.

However, the original and seemingly simple idea of the CP has been the subject of controversy, with attitudes to it ranging from reverence (Brown, 1985) to dismissal (Borg and Shye, 1995). Brown (1985) stated with force that the notion derived from the CP - that conceptual similarity led to empirical similarity - was a 'an absolutely critical assumption' of Facet Theory (Brown, 1985, p. 20). Yet the CP was criticised by Borg and Lingoes (1987), ignored by Shye and Elizur (1994) and roundly rejected by Borg and Shye (1995).

Borg and Shye (1995) suggested that the CP was a primitive principle of correspondence between design and data, and that it could be superseded by more specific principles, namely the Principle of Empirical Nontriviality, the Principle of Formal Control of Variance and the Principle of Discriminability. Empirical Nontriviality means that definitions are linked to some regularity in empirical data. Formal Control of Variance refers to the principle that differences in sources of variance must be related to the design of facets. Discriminability asks whether distinct regions of items or respondents or both can be identified.

Yet it is possible that the term 'principle' may actually be a misnomer, even though Foa (1965) later only referred to Contiguity in terms of the 'Contiguity Principle'. We know that the most basic meaning of the term 'principle' is 'A general, basic maxim; a fundamental truth.' (Reber, 1985, p. 574). This was demonstrated by the quotation above, which stated that the CP 'suggests the [Contiguity] *hypothesis*' (Foa, 1958, p. 233, emphasis added). For this reason, when the Contiguity 'Principle' is applied as an empirical test of fit between definition and data then it should be understood instead as 'hypothesis'. The practical implications of the CP as an hypothesis is simply that greater similarity in faceted definition implies greater similarity in empirical data, *provided that the definition is valid.* This hypothesis is merely a suggestion of the pattern of similarity between definition and data. The precise nature of the relation could be any of the four different types of monotonicity (Guttman, 1986). Alternatively, perfectly linear or even curvilinear functions could equally used to describe the relation, with the increasing strictness indicating stronger correspondence given the nature of the data.

Recasting contiguity as an empirical test means that it is not a descriptive index in the style of Stevens' levels of measurement, something that cannot be 'assumed' to hold for a study (*pace* Brown, 1985). Let us consider what would happen if the original hypothesis of Contiguity were shown not to be true for a set of empirical observations made on a particular design. The empirical representation of the facets would not be in distinct regions (the Principle of Discriminability). This would mean that the explained variance in the correlation matrix was not due to the faceted definition (the Principle of Formal Control of Variance). The absence of regularity in the data due to the design (the Principle of Empirical Nontriviality) would manifest in the empirical plot and in the correlation matrix. In other words, the principles would be worthless, because contiguity was not proved.

The idea that Contiguity is an hypothesis is also implicitly recognised in the whole structural hypothesis testing procedure in Facet Theory itself. For there to exist distinguishable regions in empirical data, this implies that the design must also be conceptually distinct elements in well-formed facets. If there is no correspondence between design and data, then there are no contiguous regions and the facets have not conceptually captured the domain of interest adequately. This process is clearly a falsifiable process of hypothesis testing involving Contiguity and was clearly intended by Guttman's widely cited definition of theory (e.g. in Brown, 1985; Shye and Elizur, 1995; Donald, 1995). The definition stated:

> A theory is an hypothesis of a correspondence between a definitional system for a universe of observations and an aspect of the empirical structure of those observations, and includes a rationale for such an hypothesis. (Guttman, 1982a, p. 335; originally cited by Gratch, 1973)

The phrase 'hypothesis of correspondence' can readily be seen to refer to the notion of Contiguity. Furthermore, if the Contiguity mapping does not stand up to empirical scrutiny for the definitions and data, then those contiguous regions that are found in the representation suggest - though not prove - the existence of other structures. They can then act as springboards to modify future MSs.

Geometric Representations of Faceted Structure

So far, it has been stated that the empirical 'closeness' of constituents of a structure can be measured by measuring the similarity in its attributes. These similarities in attributes were ideally constructed from correlating or associating items using faceted Mapping Sentences. If this closeness is then translated into a geometric closeness in the form of a line, plane, cube, etc., then the structure can easily be observed in the representation, provided there exists Contiguity between structural hypothesis and regional hypotheses. Contiguity in the geometric space can be demonstrated if distinct regions of the space could be 'partitioned' according to the structural hypotheses. The regions are content sub-universes (Shye and Elizur, 1995), designating a different part of the content universe.

Each actual or hypothetical point within a region would be classified as being part of the same element of the facet. Any point within a geometric space will be classified differently for each hypothesised facet. As Lingoes put it, 'Each partitioning of the space imposed by the investigator implies a particular classification of subsets of points ... In constructing these partitions it is crucial that there be some *compelling* reason for doing so.' (Lingoes, 1977b, p. 115, emphasis in original).

All, some or none of the facets may partition as hypothesised. Where there is no Contiguity for some or all proposed facets, then possibly the facet, the reliability of the items purported to test it or the respondents reactions to the item must be investigated to understand why. All this knowledge feeds back into the cumulative body of knowledge and informs future replications, thereby overcoming the objection that '[d]ata analysis remains barren unless a correspondence is established with the definitional system of the observations' (Levy and Guttman, 1975, p. 370).

The regional interpretation of contiguities in the geometric space is the method of identifying structures, as schematically suggested in Figure 3.2.

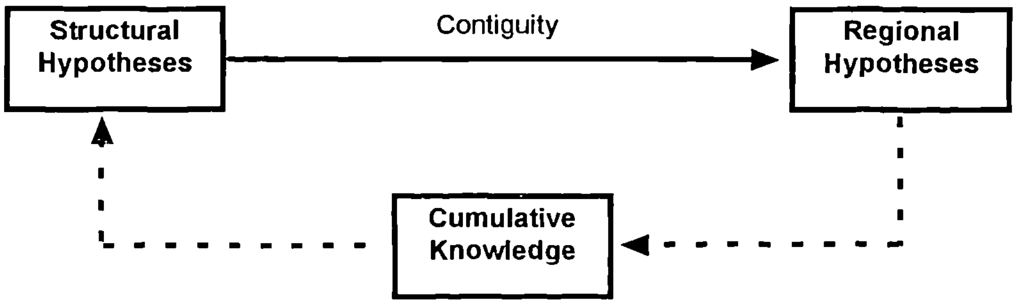


Figure 3.2. Process of structural hypothesis testing

As can be seen from this diagram, the regional hypotheses feed back into theory by adding to knowledge, which informs future research.

The boundaries of regional hypotheses in the plot denote the membership of points contained in them to the facet elements. The quality and strength of regional interpretation is increased where the number of items is high. This is important in applied research though may be less crucial in exploratory or theory-building research when the rejection of tentative hypotheses is not crucial.

The actual shape of the partitioned regions suggest different relations with the elements of a facet on account of the implied differences in similarities or correlations. Consider an MS containing fully formed domain facets A, with three elements and B, with four elements and Contiguity between data and definitions. In quantitative terms, facet A has order $a_1 > a_2 > a_3$ but for facet B there is no order i.e. $b_1, = b_2 = b_3 = b_4$. Since the spatial configuration of facets in a geometric space 'comes in part from consideration of order' (Levy, 1985, p. 74), there are two ways to represent quantitative order in facet A geometrically: as concentric rings or as parallel hyperplanes. In the former, the facet A has what is termed a radial or modulating function; in the latter, an axial function (e.g. Borg and Shye, 1995, p. 130; Shye and Elizur, 1994, p. 121-122). These are shown in Figure 3.3.

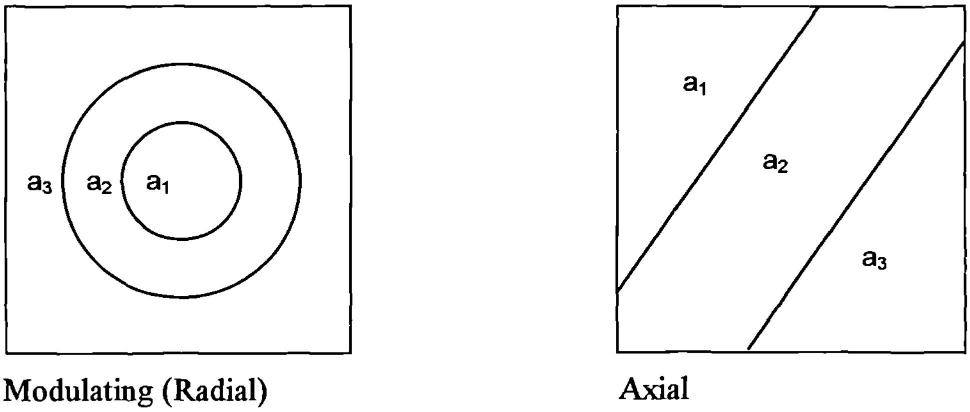


Figure 3.3. Examples of modulating (radial) and axial facets

The way to represent a qualitative differential order but quantitative equivalence - as in facet B - is as a polar or angular facet. In this domain (content) facet with no quantitative order, each element of the facet 'corresponds to a different direction in SSA space, emanating from a common origin' (Levy, 1985, p. 74). The meaning of the elements relative to each other in the SSA space is therefore qualitative or categorical, as shown in Figure 3.4.

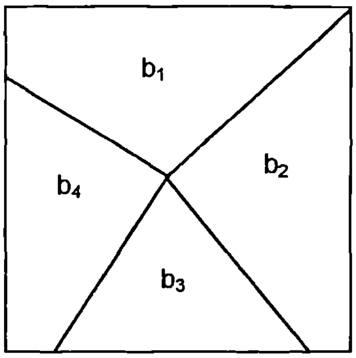


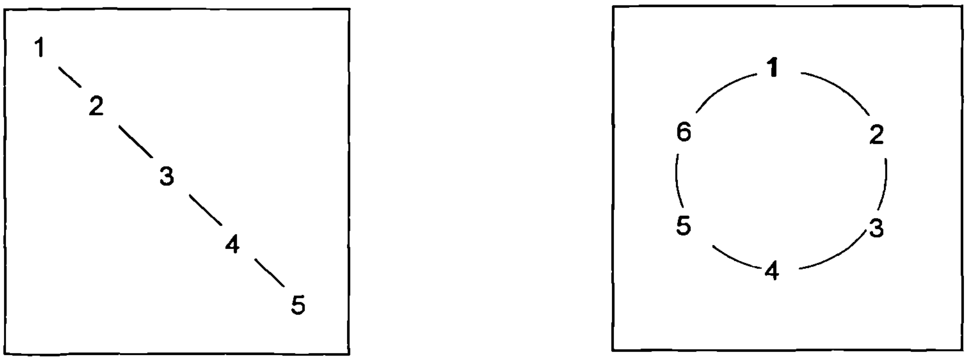
Figure 3.4. Example of polar (angular) facet

It is possible to combine many facets together to form new configurations where each partitioned facet describes a different 'facet' of the content universe. The most important of these is the radex, identified originally by Guttman (1954b) and consisting of facets playing angular and modulating roles, which is shown in chapter

13 to have a key place in understanding criminal action. The substantive implications of partitioning the plots in accordance with structural hypotheses has already been discussed.

Spatial Configurations of Data Points and Regional Interpretation

Regional interpretations are a more general form of understanding spatial configurations of points. More specific structure can be tested for existence in MDS representations. Two notable configurations are the simplex and the circumplex, given in Figure 3.5.



Simplex Circumplex

Figure 3.5. Example of simplex and circumplex configurations

In the simplex, the items are aligned along a straight line in two dimensions, or fit reasonably a unidimensional representation. The structure implied by a simplex is that the items are quantitatively ordered on some dimension, increasing in intensity. The requirement of the circumplex is that items fit around a circle shape, such as the Shepard (1978) colour circle introduced later in chapter 7 and the Wiggins (1979) interpersonal wheel explored in chapter 10. This assumes that items are equal in intensity, and contain a 'blend' of variance from each content subuniverse, to coin a phrase from Block (1995, p. 189). Naturally, this requires a two dimensional representation at least.

The values for finding the simplex pattern in a two-way one-mode Similarities matrix were given by Guttman (1954b). For the simplex in Figure 3.5, the ranked values of would have a 'contour' of tied values going from the diagonal to the bottom left corner, as illustrated in Figure 3.6.

	t_1	t_2	t_3	t_4	t_5
t_1	-				
t_2	1	-			
t_3	2	1	-		
t_4	3	2	1	-	
t_5	4	3	2	1	-

Figure 3.6. Simplex pattern for configuration in Figure 3.5
Values are ranked on similarity

This would create the perfect equally-spaced data simplex, of which the Spearman (1904) 'tetrad difference' was a special case where the ratio between two pairs of adjacent tests is equal.

Lingoes and Borg (1977) stated that the spatial representation of both equally- and nonequally-spaced simplicial matrices is that of a straight line which may be one-dimensional, which may 'curve' to form a 'horseshoe' or 'C' shape in two dimensions. This was termed a 'geometric simplex' by Borg and Lingoes (1987). (As is later suggested in chapter 7, the use of local monotonicity may help in testing for a simplex. This is because it maps more closely adjacent ('conjoint tetrad') cells which determined the simplex configuration at the expense of the non-adjacent ('disjoint tetrad') cells.)

For the equally-spaced circumplex in Figure 3.5, the ranked values with diagonals ignored in Figure 3.7 would be required.

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	-					
t_2	1	-				
t_3	2	1	-			
t_4	3	2	1	-		
t_5	2	3	2	1	-	
t_6	1	2	3	2	1	-

Figure 3.7. Circumplex pattern for configuration in Figure 3.5
Values are ranked on similarity

In this data circumplex there is a gradual 'sloping off' then increase in similarity towards the lowest left cell. This represents the that fact that the tests become more dissimilar, then more similar going down the test battery from t_1 to t_6.

As will be appreciated from Figures 3.6 and 3.7, the requirements of data values to fit perfectly a simplex or circumplex pattern are exacting. However, deviation from the ideal into a 'quasi-simplex' or 'quasi-circumplex' can be tolerated up to a point, after which the structural hypotheses implied by these representations become weak. Where data sets are noisier - or are not psychophysical as in Shepard or constructed as amalgams of items as in Wiggins - then these patterns may be too strict.

An alternative to rejecting these spatial configurations is to use regional interpretation instead. A circular data point configuration can equally be interpreted as a regional polar facet, as in Figure 3.4, while a simplex can be interpreted as a regional axial facet, provided that the simplex does not curve back on itself too much. Where the simplex does do this then there must be a stronger understanding of the meaning of the representation to test for the simplex structure. This substantive rigour may be justified and achieved through a computational weakening, as is shown in chapter 7.

To summarise, regional structures are more general than spatial data point configurations and are suited to lower reliability data. However, the recognition of data points is necessary in Empirical Study 7.7.

Techniques for Geometric Representations of Data

So far, the empirical representation of data has been considered in an abstract sense. To achieve these geometric representations, what is required is a technique which can translate the similarity information into distance information in the geometric space. Such techniques are generically known as Multidimensional Scaling (MDS) techniques. The nature of the geometric space produced by MDS can be best explained using a geographical analogy for the set of items on which the sample has been tested. In this analogy, the set of tests is a series of point estimates of height in a map, using the landscape around it for reference. The overall design of the map is an approximation of the empirical reality of the landscape, the content universe under examination. The interrelationship of variables as given by the similarity matrix connects the space together in an analogous way as the calculation of the height of all the points is done by a method of triangulation.

The accuracy of the prediction of unknown points is dependent on two factors. These are firstly the number of predictors, such that the greater the number of predictors the greater the accuracy, and secondly the closeness of the predictors to the dependent ('predictand'; Shye, 1978) variable, such that closer predictors are more accurate than distant ones. A third factor, Localised Spatial Bonding, is explored in chapter 7.

If the predictions were applied not to existing points in the plot but the 'empty' space in the plot, then this would give an indication of what would be in the space if an item had specifically tested that idea. The 'gaps' between points in the MDS space are not empty but an extension of existing points, a property termed the Continuity Principle by Borg and Shye (1995). Returning to the geographical analogy, an estimate of the gradient between two height points on a map is assumed to be regular, provided that the points are not physically too far apart and the terrain too unpredictable. The gradient is calculated simply as a linear interpolation between the two points.

Therefore under conditions of Contiguity, all the hypothetical 'unfilled' space is assumed to be a regular continuation of the properties indicated by points in the space that are 'filled' by variables. So for example the hypothetical point halfway between two variables will contain the average of the properties of both the points. The advantage of perceiving the space as continuous rather than discrete is that items may be created that are hypothesised as being that part of the space. The phrasing of such items can be done in conjunction with the meaning of the surrounding items.

Summary of Chapter 3

Facet Theory was suggested to offer a structure-seeking approach to research. It is concerned with *a priori* definitions, hypotheses derived from those definitions and geometric representations to test these hypotheses. The key link between definition and data is Contiguity. In the ideal research case, this can be done formally using Mapping Sentences.

The faceted approach to structure and representation in the ideal case is as follows. Structure indicates what are the constituent parts of a domain and how - if at all - they are ordered. A researcher proposes structural hypotheses to explain the structure. In ideal circumstances, in the faceted approach, these are derived from Mapping Sentences. Observations are made according to the rational item selection procedure, from which data in correlation or association matrices are created. These are mapped into a space of known dimensionality to create a geometric representation. If there exists Contiguity between the structural hypotheses and the empirically-obtained representation, then the space can be partitioned into regions in support of the structural regional hypotheses.

Some geometric regional and data point configurations of structural hypotheses were presented, showing quantitative and qualitative order. MDS techniques which represent these orders were introduced.

Chapter 4
the relevance of faceted representations of criminal actions

Non-Metric Representations of Information

In the previous chapter, the faceted approach to mapping structure onto representation was examined in the ideal case. The ideal case would involve such things as having a Mapping Sentences with which to devise items, high reliability of information sources in terms of both items and respondents, and availability of representative or non-artificially selected samples. However, the nature of the information on criminal actions, as shown in chapter 1, means that the domain is far from ideal.

It has already been suggested that one way to overcome this inherent problem is to analyse separately the constituents of action - namely behaviour and intention - and then to integrate them to a limited degree, given the quality of data. Another complimentary way to overcome 'noisiness' in information sources is to reproduce the structure of a domain in a more modest representation - or using Stevens' scales of measurement, this implies the 'dropping' of a level. This would accept that the actual values for the similarities or correlations carry information on the domain but at a weaker grade than would have been expected in the ideal case. Thus the effects of an extreme outlier in either a variable or a set of cases would be diminished, which would have otherwise skewed the distribution of values in the association matrix.

This reduction of information can be achieved by taking the absolute values in the association or correlation matrix and ranking them. The ranked values are then translated into the geometric space as distances, attempting to preserve the order of values rather than absolute values. This will reduce error where the specific data values derived from the information are meant to be an indication of the degree of association or correlation, rather than a precise measure. In the ideal case however this transformation may not be necessary, since information sources may be stronger, respondents may be more reliable, there will be no outliers, etc. The additional and possibly confounding factor is the coefficient, the importance of which is shown in chapter 8.

There are two distinct types of MDS procedures: those which translate absolute values in the association matrix, known as metric MDS, and those which translated ranked values, known as non-metric MDS (or 'quasi-non-metric'; Coxon and Jones, 1980). If the distribution of associations genuinely were interval or ratio level of measurement, reducing associations to an ordinal level may indeed throw out information. Therefore metric MDS would be more appropriate since it takes the uniform distribution into account. However, if there are possible effects from outliers and unsystematic error in the distribution then the assumption may not be justified, and the use of metric MDS could add extrinsic or statistical non-substantive error (Shye and Elizur, 1994).

Moreover, it has been shown that the ranking of similarities scaled with non-metric MDS to a set of Euclidean distances can still lead to the accurate recovery of the original interval matrix of similarities (Young, 1970). Numerous examples exist of the reliable reconstruction of maps (i.e. representations) where the distances between towns were ranked and then put into non-metric MDS (e.g. Coxon, 1982, for Scotland; Borg and Lingoes, 1987, for Germany; Cox and Cox, 1994, for England). Therefore not much is at risk by using ordinal scaling.

Clearly, then, the importance of non-metric scaling as an extremely close approximation to metric scaling must be emphasised. While at the same time statistically there is no great advantage to metric scaling, substantively there is an advantage to non-metric scaling. Shye stated that 'The nonmetric mapping appears to be good way to avoid attaching undue importance to the numerical values of coefficients' (Shye, 1978a, p. 14). Subsequent research could build on existing knowledge of the relationships in the content universe, Shye continued, and hence more strict mapping conditions could be specified *a priori* in future replications as knowledge accumulated. This could be realised through more facets, more structs and stricter translations from similarities to distances such as linear or power functions.

The role of non-metric MDS in the faceted approach is fundamental (Shye and Elizur, 1994; Canter, 1985). Guttman (1967) stated that 'a healthy consequence of nonmetric analysis is that it will force us to face more squarely the problem of substantive theory construction' (Guttman, 1967, p. 81), which is a prime concern of Facet Theory. The use of non-metric procedures in the domain of criminal actions follows the assertion of Shye (1978a) that successful non-metric scaling should precede attempts at metric scaling.

This particular rationale for non-metric MDS does however contrast with the emphasis placed by Guttman on the use of ranked inputs. When asked why input was ranked, Guttman (1979, p. 4) replied that 'Lingoes chose to do this [i.e. rank the input] for convenience in standardizing the programs'. Guttman further asserted that later programs in the SSA series should use absolute rather than rank values.

Nevertheless Lingoes also commented that

> To reproduce order information, then, requires a simpler and more direct representation of one's data, which facilitates the interpretative process. If our interest is in patterns or configurations, the most natural concept for revealing them is order and the appropriate method for analysis is one which focuses on monontonic transformations. (Lingoes, 1972, p. 52)

One such method that translates ranked similarities matrix as distance information in a geometric space is Smallest Space Analysis-I (SSA-I; Lingoes, 1973). Since Facet Theory is based around the search for regional structure, the non-metric SSA-I is well suited to the task. The mechanics of this procedure are explored in greater detail in the next chapter.

Secondary Information and Partial Content Universes in the Faceted Approach

So far the structural alternative to measurement by fiat and scale construction that pervades much of quantitative psychology has started with a universe of stimuli sampled with a Mapping Sentence (MS). The MSs have been used to construct items such as in a questionnaire that are then tested on the sample. The non-metric faceted approach to research has particular relevance when using secondary or unreliable information, as is found with sources on criminal actions. Furthermore, partial content universes - as introduced in chapter 1 - may mean that not all structuples can be represented, or observations on some parts of the content universe may be poor or missing. Secondary information means recorded observations that have been conceptualised as relations in some way and at some stages in the research process, as modelled with the expanded CRM taken from Coombs (1964).

However, this is not a problem for the faceted approach, and it is proposed that the approach is in fact especially suitable in these cases. It has already been suggested that the faceted approach with non-metric MDS can cope with some of the unreliability of the criminal actions information. Published examples demonstrate other ways in which secondary information and partial sub-universes can be fruitfully used.

For example, the first public demonstration of a facet was in fact a re-analysis of some of Thurstone's data on intelligence testing (Guttman, 1964), even though the precepts of Facet Theory were traced by Canter (1983) to the 1940s. Therefore contiguity in the analysis of secondary data does not represent a problem, as was shown by Guttman (1964) who put Thurstone's original data into the then new SSA-I program. The data here then were of the most restricted kind, since the items are already constructed. Guttman cited an existing MS that hypothesised structure for intelligence tests, unveiling a 'faceted definition' of intelligence. This was a statement about how empirical observations in intelligence testing - namely the results of IQ tests - would be ordered on its facets. Though other facets may also order the observations, but two were examined: mode of communication and substance of communication. Mode of communication had numerical, verbal and geometrical elements; substance of communication had analytical or achievement. These facets were derived from consideration of the existing literature, and formed structural hypotheses.

This way of defining the content universe was clearly the precursor to the MS. Guttman classified each intelligence test under the 'facetisation' (Shye and Elizur, 1994) of a one-to-many mapping of faceted structuples to tests. In terms of the expanded CRM, the original researcher has made the decisions up to and including phase 2, where the stimuli have been designed, valid responses taken and observations are recorded as Quadrant IV similarities data (Coombs, 1964).

The SSA-I program translated the proximity matrix of correlations between tests into distances in space such that the greater the distance between two then the lower the correlation between them. The representation in the SSA-I solution demonstrated Contiguity and therefore could be successfully partitioned by Guttman according to the faceted structure hypothesised. Each intelligence test item was shown to be in a distinct part of the plot. The creation of the MS was an interactive procedure of previous proposals and tests of structural hypotheses.

Guttman suggested that the facets in the two dimensional plane were in the configuration of the radex. That is to say, the facet of 'mode of communication' was a circumplex and the facet 'substance of communication' was a simplex. The circumplex indicated a circular qualitative order, while the simplex indicated an additive quantitative order. These configurations were obtained by examining how the points in space related together, given the *a priori* statement in the faceted definition. These spatial configurations were examined in chapter 3 and also later in chapter 7, but the important point to be made here was that it was a hypothesised structure supported empirically.

The radex itself had been suggested by Guttman ten years earlier (Guttman, 1954b) as providing a means of integrating the theory of intelligence testing. This was the first presentation of these spatial configurations as a results of the SSA-I program. Guttman stated that the radex 'didn't catch hold for several reasons' (Guttman, 1964: 29) - one of which was the fact that there was no computer program to perform the analysis. Seen in the context of measurement as structural hypothesis testing, the idea of the radex needed the computer program to allow it to be demonstrated more easily in larger - and more reliable - data sets than could be calculated by hand. Furthermore, the program needed some systematic way of understanding the dimensional solution it produced. In fact, Guttman suggested that 'Without the faceted definitional system, it would be very difficult to interpret a [multidimensional scaling] plot' (Guttman, 1964, p. 33). Nevertheless, an important structure was found using the faceted approach, secondary information and a non-metric representation.

Where the choice of items or creation of data has been sampled highly selectively then the content universe may only be partially covered in the information source, as was suggested in chapter 1. The solution to the methodological problem of finding contiguity in the partial content universe for Canter and Heritage (1990) was to construct items according to structural hypotheses indirectly. The literature was reviewed extensively and various themes of rape behaviour were drawn from existing theories. The types derived from the literature informed the structural nature of the plot. Distinct inherent contradictions in the types were noted, such as the intention of the offence being assertion of power (Groth, 1979) or an attempt at intimacy (Marshall, 1989). Additionally, the 'confusion of action and person' (Canter and Heritage, 1990, p. 188) was noted in some of the empirically-supported typologies due to the clinical sample, with the understanding of intention being in terms of pyschodynamic motivation.

The items were constructed firstly so as to cover the sorts of behaviour shown by the offenders of different types, and secondly so that the information would have been noted by investigating officers and thus be found in existing case files. (Chapter 9 addresses the issue of item design directly.) A thorough understanding of both the wider theoretical content universe and the practical partial content universe was paramount. The use of many items to test whether the themes are competing or complementary by examining their contiguities acts as a secondary structural hypothesis, where the inclusion of each behaviour as part of a theme in the definitional system must be argued substantively.

'Gaps' in the plot of sexual assaults implied discontinuities in the sampling of the partial content universe, according to the Continuity Principle of Borg and Shye (1995). These may be filled by cross-reference to other points in the plot that identify the facet structure. This requires of course that the plot demonstrates Contiguity between the structure of the partial content universe and the geometric representation.

To conclude this chapter, the implications of a regional interpretation in terms of theory construction and cumulative science are examined.

Structure and Cumulative Science

The faceted approach offers guidance on understanding measurement solutions both at the micro-level and at the macro-level - where many results are integrated and theories constructed. The combined contribution is towards cumulative knowledge.

As Shye put it:

> Unfortunately, social and psychological investigators, in their attempt to attain the rigor of the physical sciences, often resort prematurely to quantitative trimmings in their studies while neglecting another feature essential to all empirical science: formulating a reliable definitional basis for carrying out observations. ... As a results, meaningful replications of empirical studies are usually impossible and little knowledge is actually accumulated towards the formulation of scientific laws. (Shye, 1978a, p. 4)

Guttman suggested that 'replication was the essence of science' (Guttman, 1977, p. 86). Consequently, hypothesis testing should not be carried out as if only one experiment was ever carried out. In the interests of a cumulative science, Guttman argued, the null hypothesis should be dynamic and not a statistical given. For example, if an experiment has rejected a null hypothesis of 'no difference', then surely experiments will not be performed blind of this fact. Future research attempts to replicate that finding by testing its existence. But this can only be achieved by stating the previous research finding as a null hypothesis, not an alternative hypothesis.

It is fair to say that the use of significance-based hypothesis testing has always been controversial. It was originally designed by Fisher in the 1930s to compare the relative effects on crop production from factors such as fertiliser, and now forms the hegemonic methodology in one form or another for nearly all social research. This despite repeated pleas to recognise its shortcomings and move to more appropriate methods (e.g. Cohen, 1962; Guttman, 1977; Carver, 1978; Schmidt, 1996). In fact, at the present moment a Task Force has been set up by the American Psychological Association Board of Scientific Affairs to investigate the state of null hypothesis significance testing in psychology and how better research practice can be taught.

There is one particular difficulty faced by statistical significance testing in real-world research (RWR), such as the domain of criminal actions. This is its bias towards finding a null hypothesis of no effect, and minimising the error associated with falsely reporting that there is an experimental effect (i.e. Type I error, α). The bias is towards a conservative estimate of not reporting any effect unless it is certain, which is important where - strictly speaking - theories could be rejected contrary evidence is reported. However, in real-world research (RWR) the bias is towards finding out what does rather than does not work - testing the experimental hypothesis. Therefore the issue of statistical power is more relevant, and minimising the false reporting of no effect when in fact there is (i.e. Type II error, β). Despite this, Schmidt (1996) stated that much RWR and virtually all academic research does not concern itself explicitly with statistical power. As is shown in chapter 6, the faceted approach emphasises the importance of power.

Various solutions are available to counter this major shortcoming with statistical significance testing in RWR. These include the use of point estimates and confidence intervals, as were originally used before Fisher's work. Also widely touted as a solution is meta-analysis, where several studies are amalgamated to increase statistical power. For RWR, however, these solutions do not get away from the fact that many statistical procedures are simply inappropriate given contextual issues. There are no guarantees that other heavily statistical methods may overcome such difficulties.

Lingoes (1981) stated that '[s]omething more than the development of new techniques for blindly analysing data [is] needed for stemming the needless stream of one forgettable empirical exploratory study after another'. Canter added that this 'something more' must connect with 'daily experience and human action ... [and] be part of the cumulative development of systematic explanations and understanding' (Canter, 1985, p. 10).

Elsewhere, Canter (1983) suggested that there were four key aspects for practical problem solving that must be addressed by effective applied social research. These were:

1. The categorical nature of policy-making and action-taking.
2. The multivariate structure of human experience and action.
3. Often action is facilitated if the options available can be specified.
4. The concepts being examined need to be specified in other than operational terms.

These aspects are different manifestations of the peculiar context of RWR. The only time RWR is commissioned is when a specific shortcoming has been identified or a specific problem needs to be addressed. Consequently, research should be framed by the context for which it will be used. This requires operationalisation using Canter (1993) used the notion of developing 'organic' research projects for which 'the context within which the data is being collected and drawing upon that context in a clear-sighted way' (Canter, 1993, pp. 46-7). Specifically, Canter (1993) suggested that 'organic' research must be appropriate to the context, in order to be scientifically - and economically - parsimonious. Therefore observers or data gathering procedures must be non-obtrusive or there is the possibility of confounding the very subject under study. This could be achieved by taking advantage of peculiarities of the context in ethnographic-style work, or by using of existing information or records held by institutions. Finally, issues under examination must be defined within the context so that analysis is directly relevant. Consideration of such aspects allow the link between RWR and action to be better integrated.

The faceted approach is a methodology which addresses all these issues and particular difficulties of RWR. It is a multivariate approach and can contribute directly and helpfully towards police investigations. It is eclectic in the formulation of hypotheses and useful in the confirmation or rejection of these hypotheses.

Summary of Chapter 4

This chapter argued that the faceted approach had relevance for the creation of representations and analysis of structure in the domain of criminal actions. This domain was suggested to be far from the ideal case, being that it is often contains unreliable sources, secondary information and partial content universes. Nevertheless, non-metric MDS and regional interpretation can overcome many of these problems and still contribute towards a cumulative body of knowledge. It was stated that the approach is also geared towards the recovery of information relevant for real-world contexts, taking into consideration the nature of applied research.

Chapter 5
the method of ssa-i and other mds programs

A History and Typology of Scaling Methods

The previous chapter introduced SSA-I as a suitable procedure for analysing criminal actions. This chapter expands on the method of SSA-I itself, exploring how it achieves a geometric representation of data and comparing it to similar MDS models.

Two classifications of types of MDS procedures are those of MacCallum (1988) and Coxon (1982). MacCallum characterised MDS procedures on the basis of 'model' and 'method'. The 'model' refers to the 'set of rules specifying a correspondence between data and some parsimonious representation of the structure of the data' (MacCallum, 1988, p. 424), such as metric or nonmetric. The 'method' on other hand refers to the 'procedure for fitting the model to the data' (MacCallum, 1988, p. 424), concerning algorithm and fit. Coxon, however, suggested the various procedures should be understood in terms of data, transformation and model. 'Data' concerns the input to the program, the 'transformation' is roughly what MacCallum meant by 'model' and Coxon's 'model' is the nature of the output of the program.

There are advantages to both descriptions, and when combined, the explanatory power is increased. MDS procedures may be characterised on the following facets:

1. *model*: the nature of the distances transformed, as in Coxon's 'model',
2. *transformation*: the scaling function on the data input, as in MacCallum's 'model' and Coxon's 'transformation',
3. *data input*: the matrix entered for analysis, as in Coxon's 'data', and
4. *algorithm*: the computational procedure to do the scaling, as in MacCallum's 'method'.

The historical development of MDS procedures can be seen primarily as a new exploration of each of these successive facets. The first stage was the classical or metric scaling approach, also characterised by Shepard (1972) as the 'Princeton' approach which specifically involved Gulliksen, Messick, Abelson and Torgerson in the 1950s. The key distinguishing feature of this phase of research was the move away from analysis of vectors in geometric space - as in factor analysis - towards the analysis of distances in a geometric space. In terms of the scheme, this phase explored new forms of geometric *model.*

This new form of model was taken further in the second phase of Shepard (1972) concerned the *transformation.* Shepard developed the 'analysis of proxmities' for spatial points such that the data were connected by some 'notion of psychological nearness, closeness, or degree of proximity' (Shepard, 1962, p. 126). Termed by Shepard (1972) the 'Shepard-Kruskal' approach, it involved transformation functions other than linear, principally the non-metric monotonic transformation. The approach was pioneered and the algorithm further developed by researchers at the Bell Telephone Laboratories such as Kruskal, Carroll and Wish in the 1960s. Shepard also mentioned the contribution to the non-metric approach of Lingoes and Guttman in Michigan and Roskam in Nijmegen. Young (1987) suggested that this stage seemed somewhat 'magical' in its ability to use weaker non-linear transformations to achieve virtually the same quality representation as did the 'Princeton' metric approach.

Young (1987) stated that the third phase in the 1970s was the work on individual differences scaling, as pioneered by Carroll and Chang (1970). This work was centred around the generalisation of *data input* part of the scheme to include individual contributions to the geometric representation as well the group aggregate. It was shown that scaling programs need not be restricted to low order arrays and could be taken to higher, more complex arrays of arrays (Arabie, Carroll and DeSarbo, 1987). One program, ALSCAL, was put forward as a unified procedure encompassing all possibilities of data input as well as other combinations of representation and transformation (Takane, Young and de Leeuw, 1977; Young, 1981).

Young (1987) suggested that the most recent phase in the 1980s was the work in Maximum Likelihood Estimation (MLE) scaling by Ramsay (1982). This development has centred mainly around new forms for the *algorithm*. Maximum likelihood scaling proposes an error distribution for a scaling solution, which if appropriately used allows inference about the correct number of dimensions and the bounds of fit for individual points in the configuration.

The Development and Place of Guttman-Lingoes SSA-I

Within the second phase of Shepard (1972) - concerning the development of MDS transformations - there were two main types of program developed to perform ordinal scaling of two-way one-mode matrices. These were the programs MDSCAL (Kruskal, 1964b) and G-L SSA-I (Guttman, 1968). Lingoes and Roskam (1968) stated that these two MDS programs were developed independently, but both were intended as improvements on the computational procedure put forward by Shepard (1962). This was because both styles incorporated statistical loss functions implemented as computational procedures to improve the solutions, hence better achieving the goal of MDS of decreasing dimensionality.

G-L SSA-I was part of the Guttman-Lingoes 'Smallest Space Analysis' (G-L SSA; Guttman, 1968). Although published after Kruskal's MDSCAL (Kruskal, 1964b), it was reported that Guttman had claimed his 'first thoughts on the topic of nonmetric multi-dimensional scaling occurred in the late 1930s' (Young, 1987, p. 23). It is clear that Guttman's thoughts on computational aspects of the program series were given earlier than 1968, for example with the program outline for G-L SSA-I being given in Lingoes (1965), and Guttman (1964) stating that the routines for a suite of programs were in press at *Psychometrika* and would be published in 1965, even though they appeared in 1968.

A suite of eight programs was available in the G-L SSA series consisting of four each for the analysis of square and rectangular data matrices. Two other series types were also available: three 'Multidimensional Scalogram Analysis' (MSA) procedures for analysis of qualitative data (Lingoes, 1968a); and three in the Conjoint Measurement (CM) series, totalling 14 programs to perform a variety of analyses on different types of data input (Lingoes, 1968b). These 14 programs were implemented jointly by Guttman and Lingoes, thereby initiating the addition of the rubric 'G-L' to all programs (Guttman, 1967b).

Roskam and Lingoes (1970) introduced the first of a new series of programs starting with MINISSA-I, denoting '*M*ichigan-*I*srael-*N*ijmegen-*I*ntegrated-*S*mallest-*S*pace-*A*nalysis'. The '*N*' was also intended to denote 'New Jersey', in honour of Kruskal (Roskam and Lingoes, 1981). MINISSA allowed the choice of algorithms and loss functions of Kruskal, Roskam and Guttman - hence 'Integrated' (Roskam and Lingoes, 1970, p. 204). MINISSA-I was proposed to have three advantages over MDSCAL (Kruskal, 1964a) and the early G-L SSA-I (Lingoes, 1965), namely finding globally optimal solutions, computational speed and a range of options for scaling (Roskam and Lingoes, 1970; Lingoes and Roskam, 1973). Roskam and Lingoes (1970) also released an updated version of SSA-I, incorporating the improvements in MINISSA-I though without the full range of options.

After this, Roskam and Lingoes then separately developed their own program series in Nijmegen and Michigan respectively. Lingoes (1973) specified that his MINISSA-I version was to be known as MINISSA-I (M), referring to Michigan. This was 'equivalent to SSA-I' (Lingoes, 1973, p. 43), presumably referring to the SSA-I revised and updated in Lingoes and Roskam (1973). Collectively, the Nijmegen series is known as MINISSA (N), according to Lingoes and Roskam (1973). It was noted that 'The main features which distinguish MINISSA-I (M) from MINISSA-I (N) are the variety of options offered for data input and analysis' (Lingoes and Roskam, 1973,p. 81).

A shortened MINISSA-I was also designated as MNSSAST by Roskam (1977). Roskam and Lingoes (1981) stated that the original unshortened MINISSA-I developed by Roskam was known in the Nijmegen series as MNSSAOR. (However, Roskam and Lingoes (1981) had reported that Roskam had previously renamed his versions such that SSA-II was MINICPA and SSAR-II was MINIRSA.)

Lingoes (1972) had attempted to redesignate the notation of the program series into more descriptive notation, such that SSA-I would become 'MDA-U', referring to 'monotone distance analysis with matrix unconditional', and SSA-II would become 'MDA-C', 'monotone distance analysis with matrix conditional'. This plea was not heeded, although later additions to the series such as SSAR-V were also given Lingoes' redesignation in Lingoes (1977c).

The time line in Figure 5.1 shows the important developments for SSA-I and for non-metric multidimensional scaling more generally.

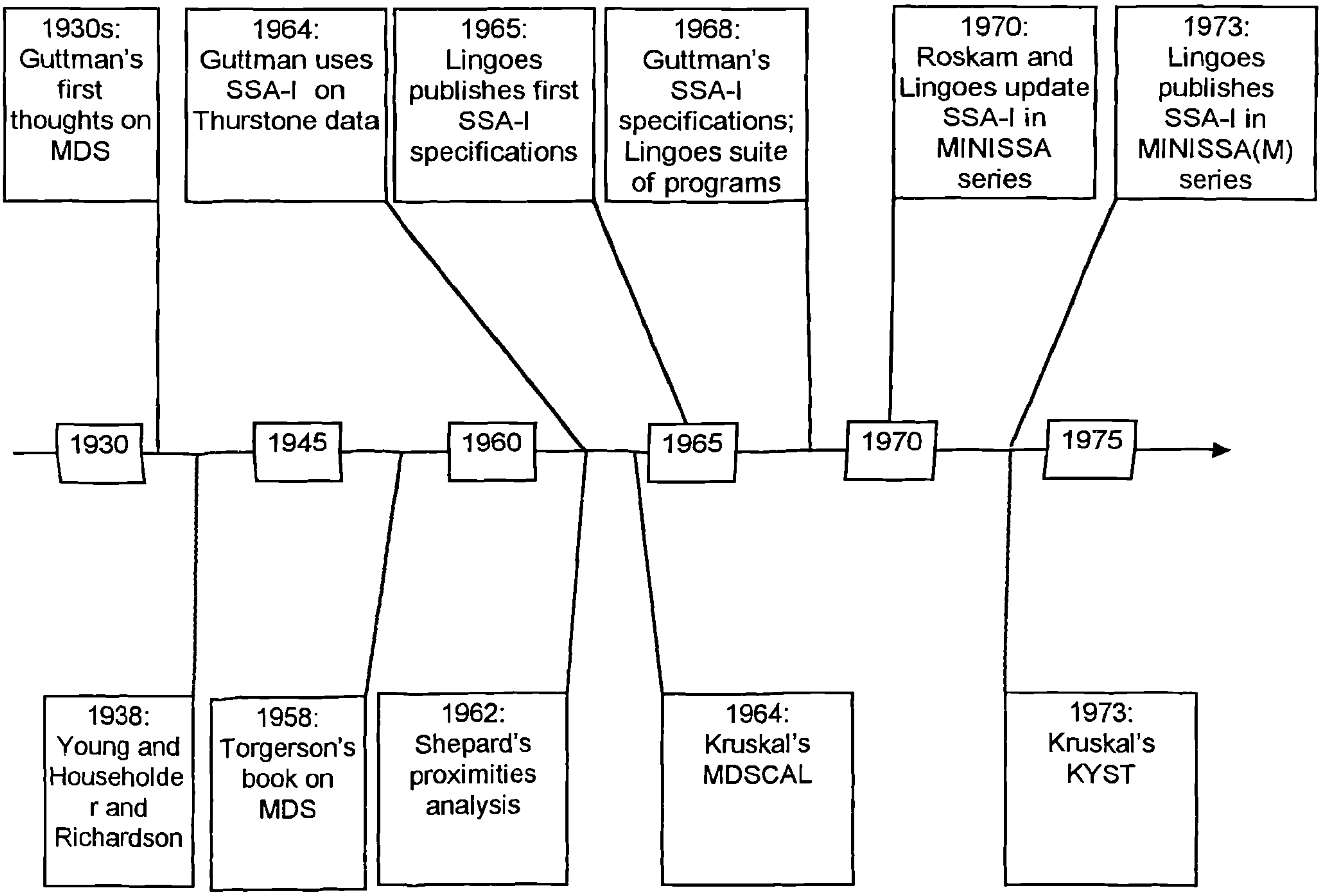


Figure 5.1. Time line for development of SSA-I and other MDS programs

Given this minefield of acronyms and developments, it was felt important to establish comparability between the SSA-I program implemented at Liverpool and other published examples.

Test Runs on the SSA Program Available at The University Of Liverpool

The original precursors to the SSA-I program available on the University of Liverpool UNIX mainframe were from Lingoes (1965) in collaboration with Guttman. Guttman and Lingoes developed the programs to the specifications of Guttman (1968) which form the basis of the Coefficient of Alienation in the Liverpool SSA-I. Lingoes and Roskam went on to work together on their own MINISSA program series, culminating in MINISSA-I (M) by Lingoes (1973). This was an updated version of G-

L SSA-I with better algorithms and more computational power (Roskam and Lingoes, 1981), which are used in the Liverpool SSA-I.

The Liverpool SSA-I is functionally equivalent to MINISSA-I (M), since it has the full range of options in Lingoes and Roskam (1973, pp. 81-93). It therefore incorporates:

- input matrix of similarities or dissimilarities,
- local or global monotonicity,
- minimisation of alienation with semi-strong monotonicity or stress with weak monotonicity,
- choice of user defined starting configuration or Guttman-Lingoes-Roskam quasi non-metric configuration, and
- Euclidean or Manhattan distance metric.

The Liverpool SSA-I was given a 'pseudo-interactive' interface by Sean Hammond in 1984 which creates an input batch file fed into the main program which then returns an output file in UNIX. It also has the following function added:

- calculation of one-mode association matrix from two-mode data matrix using Jaccard's, Yule's Q or G Index for dichotomous data, or Pearson's *r* for polychotomous data.

More recently, a fully interactive Windows version LiFA has been implemented by Malcolm Huntley which updates the results of analyses continuously so the user has more control and can perform on-line analyses quickly. The above functions and options are all available on the LiFA SSA-I, which is identical to the UNIX SSA-I other than in its user interface.

Test runs with data supplied in Lingoes (1973, p. 75) show that the Liverpool UNIX SSA-I does produce a similar output to MINISSA-I (M). However, the Liverpool configuration was reflected through the second dimension (y-axis). Once the sign of the first dimensional coordinate (x coordinate) is changed to simulate the reflection in $x = 0$ then there is a tolerance of roughly $\pm$ 1.5 in the value of the coordinate from -100 to +100. This does not crucially change the configuration or the regional interpretation of it, but it does change in the alienation value at the fourth decimal place.

Another test run using data from Guthrie (1973) given in Borg and Lingoes (1987) revealed alienation different again at the fourth decimal place and a virtually identical configuration, except for a sign change in the y coordinate, indicating a reflection in the $y = 0$ line. This was a similar reflection in the configuration to the Lingoes (1973) test data above, except in a different axis.

However, a test run with data taken from Schiffman, Reynolds and Young (1981, p. 89) gave a similar configuration without reflection. The alienation value was identical to five decimal places and the difference in coordinates was only to the third decimal place. It should be noted that this small change coordinates may not affect alienation to any extent unless there are many points involved in the calculation.

From these test comparisons, it can be seen that the Liverpool UNIX SSA-I gives identical representations to other implementations, though the procedures are slightly computationally different. The reflections cited above are permissible and are invariant with respect to both the configuration and the axes' accounting for variance. Huntley (personal communication) has suggested that the slight differences in values may be due to precision and floating point differences between implementations, and it should be noted that Liverpool UNIX SSA-I is stable over all replications known to the author.

The Nature of Smallest Space Analysis

The mechanical process by which SSA-I - including the Liverpool version - actually goes about constructing a solution is best thought of as a series of stages, connected through a flow chart (e.g. Coxon, 1982). This is shown in Figure 5.2. As the flow chart suggests, an SSA-I run may be iterative, though not inevitably. In practical terms, however, the program will be iterative unless the initial approximation is extremely good, or perhaps if there were few variables and many dimensions.

The broad features of SSA-I will now be explained, with the main differences between SSA-I and other MDS program equivalents explored in a later section.

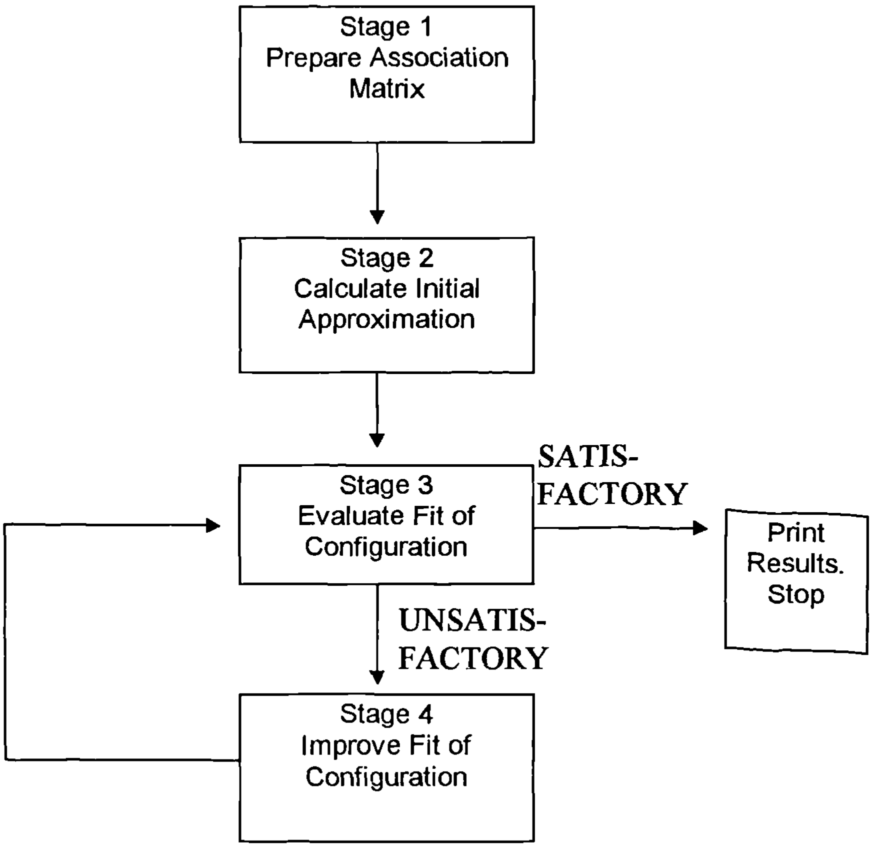


Figure 5.2. Flow chart to describe SSA-I iterative process (adapted from Coxon, 1982)

Stage 1: Prepare Association Matrix

SSA-I employs an unconditional, symmetric, one-mode two-way matrix. The matrix contains associations between points that may have been measured in a variety of ways. The key is that each value but must be either an index of how closely related on unrelated points are. The relation between objects or variables may therefore be of either similarity or dissimilarity provided it is consistent. These associations may have been derived from a two-mode two-way matrix, using correlation or association coefficients. In the present text, 'association' rather than 'proximity' will be used to describe this input matrix to emphasise the fact that most of the two-way one-mode matrices used in the thesis have usually been derived by associating or correlating variables in a two-way two-mode matrix, rather than as a direct judgement of similarity or dissimilarity.

SSA-I requires objects in the one-mode two-way matrix to be symmetric. Two objects i and j are symmetric if their similarities are related such that $s_{ij} = s_{ji}$. If all the objects in a one-mode matrix are symmetric, then relations may be summarised by reproducing only one half of the association matrix, typically the lower half. The shape of this lower-half matrix is triangular. The values of the triangular association matrix are then ranked by SSA-I, being a non-metric procedure.

Stage 2: Calculate Initial Approximation

The second stage is concerned with translating the ranked values in the triangular matrix into a set of distances which can then be displayed in a geometric representation. It must be noted that the separate substantive issue of the Contiguity between conceptual similarities in the domain and the empirical closeness of similarity values is *not* addressed by the program.

However, the translation of points into the geometric space cannot be achieved straightaway by the SSA-I program: first, an approximation to the solution must be calculated to start the ball rolling. Therefore the second stage is concerned with finding a trial starting configuration for the data, requiring an 'initialisation routine' (Schiffman, Reynolds and Young, 1981). To the extent that the overall process is algorithmic, this stage is heuristic. Nevertheless, it is important to obtain an accurate rational initialisation routine (Lingoes and Roskam, 1973). In SSA-I, this is a principle components analysis of the ranked association matrix where the first *m* components of the matrix are taken to give co-ordinates to the *m* dimensions of that particular run (Guttman, 1968). This only occurs for the first dimensionality chosen for a run in SSA-I, as 'dimensional slurring' occurs at higher dimensionalities when the initial configuration in the higher dimensionality is based on the lower dimensional final configuration (Lee and Canter, submitted).

Stage 3: Evaluate Configuration

The third stage is to evaluate how well the present configuration matches the target of an ideal, perfect configuration. This stage forms the first phase of the 'double-phase' algorithm described by Guttman (1968), compared to playing 'Ping-Pong' by Roskam and Lingoes (1981).

Following the nomenclature of Shepard (1972) and Borg and Shye (1995), the ideal perfect configuration for SSA-I would be where the ranked matrix of associations were transformed into distances in a geometric space such that if a similarity were ordered $s_{ij} > s_{kl}$ then corresponding distances would be $d_{ij} < d_{kl}$ if i, j, k and l are variables. In SSA-I, if similarities were different then the distances should not be equal i.e. if $s_{ij} = s_{kl}$ then $d_{ij} < = > d_{kl}$.. However, if similarities were equal then the distances may or may not be equal -termed the primary approach (Kruskal, 1964a). The ordering is therefore semi-strong monotonicity (Guttman, 1986). Consequently, points (variables, attributes) that are highly associated in the association matrix are placed closer together in the geometric space.

The values of the distances in m dimensions are calculated using the standard Euclidean metric distance function. City-Block metric may also be used in SSA-I, where this is substantively justified. The Euclidean distance d_{jk} between Cartesian points j and k is given in Formula 5.1.

$$d_{jk} = \sqrt{\sum_{m} (x_{jm} - x_{km})^2}$$

Formula 5.1. Euclidean distance between two points j and k

The distances between each and every point is calculated as a distance matrix. Naturally, the distance matrix is the same size or rank as the association matrix. SSA-I then uses a procedure known as Guttman's Rank Image principle to evaluate the configuration (e.g. Guttman, 1968). The following description of the Rank Image principle is taken from Coxon (1982) and Borg and Lingoes (1987).

The ranks of values in the association matrix are rearranged so that they are ranked in order of size, as if in a long line. The order in which the associations of pairs of variables is noted. The distances in the current distance matrix are matched with the corresponding association point pair in the line. In a good configuration, the ranks of association values numbered 1 ... n (n - 1) / 2 would be roughly matched by a steady increase in distances values, since high association is ideally represented by small distance. To evaluate how well this has been achieved, a duplicate of the line of distances is made and then rearranged so that it is in perfect rank order, just as the original associations were. This new line is the set of 'disparities' or 'fitted values', known as d^* for Guttman's rank image principle. These are also denoted by Borg and Lingoes (1987) as 'rank-images'.

The more the line of these disparities has to be shuffled reflects the mismatch between ranked associations and the corresponding distance line. If the configuration was perfect, the disparities would not have to be shuffled at all. If the configuration was absolutely wrong, the disparities would have to be completely reversed. The statistical measure of the error is simply to take the squared difference between the ranked values in the line of distances and the corresponding ranked values in the line of disparities. The sum of the squared distances raw phi, ϕ_0, is an overall measure of fit between distances and disparities (Roskam and Lingoes, 1981).

$$\phi_0 = \sum_{jk} (d_{jk} - d^*_{jk})^2$$

Formula 5.2. Raw ϕ_0 for distances d and disparities d^*

The error measure ϕ_0 is normalised so as to be independent of the size of configuration, otherwise it could always be decreased and therefore improved by reducing the scale of the configuration. This would result in all points being very tightly 'clustered' in a degenerate solution so the norming factor is the sum of each distance in the distance matrix squared, as given in Formula 5.3.

$$\phi = \frac{\phi_0}{\sum_{jk} d_{jk}^2}$$

Formula 5.3. Normalised ϕ for distances d and raw ϕ_0

The calculation of error in the Kruskal approach is known as stress (Kruskal, 1964a). Since stress is to be later iteratively minimised, it helps computationally if the value is made systematically larger. Therefore the squareroot of stress is taken, which known as stress_1 (e.g. Coxon, 1982) or SFORM 1 (e.g. Lingoes and Roskam, 1981).

When used with rank-image disparities, the form of ϕ_0 can readily be converted into Guttman's monotonicity coefficient μ (Guttman, 1968) in its semi-strong form. μ is given in Formula 5.4.

$$\mu = \frac{\sum_{jk} d_{jk} \cdot d^*_{jk}}{\sum_{jk} d_{jk}^2}$$

Formula 5.4. Guttman's μ for distances d and disparities d^*

Guttman's semi-strong μ can be equivalent to Pearson's ϕ, the Product Moment Correlation Coefficient when the relationship between the bivariate distributions under examination is perfectly linear. Guttman's μ varies between -1 and 1. Since μ can approach unity even when there is a large amount of scatter in the data, for practical reasons it is preferable to subtract μ from 1 and then squareroot it to obtain K, the Coefficient of Alienation. This is given in Formula 5.5.

$$K = \sqrt{1 - \mu^2}$$

Formula 5.5. Alienation K using Guttman's μ

Thus alienation K acts as a measure of the 'unexplained variation' (Borg and Lingoes, 1987, p. 46) and ranges from 0 and 1, where 0 indicates a perfect fit i.e. no variance unexplained.

The evaluation of fit at the third stage between output distances and input associations (proximities) by means of disparities leads to one of two outcomes: that the alienation value is satisfactory and therefore the program may stop, or that the alienation value is not satisfactory and therefore the program must attempt to decrease alienation. For alienation to be satisfactory at this stage after only the initial approximation, one of two conditions must hold (Coxon, 1982). The first is that alienation is zero and the second is that alienation is 'acceptably close' to zero. However, it is unlikely - though not impossible - that the program would stop at this stage, if the initial approximation were accurate. If there were no more iterations the results are then plotted and printed. The orientation of the output is calculated by extracting m independent principal components for the m-dimensional solution (Huntley, personal communication). The first axis thus accounts for the most variance, the second accounts for the next most, etc. Some approaches take the fact that principal components are orthogonal in the space into consideration in the interpretation of the plot.

If alienation is not satisfactory the SSA-I program tries to improve the configuration and therefore decrease the Coefficient of Alienation.

Stage 4: Improve Configuration

The fourth stage of SSA-I is an attempt to find what would be the best step to improve the plot. This is the first stage of the Guttman (1968) double-phase algorithm - the ping of Roskam and Lingoes (1981) - namely finding a new configuration.

The alienation value is composed of the errors in the transformation between associations and distances. The square difference between disparity and distance shows which point pair(s) contributes most to the overall sum of squares of error, ϕ_0. There is then a possibility of prioritisation for improvementusing local monotonicity, which is shown in chapter 7 as being necessary for the faceted approach to testing structure. Differential calculus is used to derive the magnitude and direction of the movement required to improve these badly fitting points. The shuffling process of fitting disparities to distances is known as the 'soft-squeeze' (Guttman, 1968).

The appropriate changes are made in the geometric configuration in the hope of improving the fit and therefore the decreasing alienation. In SSA-I, this improvement and evaluation loop is performed five times to check that the proposed change to the configuration is worthwhile (Lingoes and Roskam, 1981, p. 367). This allows the program to be 'far-sighted', so it can overcome short-term losses (i.e. increase in alienation) for the sake of long-term gains (i.e. decrease in alienation). This also serves as a counter-measure against local minima, where the program gets stuck in a rut of low alienation which in fact could be lowered further if the program was given a 'push'.

Stage 5: Re-evaluate Configuration (Repeat Stage 3)

The fifth stage is essentially to return to the same as the third stage above, namely to calculate the distances in the current configuration and evaluate the fit with the original association matrix using the Rank Image Principle. Then the program improves the configuration again, if necessary. This is the iterative loop indicated on the flow chart in Figure 5.2, and is undertaken by SSA-I if the current alienation value is not satisfactory and it shows promise of improving within five iterations.

The conditions under which alienation may be termed satisfactory are again that alienation is zero or 'acceptably close' to zero or that alienation has not improved sufficiently in this iteration to continue (Coxon, 1982).

These iterations are done in an attempt to achieve a global minimum, which the absolute best solution that can be found. However, if the iterations fall into a local minima, where the solution seems to achieve a global minimum but in fact fit could be improved, then the solution given by the program is sub-optimal. This is demonstrated Figure 5.3.

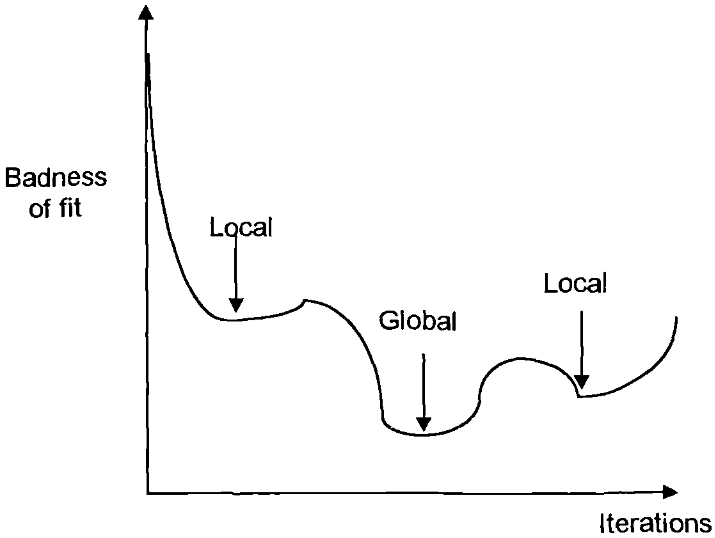


Figure 5.3. Change in fit during iterations in hypothetical MDS run, showing minima (adapted from Coxon, 1982)

The iterations of the double phase algorithm of improvement and evaluation decrease alienation exponentially such that the last few iterations make little difference to the actual distances in the plot (Coxon, 1982). Once the iterations have stopped, when alienation is satisfactory then the results are plotted and printed, having been plotted according to principal components (Huntley, personal communication).

The Place of Smallest Space Analysis in Relation to Other Scaling Methods

As was noted before, SSA-I is part of a larger family of MDS procedures, not least on account of the differences within SSA and MINISSA series. There are several key variations, and several key similarities. This section compares and contrasts SSA-I with other MDS programs, most notably FSSA (Shye, 1991) and MINISSA (N) (Roskam and Lingoes, 1981) from the MDS (X) series (Coxon, 1982). It outlines key changes in accordance with the stages used above.

Differences in Stage 1 - Prepare Association Matrix

For the metric models of the 'Princeton' approach (Shepard, 1972), the transformation from data to distances preserved the ratios between the distances and similarities. Consider as an example, a metric transformation for one-mode triangular matrices. If objects *i*, *j*, *k* and *l* are variables in the association matrix then the transformation is such that $s_{ij} / s_{kl} = d_{ij} / d_{kl}$,.

By contrast, the 'Shepard-Kruskal' (Shepard, 1972) used a transformation of one-mode, two-way matrices of similarities into distances in a geometric space which was monotone, not linear. Thus if the similarities are ordered $s_{ij} > s_{kl}$ then the distances are ordered $d_{ij} < d_{kl}$. The metric transformation does not rank the association matrix.

Another important difference among MDS programs is in the nature of the data input matrix. While SSA-I uses a symmetric one-mode two-way matrix, other program use higher modes and ways. INDSCAL (Carroll and Change, 1970) for example could use a two-mode three-way matrix for separate respondents, comprising of an array of the respondents one-mode two-way matrices. These could then be analysed to investigate the 'weights' each individual places on the dimensions derived from the overall aggregated plot of all respondents.

While SSA-I removes a mode from two-way two-mode matrices, other programs may employ the original, two-mode two-way matrix, and some programs use higher than two-mode two-way matrices (see e.g. Kruskal and Wish, 1978; Schiffman, Reynolds and Young, 1981; Arabie, Carroll and DeSarbo, 1987). In fact, the ALSCAL program (Takane, Young and de Leeuw, 1977) employs the same routines for all the different input matrices and styles of analysis.

Additionally, non-symmetric matrices may also be used by other programs.

Differences in Stage 2 - Calculation of Initial Approximation

There are differences in the 'initialisation routine' used by MDS programs to calculate the initial approximation. As stated before, SSA-I uses a principle components analysis of the ranked association matrix where the first *m* components of the matrix are taken to give co-ordinates to the *m* dimensions of that particular run (Guttman, 1968). This has been described as a 'quasi non-metric initial configuration' (Coxon, 1982). When SSA-I is required to increase dimensionality to $m + 1$ then it uses the *m* principle axes of the previous solution as an initial approximation in $m + 1$. There is a separate issue in itself as to whether this is acceptable in terms of avoiding local minima, discussed in Lee and Canter (submitted).

Other programs work from high to low dimensionality, such as MINISSA (N) in the MDS (X) series (Coxon, 1982) and MDSCAL. In these programs moving from *m* to $m - 1$ dimensions involves removing the *m*th co-ordinate from the final configuration to use the first $m - 1$ co-ordinates as an initial approximation for the lower dimensionality.

Kruskal (1964a; 1964b) originally suggested using an arbitrary start for MDSCAL, plotting the points in an 'L-shape'. However, the inappropriateness of this start was demonstrated by Lingoes and Roskam (1973), and consequently the revision to MDSCAL by Kruskal, Young and Seery (1973), KYST (*K*ruskal-*Y*oung-*S*hepard-*T*orgerson), took this into consideration. KYST uses a metric initial configuration - even though the association matrix may then be ranked - which avoids local minima except where the linear transformation is severely violated (Arabie and Boorman, 1973; *op. cit.* Coxon, 1982).

Differences in Stage 3 - Evaluation of Configuration

SSA-I uses the double-phase algorithm, though earlier versions used a single-phase algorithm to evaluate and improve at the same time (Lingoes and Roskam, 1973). Furthermore, it uses the 'soft-squeeze' method for minimising alienation with strong monotonicity. The Kruskal (1964a; 1964b) method uses the 'hard-squeeze' method.

Had tied similarities been intended to be tied as distances then it would require strong monotonicity with rank-image disparities (Guttman, 1968; 1986), which would follow the secondary approach (Kruskal, 1964a) as opposed to the primary approach used in semi-strong monotonicity. Coxon (1982) termed primary and secondary tying options the 'indeterminate' and 'equivalence' approaches respectively. Kruskal pointed out that it was 'deep inside the algorithm for finding the [distances]' where this distinction has its computational importance (Kruskal, 1964a, p. 22).

Weak monotonicity (Guttman, 1986) may be used instead of semi-strong and is most associated with the Kruskal approach. Thus if similarities are ordered $s_{ij} > s_{kl}$ then the distances are ordered $d_{ij} <= d_{kl}$ if i, j, k and l are variables. Thus similarities that were not identical may be tied as distances. Weak monotonicity uses the primary approach to ties. Metric scaling always uses the secondary approach (Borg and Lingoes, 1987, p. 39).

The resultant value sum of squared differences between distances and disparities from Kruskal's monotone regression is different to that obtained from rank-images. This is because the rank-images are stricter in monotonicity than monotone regressed disparities, even though calculation of the actual sum of square is identical. For Kruskal (1964a) and MDSCAL/KYST, this is termed 'raw stress'. Similar to the raw ϕ_0 obtained from rank-images, raw stress is normalised so as to be independent of the scale of configuration, lest degeneracy in the solution should occur. The norming denominator for raw stress is the sum of each distance in the distance matrix squared, giving the error measure known as stress (Kruskal, 1964a). The squareroot of stress is taken, which known as stress_1 (e.g. Coxon, 1982) or SFORM 1 (e.g. Lingoes and Roskam, 1981). Squarerooting the value helps computationally by making it larger, since it is to be later iteratively minimised just as K, the Coefficient of Alienation in SSA-I (Guttman, 1968).

The calculation of error of fit between distances and disparities can be either stress_1 or K for monotone regression and rank-images. However, by convention stress_1 is for monotone regression while K is for rank-images, though the actual value of these two can sometimes be identical or at least monotone with each other (Roskam and Lingoes, 1981). In other words, the important feature is the method of creating the disparities, not the measure of fit. Furthermore, the minimisation of these error measurements is not exclusive: MINISSA (N) in the MDS (X) series (Coxon,

1982) uses both principles and coefficients of fit in succession, starting with a rank-image soft-squeeze of K followed by a monotone regression hard-squeeze of $stress_1$.

Fit using monotone regression may be seen in the 'Shepard diagram' of original associations or correlations against actual distances. This scatter plot shows the disparities (fitted distances) with the monotone disparity line typically curvilinear in shape and going through the centre of the points in the scatter plot. In the monotone regression principle of Kruskal (1964a), the error from all points to the disparity line is minimised in a similar way to the least-squares principle in linear regression. The obvious difference is that the fitted line is linear in regression, but is monotone in monotone regression.

Differences in Stage 4 - Improvement of Configuration

Both SSA-I, MDSCAL and KYST use the same method to improve the configuration, namely the method of steepest descent (Kruskal, 1964b). As with the choice of measurement of departure from distances to disparities, the principle is the same for these different programs. However, other methods have been developed for improving the configuration, and newer ones are still being developed (e.g. simulated annealing).

Some SSA packages offer the user the ability to weight some values in the data input so they may be prioritised for optimisation of fit. This is termed local monotonicity, where item pairs of high association - therefore short distance - are prioritised for improvement, while in global monotonicity item pairs are prioritised purely on the basis of the size of the error. A hypothetical distal monotonicity was posited by Lingoes and Roskam to be readily programmable from the formula from local monotonicity, the meaning of which was to favour long distances at the expense of short ones in the opposite way to local monotonicity. The choice of local as opposed to the more usual global monotonicity increases alienation, therefore requiring some justifiable reason which was made on substantive grounds.

The importance of local monotonicity has not been explored elsewhere in terms of the faceted approach to finding regional structure. Although FSSA (Shye, 1991) offers a range of 11 locality weightings, the default value of +2 local weighting has been taken as a matter of course. Liverpool SSA-I offers one weighting value of local monotonicity. MINISSA in the MDS (X) series (Coxon, 1982) does not even offer any weighting.

The idea is later developed that the impact of local monotonicity is important in the search for regional structures in the context of facet theoretic interpretations of SSA-I. Briefly, this is because of the way in which partitions are made in the continuous space at the local level, namely between points close in the space. It is essential therefore that the information on close (i.e. highly similar) points is accurately portrayed. Where there are few data points in the space or a subset of points is crucial to hypothesised structure, then the placement of the high similarities is important. This is shown in the context of the FBI data set, introduced in chapter 7. The significance of local monotonicity is also demonstrated in Empirical Study 7.7 for an established non-criminal actions example from Shepard (1974) using data from Levelt, van de Geer and Plomp (1966).

Differences in Stage 5 - Re-evaluation of Configuration (Repeat Stage 3)

SSA-I has a maximum iteration size of 100, and no control over the minimum improvement before stopping. In some programs such as ALSCAL (Takane, Young and deLeeuw, 1977) it is possible to specify the number of iterations before termination, the minimum change in fit (SSTRESS for ALSCAL) before termination or the minimum acceptable SSTRESS value.

Summary of Chapter 5

MDS procedures were distinguished in terms of model, transformation, data input and algorithm. The historical development of MDS in these phases was summarised. Particular attention was paid to the procedure mostly used in this thesis, SSA-I. Different implementations of SSA-I were compared to the Liverpool implementation, and no major differences were found that would threaten the validity of structural hypothesis testing or values of alienation to two decimal places. The stages of SSA-I were summarised and compared to other MDS methods. The implications of some of these differences were highlighted.

Chapter 6
error in representation and its impact on structural hypotheses

The Parameters of MDS and their Impact on Badness of Fit

One of the central principles running through MDS is the 'correctness' of the representation. This is usually gauged in terms of the badness of fit and the best dimensionality for the representation given the fit in higher and lower dimensionalities. In one respect, fit is purely an atheoretical concern, varying within parameters such as dimensionality and the treatment of tied values. However, it is argued in this chapter that in domains such as that of criminal actions, fit *per se* as the criterion of acceptability not as important as other considerations in the search for structure in geometric representations. An analogy with statistical hypothesis testing reveals that conclusions about structure made on the basis of representations are dependent on the potential errors in the conclusions due to the conditional nature of the structural hypothesis. Firstly, though, the influence of different parameters on fit is assessed.

If there are identical values in the similarity or association matrix then the treatment of these tied values has an impact on non-metric scaling. Kruskal (1964a) suggested that there were two ways to deal with ties: the primary and secondary approaches, as explored in chapter 5. The primary approach will break ties if this will help improve goodness of fit, unlike the secondary approach which preserves them at the expense of fit. Stenson and Knoll (1969) showed that the secondary method produced higher stress in MDSCAL than the primary method. However, the magnitude of the difference was low and related to the amount of tied values. Stenson and Knoll placed the associations between 30 variables into 10 or 50 discrete categories, corresponding to 'rough' and 'fine' coarseness with 10% and 2% tying respectively. With the secondary method of preserving ties, the difference between rough and fine stress was at the third decimal place; with the primary method of breaking ties, the difference between stress with rough and fine coarseness was never greater than 0.01. The difference between stress with secondary and primary approaches with fine coarseness was extremely small, and hardly difference with untied (i.e. extremely fine) data.

The conclusion from Stenson and Knoll (1969) - and confirmed by Lingoes and Roskam (1973) - is as follows. The coarseness of the grouping, namely the amount of tied values, does not have a great impact on stress unless the data are extremely coarse (i.e. 10 or less discrete values). In these cases, the choice of tying approach becomes important. Where this applies, under the Smallest Space Principle (Shye and Elizur, 1994) Kruskal's primary approach should be used since 'permitting untying of ties enables a smaller space to be attained than otherwise' (Guttman, 1968, p. 477). However, there may be cases where the preservation of ties is important, such as in Social Network Analysis (e.g. Scott, 1991) where the association between members of a criminal gang may be the number of phone calls made or observed associations.

The choice of local monotonicity will lead to greater alienation or stress in a plot, since it acts as an additional restriction on the algorithm and thereby diminishes degrees of freedom. In local monotonicity, the accent is on preserving the order of similar values in the association matrix, meaning that short distances tend to be more accurate. As Lingoes and Roskam put it: local monotonicity 'weights errors inversely to the size of the distances, i.e., errors in smaller distances will count more than errors in larger distances' (Lingoes and Roskam, 1973, p. 89). Global monotonicity by contrast applies no weight to the errors, meaning that the largest departure from the distance/rank-image equation is taken first, irrespective of the size of the distance.

The tying option potentially gains greater significance, when combined with local rather than global monotonicity, as explored in chapter 7.

The Necessity of Error in Scaling Real-World Information

Alienation in SSA-I (or the equivalent for error in other MDS programs) plays the role of a function to be minimised - its strict meaning is therefore be computational. Given the influence of parameters outlined above, it may be minimised down to zero in three ways. Firstly, a perfect representation for a set of data can always be found in $n - 1$ dimensions, where n is the number of variables, (Lingoes and Borg, 1987, p. 59) - even where data are random. Perfect solutions may secondly be found in fewer dimensions where variables or stimuli are perfectly positively correlated and therefore occupy identical parts of the space. This is the same as in principal components analysis when the addition of extra information that correlates perfectly with existing information does not increase the number of principle components (Kline, 1994). Zero alienation solutions may thirdly be found in fewer dimensions where all the pairs of similarities in the input matrix obey some regular uniform pattern.

When real-world data are scaled in lower dimensionalities than n - 1 some degree of error is inevitable. Therefore if the minimisation of alienation was the only goal of MDS then logically this would imply that the only acceptable solutions would be of high dimensionality. But the addition of more dimensions in MDS just to achieve zero alienation goes against the principle of parsimony. In the context of MDS parsimony would suggest that the simplest possible structure should be sought in the lowest acceptable dimensionality. This has been termed by Shye and Elizur (1994) 'the Smallest Space Principle'. A variant of this is the eponymous 'Shepard's law' (e.g. Coxon, 1982), which suggests that two dimensions should normally be adequate. In this respect, low dimensionality should be preferred on substantive and interpretative grounds, since higher dimensionalities are less interpretable. Therefore the Smallest Space Principle and the inherent possibility of unreliability mean that the minimisation of error in the configuration without regard for usefulness of the solution is undesirable. In other words, alienation is not purely a computational function to be minimised and to determine the acceptability of a solution.

The same argument that alienation has substantive implications also applies to the issue of dimensionality of the solution. As Borg and Lingoes put it 'If there are no *a priori* reasons for choosing a particular dimensionality, an SSA space of the lowest possible dimensionality is preferred.' (Borg and Lingoes, 1987, p. 59) Where structural hypotheses are proposed, there is an *a priori* reason for choosing a minimum dimensionality, giving a lower bound to the dimensionality problem.

The issue then becomes how much error can be tolerated in a configuration without it losing power and impact. The two aspects of the derived configuration - the substantive and the computational - are thus in opposition: the desire for a readily interpretable solution (i.e. low dimensionality) against the desire for a statistically accurate solution (i.e. low alienation).

Given that with real-world sources it is necessary to have some amount of error in the plot, the question arises as to what is an acceptable amount of error. But as Amar and Toledano point out asking for such guidelines on a criterion of 'good' fit is tantamount to asking: "Precisely how inexact are we allowed to be?" (Amar and Toledano, n.d., p. 146). They stated further that:

> If one is willing to tolerate inexactness in fit, one should also be willing to tolerate inexactness in the size of the coefficient which should be regarded as acceptable. (Amar and Toledano, n.d., p. 146)

In other words, by accepting anything other than perfect fit then quibbling over the criterion of fit is unjustifiable. Nevertheless, the original suggestions the acceptability of fit values made by Kruskal (1964a) remain in popular usage (Shye and Elizur, 1994). Kruskal suggested that the *st*andardised *re*sidual *s*um of *s*quares (i.e. 'stress') could be expressed as a percentage through multiplication by 100. If the result was below 5%, then the configuration was 'good'; at 20% the fit was 'poor'.

The inappropriateness of using these percentages as blanket guidelines was demonstrated in numerous 'Monte Carlo' studies (e.g. Stenson and Knoll, 1969; Klahr, 1969; Wagenaar and Padmos, 1971; Spence, 1972). The Monte Carlo studies demonstrated the unreliability of heuristics because the fit of configurations was heavily dependent on number of dimensions, as shown above, as well as the number of variables, treatment of ties, error in the data and true data dimensionality. Thus the same alienation value may well be unsatisfactory for 15 items but would be highly satisfactory for 40 items (Shye and Elizur, 1994). This is because the number of paired comparison between associations and distances increases exponentially with variables, since there are $n\ (n - 1) / 2$ paired comparisons for n variables. (It must be noted that local monotonicity does not feature at all in the Monte Carlo studies cited above, but has an impact on structural hypothesis testing as shown in chapter 7.)

The greatest use for the Monte Carlo tests has been in the context of testing the hypothesis that empirical data are random on the basis of the goodness of fit in the configuration. The lack of statistical distribution to describe the variation in fit using different parameters means that the Monte Carlo studies allows the next-best thing to statistical inference in MDS (though Ramsay, 1982, has achieved this under certain restrictive assumptions). That is to say, these studies offer a comparison of empirical fit with expected fit if the data were truly random, the 'nullest hypothesis' (Cliff, 1973, p. 484). If fit for empirically-obtained data is close to or even greater than that of random data then it can be concluded that there is no systematic variation within the data.

Heuristics on Alienation and the 'Correct' Dimensionality

The most important parameter to be determined in MDS with or without reference to Monte Carlo random values is the 'correct' dimensionality. The trend has been noted of configurations with non-zero alienation improving in fit with increasing dimensionality (e.g. Kruskal, 1964a), with the exception of degenerate lower dimensional solutions (Lingoes, 1977b). These degenerate solutions would look like a clump of points very tightly bonded with perhaps some other points at a distance. Measuring dimensionality against fit gives the 'elbowing plot' (Kruskal, 1964a), since there tends to be a characteristic dip in the plot where the gradient of the slope changes markedly from a large negative to a small negative. This 'elbow' in the plot indicates that an increase in dimensionality would not result in a marked decrease in error of fit. The meaning of this is that the configuration does not 'need' the extra dimension, or that the lower dimensionality does not constrain the distances to any marked extent. Researchers presume that the 'elbow' dimensionality is therefore the smallest space representation that preserves the structure of the data. This elbowing is analogous to the scree plot in factor analysis, (Cattell, 1966) where the first few common factors in the correlation matrix typically explain most of the variance, with subsequent factors accounting for a diminishing proportion.

The alternative shape of the elbowing plot should be a smooth exponential decay, where successive dimensionalities allowed for diminishing improvement. However Kruskal (1964a) suggested that even random data, where the exponential decay would seem more accurate, also have a distinct elbow. But it is not unusual for even reliable substantive data to have an elbow that appears to be double or even treble jointed.

There are other methods gauge the acceptability of a dimensional solution without using elbowing. Wagenaar and Padmos (1971) suggested a way of using measurement error to predict the true dimensionality of a plot. First, calculate the fractional measurement error ('noise') required to produce the empirically-obtained stress value for a dimensionality *m*. Using this level of noise, calculate the expected stress from Monte Carlo random data in dimensionality $m + 1$. If this expected value is significantly less than the obtained value in $m + 1$, then reject dimensionality *m* and accept $m + 1$. Now repeat the test for $m + 1$ with $m + 2$, using the noise level needed for the empirically-obtained stress in $m + 1$. In this way, the 'noise' in successive solutions could be estimated and when it dipped sufficiently with no appreciable gain in higher dimensionalities then the true dimensionality would be revealed. However, this is still related to the 'elbowing' phenomenon, except that the elbow has been moved from dip in stress to dip in noise.

The Wagenaar and Padmos (1971) approach was further explored and computerised by Spence and Graef (1974), using Monte Carlo data created by Spence (1970). Spence and Graef (1974) claimed this was an improvement on the method of data generation, and expanded the number of variables used beyond the 12 or less used by Wagenaar and Padmos (1971). The program developed, named *M*-SPACE, would generate the expected stress values in 1 to 5 dimensions for data that had a known true dimensionality. These expected values would then be compared to the empirically-obtained values of unknown dimensionality. The closest fit between the two sets of stress values, as measured by least-squares, would be taken as indicating the 'true' dimensionality. *M*-SPACE was later modified using the findings of Spence (1979) to include an analytical expression for calculating random stress values, rather than relying on tables.

The true dimensionality also 'reveals' itself in MDS solutions in other ways without the elbowing plot, but still following a similar principle. Like many MDS programs, SSA-I rotates final configurations to principle components such that the first axis explains most variance, the second explains the second most, etc. The true dimensionality would manifest itself in such programs when the m + 1th dimension had very little variance explained when compared to the *m*th dimension. In geometric terms, this would be a solution where the extra dimension was not used by the configuration, leaving the space 'empty'. However, pointed out in chapter 3, this space is not so much 'empty' as representing parts of the content universe that were not examined by that particular sample of variables used in the analysis. In fact, elaborated replications of existing analyses would be hypothesised to contain structure in these higher dimensionalities.

The 'empty space' criterion has also been suggested as a guide to deciding the 'correct dimensionality' in MDS programs that do not rotate to principle components. This includes FSSA (Shye, 1991), where the user is advised to switch to the preceding dimensionality if 'at least one axis is not fully used by the mapping' (Shye and Elizur, 1994, p. 125). Such a practice is less reliable than in programs that do rotate to principle components since the non-rotational programs have no orientation in space, while rotational programs have orientation suggested - but not dictated - by the axes of the first principle components.

In the Shepard diagrams of successive dimensions, the fit to the line improves since stress is a measure of fit to the disparity line. This is because there are more degrees of freedom in which to reproduce the associations, and the error between distances and disparities diminishes. If the Shepard diagram is irregular in some way, such as being too 'steppy' so that it looks like a staircase side-on, then this suggests that the plot should be rejected and a higher dimensionality inspected to see if the same occurs in the new Shepard diagram.

The choice of a particular dimensionality based on comparison with other dimensionalities with the same data may obtain the best relative dimensionality. But the most suitable dimensionality in relative terms may still not be adequate in absolute terms. This is especially important when using SSA-I and scaling up from low to high dimensionality, such as 1 through to 5 as would be required by *M*-SPACE. Local minima causing error in lower dimensionalities are 'slurred' up into the higher dimensionalities, causing them to be suboptimal when compared to solutions that would start and finish in the higher dimensionalities (Lee and Canter, submitted). Furthermore, the relatively best dimensionality may still not lead to optimal metric recovery of information if it is not the 'true' dimensionality (Spence and Graef, 1974, p. 337).

However, there are two limitations of the use of the nullest hypothesis in Monte Carlo runs and the heuristics on the adequacy of a representation. Firstly, the conditional nature of the hypothesis-testing procedure must be fully considered, and secondly, the substantive contribution to fit must not be neglected, both of which are not fully covered by a simple comparison with the nullest hypothesis.

Testing Structural Hypotheses with MDS Representations

Coombs *et al.* (1970) distinguished between two practices of scaling as a technique and scaling as a criterion. When used purely as a technique, scaling simply attempts to find the best fit between model and data. This is similar to how multiple regression equations seek to maximise variance like a function by adding and deleting predictor variables even if the items were worthless. But when scaling is used as a criterion then it tests the validity of the model as represented by the data. In other words, scaling as a technique is insensitive to departure of the representation from hypothesised structure.

To achieve the more desirable scaling as a criterion, it must be possible to reject structural hypotheses on the basis of representations - essential for the faceted approach in domains such as criminal actions. In chapter 4 the faceted approach to analysing the structure of criminal actions used the geometric representation from SSA-I to reveal that structure, provided there was Contiguity. Even though the information source for criminal actions were biased, it was suggested that the non-metric nature of SSA-I and the search for regions would reduce the impact of unreliability and recover structure. One problem so far unresolved concerns how structural hypotheses may be rejected or accepted. It is possible that a particular configuration was obtained that supported regional hypotheses that were in fact a result of an inadequate representation, or of data that were too unreliable to allow conclusions to be drawn. Alternatively, structural hypotheses could be hidden by representations with inappropriate parameters, such as too few dimensions. What must also be considered therefore is the nature of the representation and the possibility of extrinsic factors in the representation. This chapter investigates the way in which substantive concerns interact with statistical and computational concerns, and how structural hypothesis testing in the faceted approach can reduce the chance of error not just by seeking the 'true' dimensionality but with broader considerations.

There are some similarities and differences between structural hypothesis testing with the faceted approach and traditional statistical hypothesis-testing using probability significance values. Probability values with statistical tests are intended to guide the researcher, though objections have been raised objected to on the grounds of their questionable role in the scientific process (Guttman, 1977; Schmidt, 1996; also chapter 4) and a misunderstanding - even in introductory psychology texts - on their exact meaning (Dracup, 1995). The question over meaning is the myth that the probability value is the 'chance that the effect happened at random', which Booth (1994) forcefully stated was highly misleading.

According to Fisher's original work, 'statistical significance' actually meant one of two things: that the sample taken from the population was extremely rare or that the theory of random distributions did not hold for this sample. What it did *not* mean was the independent variable 'caused' the change in the dependent variable, but if all other external influences were controlled such as respondent fatigue then the researcher would be justified in concluding there was an experimental effect (Macdonald, 1997). Macdonald went on to say that 'All statistical tests do is to provide a researcher with an answer to the sceptical challenge that some particular effect could have resulted from sampling variation.' (Macdonald, 1997, p. 334) It must be remembered that for criminal actions, information bias begins with the sample. which is self-selected by offenders then themselves.

The most important similarity between statistical and structural hypothesis testing however is the conditional nature of the test. The acceptance or rejection of the null hypothesis using probability depends entirely upon the true state of the data, as theorised by Neyman and Pearson as an alternative to the null hypothesis (Macdonald, 1997). This leads to the correct understanding of the meaning of the probability value, with Dracup stating emphatically that:

> *The correct interpretation is that if the null hypothesis were true, then the probability that our experiment would produce a significant result (and the null hypothesis be rejected as a consequence) would be equal to the significance level at which the test was conducted* (Dracup, 1995, p. 359, emphasis in original)

In other words, it is the probability that the test statistic will lead to the rejection of the null hypothesis when in fact the null hypothesis is true. Table 6.1 summarises the interplay between the reality of the population and the conclusion drawn from the sample in statistical hypothesis testing.

		Reality of Population	
		Null Hypothesis True	**Null Hypothesis False**
Decision from Sample	**Accept Null Hypothesis**	True Negative	False Negative (Type II Error)
	Reject Null Hypothesis	False Positive (Type I Error)	True Positive (Power)

Table 6.1. Interaction between reality of population and decision based on sample in statistical hypothesis testing

For the False Positive, a Bayesian analysis of statistical hypothesis testing by Dracup (1995) showed that the probability values was dependent on the α level, the power of the test to detect a real experimental effect and the prior probability that the null hypothesis is true. In statistical significance testing, the first can be deliberately set by the researcher, the second can be estimated (e.g. Cohen, 1988), but the third can only be guessed. (For structural hypotheses the first cannot be used, the second can be improved but not perfected, and the third relies on a sound knowledge of previous results - where they exist.)

Similarly, test power is important in the applied setting where research is typically done for a reason and with an intention to change - which is fundamental to action research - rather than for its own sake (Ackroyd and Hughes, 1992). Factors influencing the power of a test to detect a True Positive include the test used, the effect size, the α level, the nature of the design, the direction of the hypothesis, the reliability of the measures and the sample size (Clark-Carter, 1997). Power was introduced by the Neyman-Pearson revision to the Fisherian experimental design, though the continuing ignorance of power in research was noted by Cohen (1988).

The most important difference however is that SSA-I and other non-maximum likelihood models do not have significance values associated with their plots. One way which has been widely used to gauge the acceptability of MDS solutions without the p values is through an evaluation of the fit as measured by alienation in an empirically-derived representation to reduce the chances of making an incorrect inference.

By analogy with statistical hypothesis testing, the errors inherent in doing this are as follows. A configuration must not be accepted if the regional hypotheses are artefactual of the data or the representation. If this incorrect judgement were made, it would be a type I error, a false positive. Similarly, a type II error, a false negative, would be the incorrect judgement that there is no regional interpretation when in fact this was due to the inadequacies of the representation. The conditional nature of the null hypothesis is paramount. Table 6.2 illustrates this.

		Structure	
		Structural Hypothesis False	Structural Hypothesis True
Representation	No Regional Hypotheses	True Negative	False Negative (Type II Error)
	Regional Hypotheses	False Positive (Type I Error)	True Positive (Power)

Table 6.2. Interaction between reality of population and decision based on sample in structural hypothesis testing

For a type I error, unmeaningful and random data would have to give a lower fit than other similar data in the Monte Carlo studies (assuming the Monte Carlo studies found global minima.) This situation is feasible, although it is improbable. In fact, the chances of getting this error diminish as the number of variables increases, since it becomes statistically more unlikely that larger number of variables could by chance correlate in this misleading way. The nullest hypothesis states the fit associated with a sample of correlations generally drawn at random from a normal distribution of known mean and standard deviation.

One way to diminish deliberately the type I error - rather than rely purely on decreasing probability with increasing variables - is offered by the faceted approach. This is simply to ensure that the selection criteria for items be structurally driven, preferably on the basis of Mapping Sentences where possible. This would emphasise that item selection was optimised, since facets or the elements of facets would be fully formed such that the facets collectively exhaust the domain, the elements are mutually exclusive, etc. However, if the observations were made from a partial content universe then this task is more difficult, though it can still be improved by being as fully formed as possible. Empirical Study 9.3 examines how strong structural hypotheses can be used to demonstrate that results are not meaningful due to poor design.

The consequent improvement in finding the true state of nature in the data arises from the *a priori* rejection of items that would be unrelated conceptually with the mass of other items. Given Contiguity between definition and data, such irrelevant items would increase badness of fit disproportionately, even with non-metric transformations. Identifying such items is easier where proven scientific laws exist in the behavioural sciences, such as Guttman's first law of attitude or intelligence (Gratch, 1973; cited by Shye 1978a). Without Contiguity, these items should be conspicuous by their unmeaningful contribution to the regional structure. Irrelevant items may also identified *a posteriori*, as in the practice of item analysis, by examination of the correlation matrix to examine which items were indeed highly negatively correlated with the others on the assumption of Contiguity.

For a type II error to occur, the solution would be rejected as unacceptable even though it was meaningful. This could be due to poor alienation on the basis of comparison with Monte Carlo values. However, the error could also occur where the representation was inadequate to model the structural hypotheses. The adequacy of representation - the key issue in type II error - is more of a substantive than statistical issue, and the faceted approach resolves this with structural hypotheses to guide the regional interpretation.

A minimum dimensionality for the representation may be required to test for the existence of structures in the faceted approach. Levy (1985) stated that 'Hypotheses as to the dimensionality of the SSA space are related to the types of roles of the facets.' (Levy, 1985, p. 74). Combinations facets also have minimal dimensionalities, such as a multiplex containing n axial facets requires n dimensions meaning for example a two axial faceted duplex requires at least two dimensions (Levy, 1985). Consequently, the discovery of an n dimensional faceted structure in an m dimensional space where $n < m$ implies that either the configuration has too many dimensions, other facets have not been identified in the domain or the sample is noisy. The latter condition means that higher dimensional solutions may need to be sought for lower dimensional structures with unreliable data.

Non-faceted interpretations of geometric solutions still also require minimum dimensionality. For example, the reanalysis of the Ekman data on perception of similarity of colours by Shepard (1978) required that at least two dimensions were necessary to display the circular order. The spatial configuration of data points found by Shepard was an ordered circle or circumplex. Shepard suggested that a one-dimensional solution, which presented the colours as a linear progression from high to low wavelength, was adequate in terms of the physics of the situation, but not psychophysics of perceptual judgement. In fact, Ekman's factor analysis of the data was postulated to have 5 factors explaining the variance, a clear demonstration of the unsuitability of that method for these data given Shepard's explanation of 99% of the variance with 2 geometric dimensions. Chapter 10 returns to this issue.

Shepard also addressed the damaging practice of rejecting structures on the basis of fit, suggesting that

> many users [of MDS] tend to place undue emphasis on the numerical value of departure from monotonicity (stress) to the virtual exclusion of much more important considerations of the statistical stability and substantive interpretability of the obtained configuration. (Shepard, 1974, p. 385)

To these considerations can also be added the issue of replicability of the structure in the plot in support of other known structures (Guttman, 1977).

Shepard (1974) also stated that reliance on Monte Carlo studies could lead to overinclusion of dimensions just as the Kaiser-Guttman criterion in factor analysis may be overinclusive of factors (Kline, 1994; see chapter 10). This objection to Monte Carlo runs was overcome by improvements in later studies, though. The third objection of Shepard (1974) was that configurations hypothesised to be one-dimensional could not be found in two-dimensional or three-dimensional representations. Empirical Study 7.7 presents such a representational problem.

The interpretability of plots is more important than judging the acceptability of the solution by comparison with Monte Carlo fit - something argued in the facet theoretical perspective. Clearly then substantive considerations can outweigh statistical ones. The faceted approach takes this one stage further by using substantive structural hypotheses as clues to the acceptability of the plot.

As suggested earlier, type II error may also be found in a poor representation of the plot rather than in poor sampling of the content universe. Nowhere is there a poorer sample method from the content universe than when the rationale for sampling was different to that of a researcher; namely, using a partial content universe when looking at existing observations of a phenomenon such as police crime files. This issue of adequacy of representation is therefore paramount.

The chance of a type II error may be increased where information is particularly noisy or conducted from a partial content universe, as with the criminal actions data. Therefore scaling must be done with a consideration of how the representation will model the data so as to increase power.

The Strength of Structural Hypothesis Testing

In conclusion, the experimental hypothesis (H_1) for the faceted approach states that the hypothesis of structure in the domain is correct and the SSA-I representation can be partitioned into distinct sub-universes or regions. The null hypothesis (H_0) states that the structural hypothesis has not adequately captured and conceptualised the content universe, so its hypothesised regions will not be found in the SSA-I representation. To ensure that the correct inference about structure is drawn from any given representation, adequate parameter . must be provided to model the solution, and the creation of items must be theory driven. This would prevent a false negative, that the structure was not modelled accurately by the representation when in fact it was, or a false positive, that structure was found when this was artefactual.

The criticism is sometimes made that the finding of structure in the geometric space is a certainty, since the observations were designed to do that anyway. For example, Roskam suggested that the if Facet Theory is a 'method for generating definitions for the elements of a universe of observations [then] it can only predict the confirmation of its definitions, which is trivial.' (Roskam, 1979, p. 243) But this view is ill-conceived since Facet Theory doesn't so much as *predict* the confirmation of its definitions but *allow* the confirmation or indeed disconfirmation of its definitions, namely its structural hypotheses from definitional systems. It does however allow the definitions to be logically well-formed and consistent with existing knowledge. In addition, it allows those definitions to be used for future research, moving towards the achievement of cumulative science.

Moreover, as Brown pointed out, 'this is a strength rather than a weakness of the faceted approach' (Brown, 1985, p. 53). There is no guarantee that just because the design was faceted then the results will also be faceted with regional interpretation. The use of secondary data to confirm faceted structural hypotheses - such as Guttman's original re-analysis of Thurstone's data (Guttman, *1964*) - demonstrates this. Furthermore, if Roskam's criticism were applied to Fisherian experimental design then it would deny the validity of an experiment that was set up to 'prove' a difference in the behaviour rats injected with a psychoactive drug and rats injected with saline. With an experimental hypothesis 'behaviour is affected by this drug', how else could such a hypothesis be tested?

The incorporation of substantive elements of the content universe into the statistical representation was something repeatedly emphasised by Guttman. By incorporating issues of meaning into the interpretation of MDS plots, the practice becomes scaling as a criterion, which is 'a method of testing the descriptive validity of some measurement model ... whether the data can be fitted by the model' Coombs *et al.* (1970, p. 32). This of course contrasts with scaling as a technique, where error is minimised without regard to some hypothesised model (Coombs, *et al.*, 1970). Facet Theory was developed specifically with the style of measurement intended by scaling as a criterion. This is done with the parameters and limitations in mind.

In this way, the danger is minimised for MDS techniques becoming what Coxon and Jones (1980, p. 32) termed a 'garbage-processing' method which is used when stronger multivariate models are too costly in Coombsian assumption terms (from Coombs, 1964). That is to say, the weaker measurement models which can deal with domains such as criminal actions can nevertheless create powerful theories.

Summary of Chapter 6

Parameters such as dimensionality, treatment of ties and locality of monotonicity were examined for their impact on alienation. The need for low dimensional solutions is traded off against increasing notable alienation. However, the difficulty of exact criteria for the trade-off was noted. Improvements were made by Monte Carlo random data and other heuristics for dimensionality and fit. These were still deemed unsatisfactory for the faceted approach to structural hypothesis testing, though an analogy with traditional statistical hypothesis testing with probability values may overcome this. The conditional nature of both statistical and structural hypotheses was emphasised, and the possibility of incorrect inference with False Positives or False Negatives. Ways to minimise this potential for error in the faceted approach were discussed. It is suggested that conditional structural hypothesis testing in the faceted approach allows scaling as a criterion rather than a technique, as championed in Coombs *et al.* (1970). The critique of Roskam (1979) was addressed.

Chapter 7
local monotonicity and localised spatial bonding in representations

Local Monotonicity as a Computational and Substantive Issue

This chapter introduces a new substantive data set, the FBI data set. A standard multidimensional scaling analysis is performed on these data to test a valid structural hypothesis. The results are shown to be disappointing, but a thorough examination of the results shows that the representation partially concealed the true structure. In the previous chapter, local monotonicity was introduced as a computational parameter for geometric representations in MDS. This chapter examines how local monotonicity can influence the representation of data and consequently structural hypotheses, using the FBI data and later a published source.

The FBI Data Set

The *Crime Classification Manual* (CCM; Ressler, Douglas, Burgess and Burgess, 1992) was a formalised attempt by the FBI to 'develop a crime classification system is based on completed homicide cases' (Douglas and Burgess, 1989, p. 13). The style of classification system in the CCM was suggested by Ressler *et al.* to be based along the lines of the psychiatric classification of syndromes in DSM-III-R (APA, 1987). Just as with the psychiatric classification, the classification of crimes was based on personal experience of investigating the crime as well as shared knowledge in the FBI community.

Three crimes categories of homicide, rape and arson were examined broadly each classified by Ressler *et al.* (1992) into broad groups. For homicide, the four general categories were: criminal enterprise, personal cause, group cause and sexual homicide. Within each grouping, the individual classifications were given a code, as with DSM.

Each of the different classification types was differentiated on a number of different features, and followed by a real-life example of the crime classification as an illustration. The homicide crime types were classified so that each was based on 'questions about the victim, the crime scene, and the nature of the victim-offender exchange', the 'defining characteristics of crime' (Ressler *et al.*, 1992, p.7).

The types were distinguished on the following sorts of aspects:

1. Victimology: the offender's knowledge of the victim; the victim's elevated risk of being targeted; the risks taken by the offender in choosing this victim.
2. Crime Scene Indicators: the number of crime scenes; the environment, place and time of the offence; the length of time spent at the scene; the number of offenders; the spontaneity or organisation of the crime scene; the body being dumped or concealed; the theft of items from the scene.
3. Staging: whether someone deliberately altered the crime scene in any way prior to the arrival of the police.
4. Forensic findings: the cause of death being gunshots, trauma, strangulation, etc.; overkill; mutilation of the body; evidence of sexual assault.

Ressler *et al.* (1992) differentiated 32 distinct types of homicide, which are listed in Table 7.1. It was intended by the FBI that on being presented with a homicide crime scene, the investigator would consider the various features of the scene and then check which of the classifications best fitted the crime scene. The classification in the CCM would then give clues about the characteristics of the offender - instant offender profile - and investigative suggestions such as items to look during a search. However, it is argued that the actual published FBI profiling process and the fundamental structure of the homicide classifications imply that the method is not as systematic and rigorous as would be expected.

In terms of the Coombsian Research Model from chapter 1, all the decisions about the creation of the classification were made by many people, and requirements for inclusion as a crime type included the crime occurring, being detected, being investigated and solved, and being sufficiently distinct from other crimes. As with all classifications based on incidence, self-selection and attrition, the domain is not exhaustively covered and is biased.

More importantly, the CCM is functions as a 'slimmed-down' FBI profiling process. The widely accepted and referenced statement about the FBI profiling model is that of Douglas, Ressler, Burgess and Hartman (1986), which was neatly summarised

Homicide Group	Code and Description	
Criminal enterprise homicides	101	Contract (third party) killing
	102	Gang-motivated murder
	103	Criminal competition homicide
	104	Kidnap murder
	105	Product tampering homicide
	106	Drug murder
	107.01	Individual profit murder
	107.02	Commercial profit murder
	108.01	Indiscriminate felony murder
	108.02	Situational felony murder
Personal cause homicides	121	Erotomania-motivated killing
	122.01	Spontaneous domestic homicide
	122.02	Staged domestic homicide
	123.01	Argument murder
	123.02	Conflict murder
	124	Authority killing
	125	Revenge killing
	126	Nonspecific-motive killing
	127.01	Political extremist homicide
	127.02	Religious extremist homicide
	127.03	Socioeconomic extremist homicide
	128.01	Mercy homicide
	128.02	Hero homicide
Sexual homicides	131	Organised sexual homicide
	132	Disorganised sexual homicide
	133	Mixed sexual homicide
	134	Sadistic murder
Group cause homicides	141	Cult murder
	142	Extremist murder
	142.01	Paramilitary extremist murder
	142.02	Hostage extremist murder
	143	Group excitement murder

Table 7.1. Details of homicide types used in FBI Crime Classification Manual (Ressler *et al.*, 1992)

as being simply the formula 'WHAT plus WHY equals WHO' (Pinizzotto and Finkel, 1990, p. 216). In other words, gauge the motive or the 'WHY', and the initial supposition about the crime classification of offence and offender at the 'WHAT', leading to the 'WHO'. However, the deciphering of 'which came first: the motive or the classification' is not made clear in FBI texts with the two processes presented as so intertwined as to be inseparable. The proposed motive is gained from an initial assessment of the crime scene, and it is then used to re-interpret the more detailed features of the crime scene.

Similarly, Ressler. Burgess, Hartman, Douglas and McCormack (1986b) implied that to identify an offender it is necessary to understand motivation. FBI agents 'inferred a motivational framework that included expectations, planning, and justifications for the criminal action as well as "hunches" regarding postcrime behaviors' (Ressler *et al.*, 1986b, p. 275). Ault and Reese stated that in these cases, the 'primary psychological evidence which the profiler is looking for is motive' (Ault and Reese, 1980, p. 38), with clues to motive found in evidence of the offender's planning or irrationality.

As Ressler *et al.* (1992) put it: 'Once the investigator has classified the offence (*and thus the motive*), the investigative considerations and search warrant suggestions can be used to give direction and assistance to the investigation' (Ressler *et al.*, 1992, p. 11, emphasis added). Where there are multiple motives, it was suggested to 'classify the offence according to the predominant motive.' (pp. 6-7) The crime scene evidence (WHAT) and motive (WHY) interact and reinforce each other in the 'bootstrapping process [which] is referred to as profiling' (p. 22).

Clearly, the consideration of motive is central to the classification process. It is therefore proposed that the classification of homicide types is not based around a wide range of 'defining characteristics', but instead that the differentiation of offender motives is central and crucial to the classification. In other words, motive in the CCM is proposed to be equally important as it has been in the profiling process outside the CCM.

Nevertheless, Douglas *et al.* noted that 'Motivation is a difficult factor to judge because it requires dealing with the inner thoughts and behavior of the offender' (Douglas *et al.*, 1986, p. 414). The assessment of motive would require 'going into the mind' of the murderer since the motive for the crime 'may all too often be one understood only by the perpetrator' (p. 403). For this reason, a non-metric multivariate approach to test the true structure of the biased Crime Classification Manual typology of homicide is particularly suitable. This was done in Empirical Study 7.1.

Empirical Study 7.1: FSSA of the FBI Data

It was hypothesised that the structure of the classification system was determined by the motive(s) ascribed to the offender. This structural hypothesis was backed up by an extensive examination of the FBI literature on the profiling process and classifications of crime types (Lee, in press). Under the expanded CRM of Coombs (1964) introduced in chapter 1, phases 0 through to 4 have already been decided and it seems that the material has been presented as a classification system.

No.	Item	Description
1	Victim Known	The victim was personally known to the offender.
2	Victim Risk	Objectively speaking, the victim was at an elevated risk of targeting, usually as a result of occupational hazards or behaviour.
3	Offender Risk	The offender took an exaggerated risk in performing the murder or in selecting a particular victim.
4	One Crime Scene	The homicide was committed at one and only one location.
5	No Time Post Mortem	The offender immediately left the crime scene after the victim was murdered.
6	Single Offender	Only one offender was involved in performing the murder.
7	Unplanned Crime Scene	No preparation was made in anticipation of the murder at the crime scene.
8	Body Plain View	The body was not concealed in any way by the offender.
9	Body Left at Scene	The body was left in the same location that the murder took place.
10	Items Stolen	The offender removed an item or items from the crime scene, including from the victims body.
11	Staging	An attempt was made to mislead the investigating authorities about what took place at the crime scene.
12	Overkill	An excessive level of force was used by the offender during the murder.
13	Mutilation	The body was mutilated in some way *post mortem*.
14	No Sexual Assault	No evidence of sexual activity before or after death was found at the scene.
15	No Interaction	No evidence was found suggesting that the offender interacted with the victim immediately prior to the murder.
16	Motive Commercial	The motive of the offender was to benefit financially in some way.
17	Motive Elimination	The victim was 'in the way' and was therefore eliminated as a means to an end for an ulterior gain. This includes murdering the victim symbolically in an offender's ideology.
18	Motive Revenge	The motive was to murder the victim as an end in itself due to the offender's perception of having been wronged by that individual.
19	Motive Sexual	The motive of the murder was to gain immediate sexual gratification through the death of the victim, not just to prevent the victim identifying the murderer.

Table 7.2. Details of items used to described FBI homicide types

However, since the descriptions of the crime types is included in the classification, it is possible to assess this qualitative material given in the CCM to create a multivariate assessment of the classification. Therefore each homicide crime type was coded on 19 variables, four of which were the motives explicitly mentioned in the CCM text. The details of the variables used to code the homicide types are given in Table 7.2.

A two-way two-mode matrix was created using the 32 murder types as cases and 19 variables. Each values was coded into in an ordinal trichotomy of 'usually absent', 'sometimes present' or 'usually present'. The wording of this trichotomy was chosen to reflect the imprecise language of each description of the homicide types in the CCM text, which used qualifiers such as 'usually', 'sometimes', 'often', 'rarely', etc.

The reliability of the items were tested by three independent raters using the same items in the CCM. The results of the reliability analysis using Reliability Lite (David Graper, University of Pennsylvania) are given in Table 7.3.

No.	Item	Krippendorff's α
1	Victim Known	0.640
2	Victim Risk	0.190
3	Offender Risk	0.044
4	One Crime Scene	0.361
5	No Time Post Mortem	0.033
6	Single Offender	0.745
7	Unplanned Crime Scene	0.521
8	Body Plain View	0.437
9	Body Left at Scene	0.430
10	Items Stolen	0.615
11	Staging	0.250
12	Overkill	0.525
13	Mutilation	0.480
14	No Sexual Assault	0.762
15	No Interaction	0.029
16	Motive Commercial	0.775
17	Motive Elimination	0.682
18	Motive Revenge	0.576
19	Motive Sexual	0.722

Table 7.3. Reliability of items coding the FBI homicide types

The average Krippendorff's α for these ordinal responses was 0.464, with a range of 0.03 to 0.78. These figures indicate the proportion of observed co-occurrences above chance (Krippendorff, 1980), so that on average 46.4% are above chance. This figure is remarkably good given the imprecise nature of the material, and it must be noted that the average for the 4 motives is 0.689. Of course it would have been possible to increase reliability by coding in dichotomies or by deleting those variables with low reliability, but this was felt to be unnecessary given the inherently qualitative nature of the data.

To measure the similarity of the items, a coefficient was required which exploited the ordinal trichotomous response range for the data used. Therefore the regression-free coefficient of (weak) monotonicity μ (e.g. Shye, 1985) was used. The dichotomous form of this is equivalent to Yule's Q (Shye, 1985, p. 72). μ can also be made stronger and equivalent to Pearson's r, the Product Moment Correlation Coefficient (Raveh, 1986, p. 122).

This correlation matrix was then analysed using FSSA to search for regional structure. The analysis used the default option of +2 weighting of local monotonicity, which was intended to 'favor the high similarities "somewhat" ' (Shye, personal communication). It is hypothesised that if the domain of murder types classified in the CCM were determined by motives, then the 15 non-motive variables from Table 7.2 should form meaningful regions around the 4 motive variables.

The details of the FSSA plot are represented in Figure 7.1. This shows that the four motive items do not form such the extremes of a hypothetical cross, as would be expected using a dimensional interpretation of the motives as defining the space. Searching for regional structure which would consist of four regions each containing a motive and a set of conceptually similar items is also difficult. In particular, the highly emotional Revenge motive and the more calculated and instrumental Elimination motive seem particularly out of place. It is not possible to partition the gratuitous expressive violence of the 'Overkill' item in with the Revenge motive, where it would conceptually belong. Similarly, the 'No sexual interference' item would seem more appropriate for the non-emotional Elimination motive rather than the Revenge motive into which empirically is nearer. Furthermore, the Coefficient of Alienation for this solution was 0.291, above what many might regard as an acceptable level. Therefore, it would seem that the representation does not support the hypothesised structure of the FBI Crime Classification Manual as being defined by motive.

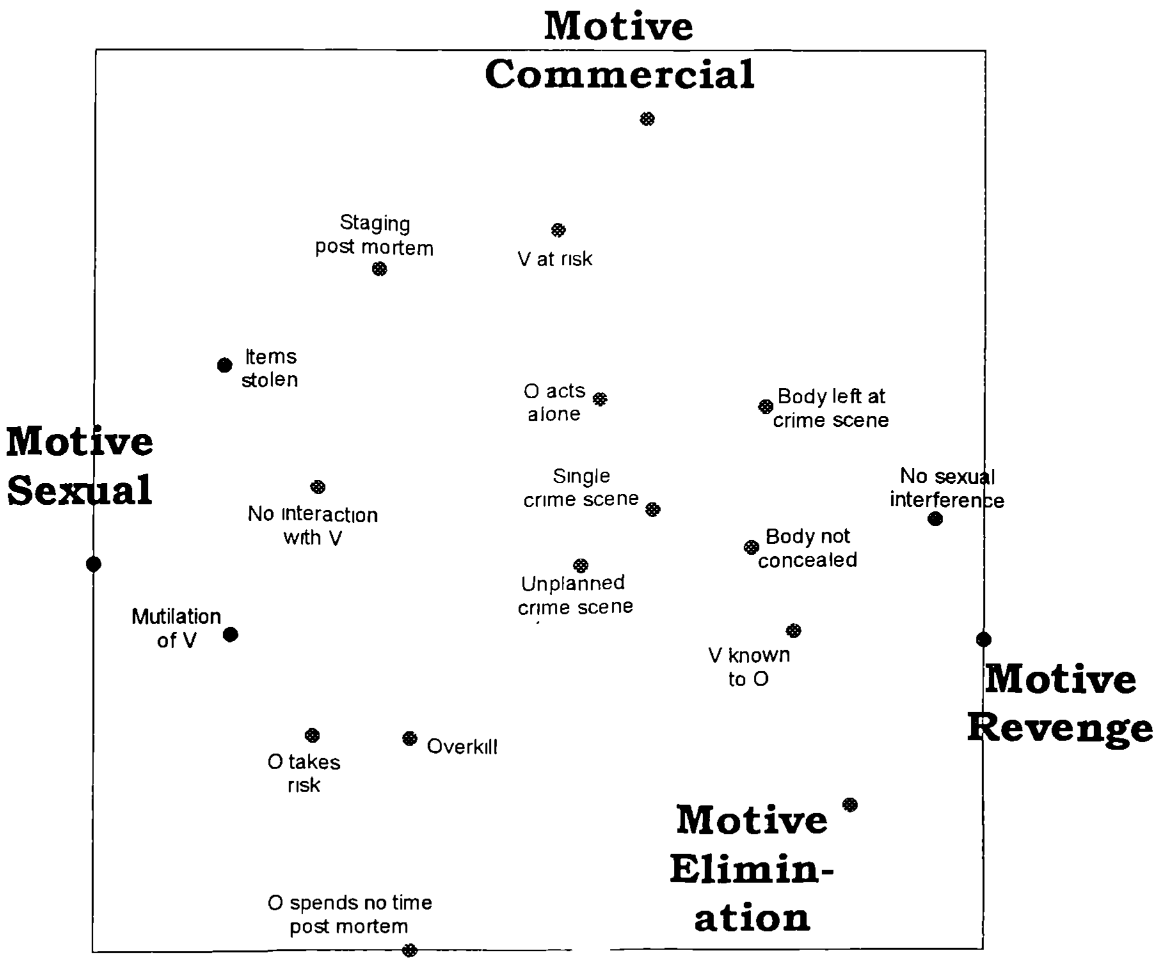


Figure 7.1. FSSA of 19 items describing FBI homicide types

Alternatively, this conclusion could be a False Negative or Type II error. In other words, the structural hypothesis about the motives determining the classification is true, but the FSSA representation conceals this and hence meaningful regional hypotheses cannot made. To postulate a reason for this error, it is necessary to develop a new concept termed Localised Spatial Bonding (LSB), a property missing in the above representation. The explanation of LSB requires first a thorough review of local monotonicity and how it influences MDS substantively, as opposed to computationally when introduced in chapters 5 and 6.

The Nature of and Need for Local Monotonicity

The representational implications of local monotonicity can be succinctly summarised using the illustration of projections of the earth. This three-dimensional spherical structure must be 'flattened' into a two-dimensional representation of a map. Maps ensures that smaller distances tend to be kept accurate whereas larger distances are more inaccurate, since the effect of the curvature of the earth is a function of the distance between two points on a map. In other words, larger distances - or the summation of many small ones - are less reliable than a short distance. In other words, error in local (short) distance information is diminished - which is how local monotonicity works. Therefore, local monotonicity may be more suitable in some cases since monotonicity mapping is required only around the 'neighbourhood' of each point.

Coxon (1982) suggested two instances where local rather than global monotonicity would be advantageous. Firstly, where data form a horseshoe or 'C' shape in a geometric solution of more than one dimension, the curvature might be associated with the data collection procedure and the number of discrete values for similarity being low (Kendall, cited by Coxon, 1982). This is not such a problem for association matrices derived from two-mode two-way matrices. The shape has been found in data where a time sequence or seriation characterises movement from one end of the horseshoe to the other (Kendall, 1971). The closeness of similar objects is essential in these cases, but there is a 'ceiling effect' whereby many comparisons are taken to be equally dissimilar, even though there could be further differentiation if there was no saturation. The use of local monotonicity is natural in these circumstances, since the lower similarities - therefore larger distances - are more unreliable and should be diminished in their contribution to the scaling solution. This is an interesting example of where systematic bias in data can be overcome by matching the measurement model to the unique characteristics of the observations and data derived from them.

A closely related effect may be found where the association matrix has been calculated from a two-way two-mode input matrix and there are many identical low values. This may occur where low overall frequencies of occurrence are found in dichotomously coded data and association coefficients are used such as Jaccard's, as often occurs with analysing criminal actions data (e.g. Canter and Heritage, 1990; Canter and Fritzon, 1998; Canter, Kirby and Hughes, in press). This particular coefficient has a range of values from 0 to 1 so any pair of items which are conceptually dissimilar are represented as a 0 - i.e. not at all similar, rather than dissimilar to some degree. An empirical instance is explored in detail in chapters 8 and 9.

The second instance of Coxon (1982) where local monotonicity has an advantage over global monotonicity was where high dimensional solutions are required to be reduced in dimensionality. In the interests of parsimony, however, and given the argument above for structural hypothesis testing, it would seem natural to prefer the smallest space automatically. This was demonstrated by Shepard (1974, p. 390) using data points placed on an object shaped like a steep parabolic dish. These points were transformed into two dimensions and appeared as an 'idealised radex' when local monotonicity was required. The important point to note is that higher dimensional structures may be found in lower dimensional representations provided that local monotonicity is used and understood in the interpretation of the representation. In other words, conclusions about the information held in larger distances was recognised as weaker than short distances. To summarise, Coxon (1982, p. 121) suggested that the criterion for global monotonicity that all data should fit distances may sometimes be more 'restrictive' than necessary - though it is less 'restrictive' in the computational sense since requiring local monotonicity is an additional parameter.

An additional third instance can be found in Borg and Lingoes (1987), who pointed out that the utility of local monotonicity in overcoming the degeneracy in geometric representations of strongly 'clustered' data. (It should be noted that 'Regions are in general not "clusters" that are discernible by "empty space" around them.' (Levy, 1985, p. 76), which is in accordance with the Shye and Elizur (1994) Continuity Principle.) 'Clusters' occur where there are distinct subsets of objects or variables with far greater between-subset variance than within-subset variance. The Shepard diagram of distance against associations of these degeneracies show a 'steppy' plot where there are only very large or very small distances irrespective of the association, even though overall alienation was good and close to zero. The geometric representation of these data would show the subsets clumped together at the edge of the plot. In this instance, the empirical fit would be extremely favourable with respect to random fit, but the substantive value of the structure would be low.

To overcome this, the use of local monotonicity was recommended by Borg and Lingoes (1987).

Another more important need for local monotonicity is when looking for regional structure in the faceted approach, requiring the achievement of localised spatial bonding (LSB).

Localised Spatial Bonding: Local Monotonicity in Regional Interpretation:

According to the Continuity Principle (Shye and Elizur, 1994), the space in the SSA-I plot is continuous. The classification of points into regions in the geometric space according to structural hypotheses is through partitioning. For an item to be a member of a region, it must add to the meaning of the region, integrated with all other points in that region. For criminal action, the behaviours are classified into themes suggestive of the intention. The boundary for the different meanings is the partition. Each region denotes a distinct content sub-universe, interrelated to the other regions by qualitative or quantitative order. This means that the continuous space is made discrete by the partitions in the regional interpretation. The importance of the partitioning for regional interpretation is paramount.

For any regional interpretation to be a valid gauge of Contiguity for the structural hypotheses displayed in the plot, it must be reliably represented so as to avoid type II error. Let us assume for the moment that the whole plot is reasonably *dense* - that is, with sufficient points spread evenly throughout. To partition into regions would require that the local sub-space be represented accurately so that if two adjacent points were placed into different regions then their relative positioning was accurate. An example of three regularly spaced points and a partition through them is given in Figure 7.2.

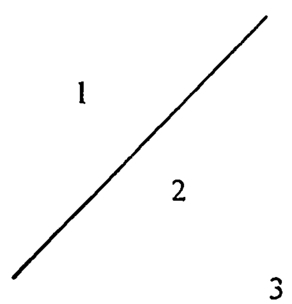


Figure 7.2. Hypothetical points and partition in SSA space

For such a partitioned plot, it is more important that the empirical similarity between points 1 and 2, s_{12}, be more accurately represented empirical similarity between points 1 and 3 , s_{13}. This is because the meaning of 1 and 2 is hypothesised to be conceptually distinct, so the accurate representation of points at the edge of partitions or 'outer points' (Zvulun, 1978) is essential. This is possible through the use of local monotonicity, which would emphasise distances d_{12} and d_{23} more than d_{13}, and therefore achieve the representational requirements. On a larger scale in a full plot of points, the distances between regions that are adjacent are more important than regions that are not adjacent and are separated by other regions.

This gives a substantive advantage to an otherwise computational parameter for the faceted approach; namely that high similarities are deliberately and accurately placed close together and regional interpretation is thus made more robust. This desire for good local monotonicity for robust regional hypotheses can be termed Localised Spatial Bonding (LSB). In a sub-space of LSB, if the items were constructed from a rational substantive perspective then these would be items of high conceptual similarity, where each adds to the overall meaning of the sub-space. In other words, a robust regional hypothesis of structure would have good LSB, the measurement of which is presented in Empirical Studies 7.4 and 7.6.

By contrast, global monotonicity applies no weight to the errors, so the largest error in the previous iteration is corrected first. This is done irrespective of the size of the empirical similarity and hence hypothesised conceptual similarity, given Contiguity. Equally, the hypothetical distal monotonicity posited by Lingoes and Roskam (1973) would favour low similarities and large distances at the expense of short ones in the opposite way to local monotonicity. This would be useful in a dimensional approach to interpreting MDS or SSA-I space, since such an approach tends to take the heuristic of examining the points furthest apart along the hypothesised dimension and suggesting what property they differentiate. The hypothetical property of *distal spatial bonding* would be useful for this approach.

LSB though requires local monotonicity so that within a small localised area, error in the monotonicity transformation is low and therefore it is accurately represented. LSB is therefore an issue of representation which has an impact on the ability to detect structure using the faceted approach and regional interpretation. One data set that requires strong LSB for its regional hypotheses to make theoretical sense is the FBI data set, which is essence is concerned with the correct representation of a subset of items in a geometric plot.

LSB would also help in the prediction of the value of existing items which may be missing from a particular case in a data set. This is because these items will be conceptually close and empirically correlated, and interpolation between these items needs to be accurate at the local level. LSB could also be used for hypothesising new items, since these are usually done in reference to what similar (local) items have in common rather than what dissimilar items do not have in common.

However, as Empirical Study 7.2 demonstrates, increasing the weighting of local monotonicity in the search for LSB causes higher alienation.

Empirical Study 7.2: Local Weighting of Monotonicity and Alienation

It was suggested in chapter 6 that local monotonicity was a parameter used to accentuate the correct representation of high similarity values in MDS at the expense of lower similarity values. Thus, for example, if there are two values which equally disobey the rank-image principle in SSA-I, then the higher similarity value is to be resolved in preference to the other. In fact, according to the weighting used then a small error with a highly similar value may be computationally more important than a large error with a highly dissimilar value. Because of this additional restriction, the use of local monotonicity would be expected to increase alienation.

To test this out, the effects of various weightings of local monotonicity on alienation were examined using FSSA (Shye, 1991) with a range of different weightings for local monotonicity. FSSA was chosen primarily since SSA-I only offers two values for monotonicity ('local' or 'global' i.e. none) while FSSA has a range of monotonicity values from +5 to -5 in steps of 1.

The substantive FBI data set was used to explore the fit under these varying conditions of locality. The results of the alienation figures are in Table 7.4.

Weighting for Local Monotonicity	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Coefficient of Alienation	185	199	215	229	244	260	276	291	305	303	309

Table 7.4. Alienation values for different local monotonicity weightings in FSSA with FBI data
Values for alienation are without decimal points

The trend for these figures can be seen more clearly in Figure 7.3.

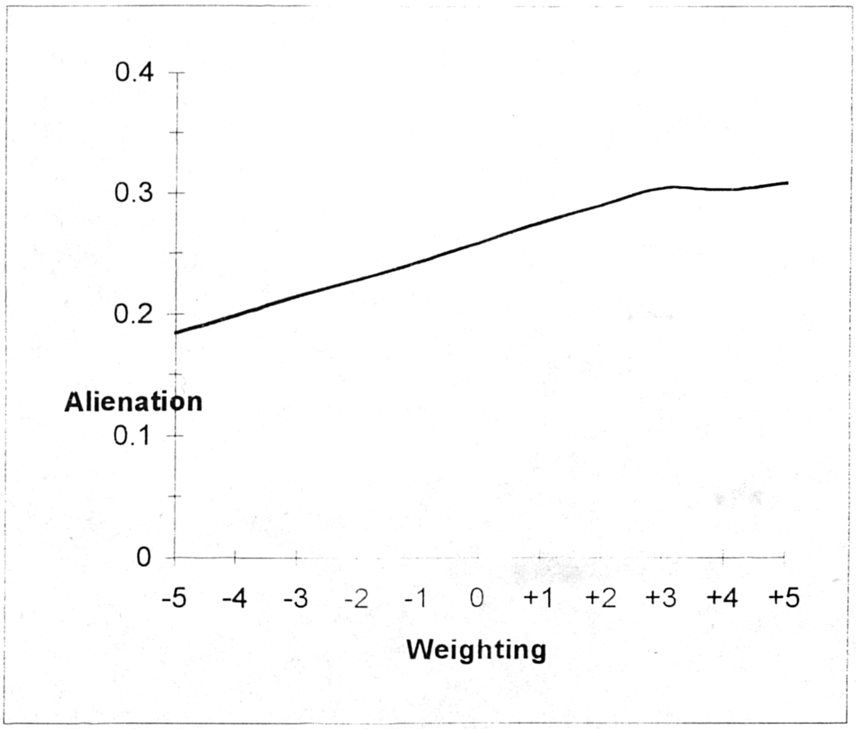


Figure 7.3. Graph of Alienation values for different local monotonicity weightings in FSSA with FBI data

Clearly, the overall trend is that as weighting of local monotonicity increases the Coefficient of Alienation also increases in a linear fashion, though with a slight blip at +4. This demonstrates that goodness of fit does indeed suffer as a result of stipulating that highly similar items are mapped more precisely. Computationally it is better to seek a solution that emphasises preserving the order of low or negative correlations and therefore large distances. Local monotonicity clearly diminishes the 'variance explained' in a solution by increasing the alienation. Local monotonicity is more statistically restrictive in terms of fit, since the weighting of smaller distances means additional constraints on finding an overall best solution.

A similar trend is realised when compared to SSA-I runs, using the options of global and local monotonicity. The differences in alienation are shown in Table 7.5.

Monotonicity	
Local	Global
260	257

Table 7.5. Alienation values for local and global monotonicity in SSA-I with FBI data
Values for Alienation are without decimal points

However, the difference in values here between local and global in SSA-I is actually less marked than to change of one unit of weighting in FSSA by one unit. This means that the choice of high local weighting in FSSA must be justified on some grounds, since the effect on the alienation is so marked.

It was suggested previously in Empirical Study 7.1 that with a weighting of +2 the alienation value of 0.291 was high and might have been seen as unacceptable; it is now suggested that the value may have to go even higher in order to obtain a better structure with good LSB. However, as the next Empirical Study 7.3 shows there is a strong substantive advantage to ignoring this computational aspect of this study.

Empirical Study 7.3: Locality of Monotonicity and Localised Spatial Bonding

The previous Empirical Study 7.2 examined only the alienation values for the FSSA solutions. Since the question is one of structure, it is perhaps more important to see how the structure may change according to the weighting.

The FBI data set was therefore analysed with FSSA using the μ coefficient - as in Empirical Study 7.1 - but using all weightings of local monotonicity from +5 to -5, including 0. The same (weak) μ correlation matrix was also put through SSA-I using local and global monotonicity conditions. The item configurations were then examined for broad consistencies in the patterns, given permissible reflections and rotations in the plot. It was found that generally speaking the configurations were highly similar, except in one crucial respect: the order of the four motive variables. The placement of the 15 non-motive variables was regionally invariant. Figure 7.4 summarises the two different orders.

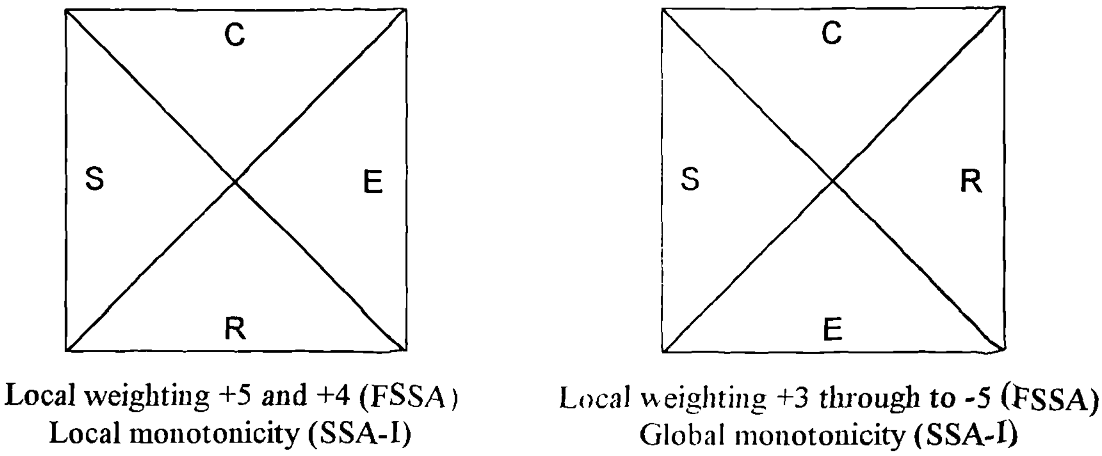


Figure 7.4. Schematic diagram of order of motives in FBI data

The order of the two motives 'swapped' between local weightings +4 and +3. There is no difference for the two configurations in terms of the type of facet to describe the regional since both could be partitioned according to a qualitative polar facet. The order of items in a circular circumplex configuration does not matter since they are quantitatively equal. However, since this is a regional interpretation then this swap has important implications for the robust partitioning of the plot into meaningfully distinct regions. As the Empirical Study 7.4 shows, a robust partitioning requires an understanding of how the local monotonicity creates LSB.

Empirical Study 7.4: LSB and the Representation of High Similarities

Local monotonicity in SSA-I or a weighting of local monotonicity in FSSA will attempt to create a good representation of high similarities. The FBI data set was taken to see if the creation of better LSB could be achieved through local monotonicity. It was hypothesised that LSB and hence more meaningful partitions would be improved by local monotonicity.

Local monotonicity emphasises good fit of short distances around a point, so more highly correlated variable pairs are put into their 'proper' place with local monotonicity. This is favourable for LSB since it requires that local subspaces preserve the true structure of the content universe in the non-metric representation.

The FSSA (Shye, 1991) association matrix for the FBI data using Guttman's coefficient of monotonicity, μ, was examined. The average correlation value for all similarities was 0.05, with a median of 0.10. It was found that for all the motives, the three highest associations between that motive variable and all other variables was positive, and usually high. Table 7.6 shows these items and their corresponding μ values.

Highest Value	Motive Type			
	Commercial	Elimination	Revenge	Sexual
1st	Victim Risk (57)	No Sexual (96)	No Sexual (=59)	Items Stolen (90)
2nd	Staging (53)	Body Plain View (66)	Single Offender (=59)	Mutilation (83)
3rd	Items Stolen (48)	Victim Known (34)	Victim Known (47)	Staging (58)

Table 7.6. Highest associations with motive variables in FBI data
Numbers in brackets are μ Coefficients without decimal points

The distances between the motive items and the three highest similarity values were physically drawn onto the FSSA solutions containing the highest local weighting and neutral weighting, namely +5 and 0. Naturally, the length of these lines should be short since the similarity values are high. Figure 7.5 and 7.6 show these, with the solid line indicating highest association, the dashed line indicating second highest and the dotted line indicating third highest.

A comparison of the results shows that irrespective of the local monotonicity criterion, the motives Revenge and Elimination overlap in the FSSA space, sharing items in common. This is also found though to a lesser extent with the Sexual and Commercial motives. This would indicate that the motives are not exclusively related to one set of items only, but share meaning with many items to a lesser degree. Any regional partitioning thus cannot be thought of as delineating strict exclusive types but rather themes with blurred boundaries.

Looking at the length of these lines in the diagrams, there is a difference between the totals involved, as measured using the derived distance matrix. The total distances (i.e. lengths of the lines marked on the plots) was 2.6% greater in the 0 local weighting solution compared to the +5 local weighting solution. In fact, just looking

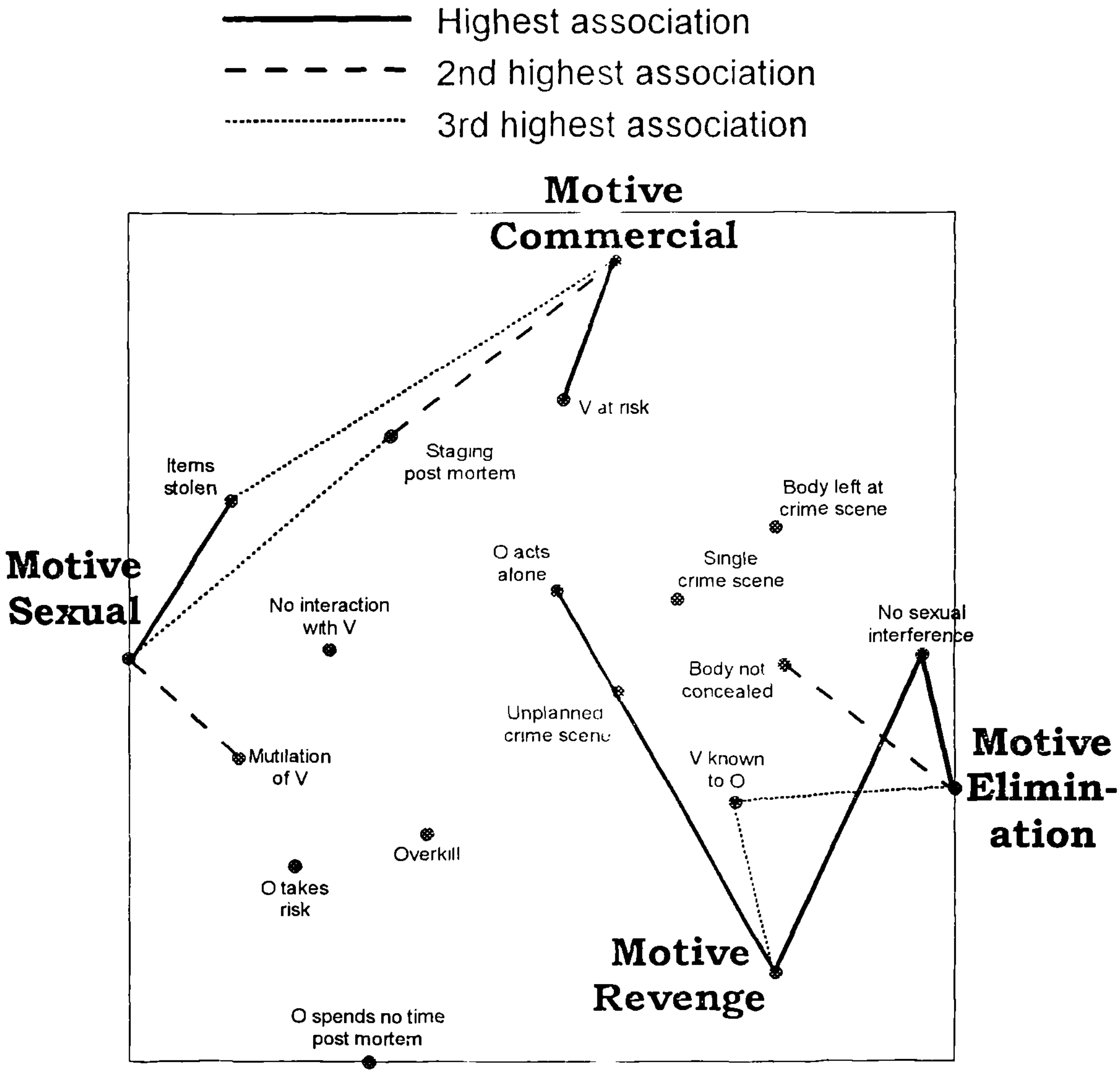


Figure 7.5. Highest similarities in +5 local monotonicity weighting for FBI motives .

at the value of the highest four similarities the 0 local weighting solution is 10.4% greater than the +5 solution. In other words, the high weighting ensures that these higher similarities are indeed represented as marginally shorter distances, increasing LSB in terms of distances.

However, the important point to note is that the +5 solution provides clues about how to find meaningful regional partitions that have conceptually inter-related items with good LSB. For example, even though the distance between the Revenge motive and the Single Offender item is large, these items are very highly similar in meaning and should be identified as being part of the same subcontent universe.

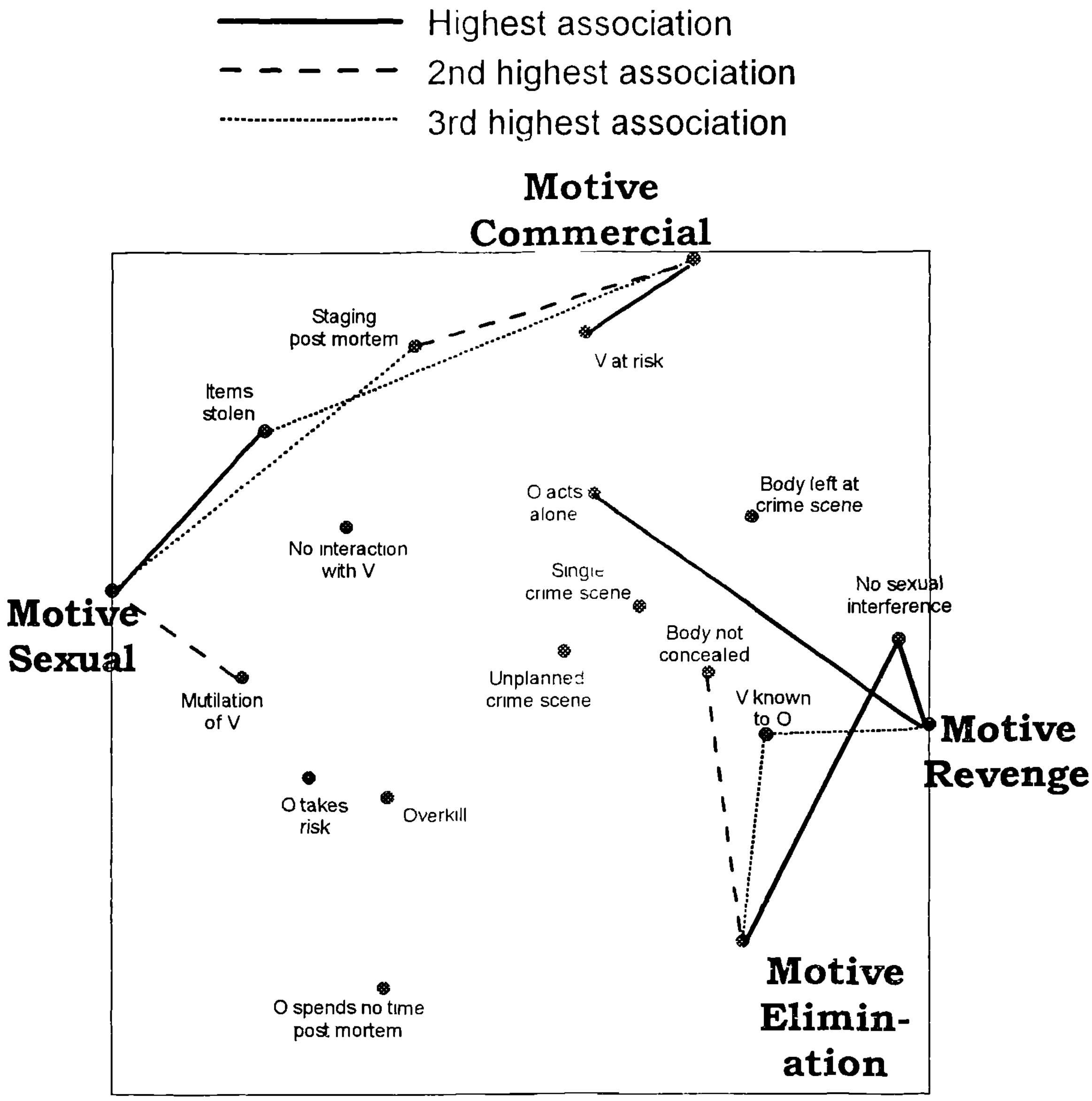


Figure 7.6. Highest similarities in 0 local (i.e. global) monotonicity weighting for FBI motives

In summary, LSB is the property of creating sound conceptual regions using local monotonicity and seeing if empirical data similarity supports conceptual proposed similarity. Since the +5 local weighting representation is better in terms of LSB and more suited to the substantive demands of the data set, Empirical Study 7.5 examined the regional interpretation of this plot.

Empirical Study 7.5: Regional Interpretation of the +5 Local Weighting FSSA of the FBI Data

The +5 local monotonicity weighting FSSA solution for the FBI data in Figure 7.5 was re-examined and regional hypotheses were made noting LSB. The regional interpretation is shown in Figure 7.7.

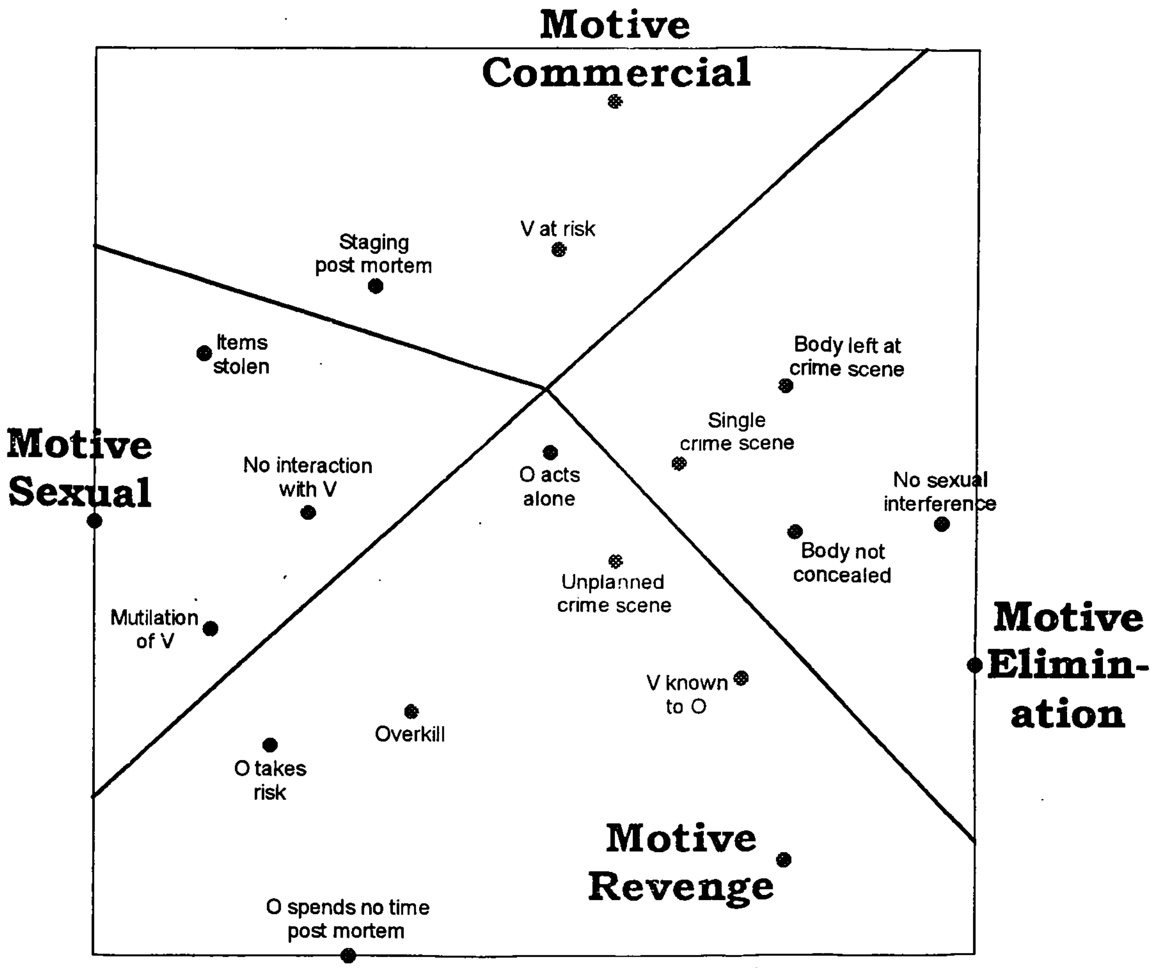


Figure 7.7. Regional interpretation of +5 local monotonicity weighting for FBI data

The solution gave a Coefficient of Alienation of 0.309, and the following regional themes were hypothesised:

- **Commercial Region**

This region contains elements of staging at the crime scene, with the victim at an elevated risk of targeting. The key variable in this region in terms of item correlation with scale total is staging. Close to this region was the item 'Single Offender', indicating that the murderer often acted alone. In other words, this theme is highly instrumental in intention and with low emotionality. Most of the homicide types that are typified by this region are the group of 'Criminal Enterprise Homicides', including Contract (Third Party) Murder, Kidnap Murder and Individual or Commercial Profit Murder.

- **Elimination Region**

There is typically no evidence of sexual assault with a motive of elimination, the body being left at the crime scene in plain view, and only one crime scene also associated though to a lesser extent. Group Cause homicides are primarily described best by region, with the absence of any sexual elements which are found on the opposite side of the FSSA space in the Sexual theme. Also represented are some Personal Cause homicides. Ideology is a typical reason for murder in this region, with homicide types of Political, Religious and Socioeconomic Extremist Homicides by individuals and groups. In these cases, victims are targeted on the basis of what they represent to the offender rather than who they are in themselves. The region may include to a lesser extent elimination of a victim who was known to the offender or offenders, such as in Drug Murder or Spontaneous Domestic Murder. For this reason the elimination region is close to the revenge region and the 'Victim Known' variable.

- **Revenge Region**

The revenge motivated theme has a single offender taking risks and using excessive force in an unplanned attack on a known victim, and immediately leaving the crime scene. Some Personal Cause homicides are explained by this theme, also with some Criminal Enterprise homicides to a limited extent. Normally the victim has some immediate personal significance to the offender and has 'wronged' the offender in some way that is worthy of retribution. Also in this region was the item 'Single Offender', meaning that the murderer tended to act alone. Key variables in this region are the offender taking risks at the crime scene. such as being easily identifiable as the murderer or attempting to attack a difficult target, and not planning out the attack. These murderers are driven by irrational and emotional rage, and naturally are opposite the cold, calculated commercial region in the SSA space. Domestic Homicides, Authority Killing and Revenge Killing are identified by this region. The intention of these homicide behaviours is directed towards that particular person and who that person is, rather than who he/she represents as in the Elimination intention. The low Cronbach's α relative to the other themes is a reminder that this region includes items of low inter-rater reliability, showing that they were less conceptually distinct in the text according to the raters, referring back to Empirical Study 7.1.

- **Sexual Region**

The sexual region contains no evidence of any meaningful interaction with the offender who subsequently mutilated of the victim's body and stole items from the scene. This is almost exclusively the Sexual Homicides of Organised, Disorganised, Mixed and Sadistic. The key items here are stealing items from the scene - the offender's 'trophies' - and mutilation of the victim, which in itself may be sexual for the offender. Notably close to this region was the variable 'Single Offender', which was accounted for by the serial sexual killer who was acting alone, namely the Sexual Homicide group of types.

This partitioning was done in an attempt to preserve good LSB. In other words, this required the examination of the most similar items to the motives and recognising that if there were Contiguity between conceptual and empirical similarity then these would form a region. Without this Contiguity, such strong regional hypotheses would not only be hard to find in the representation but also to defend as a structural interpretation of the domain.

Even though alienation is high - more than the default +2 weighting and 0 (global) weighting - this meaningful regional interpretation is a demonstration of the substantive requirements of a representation overcoming the computational aim of minimising error fit. It was necessary to use a high local monotonicity weighting to recover structure in this solution. Other data sets may have even more stringent local monotonicity and LSB need solutions requiring possibly higher weightings than +5. As Shye (personal communication) mused about local weightings 'one could also ask why stop at +5?'

Localised Spatial Bonding and Reliability

So far, LSB has been examined in terms of monotonicity transformation and the distances in the geometric space - i.e. the representation. The importance of finding this property, however, concerns the nature of the items found in the region of LSB - i.e. the structure.

Items in an area of LSB will have a high intercorrelation and act as a partial measure for the meaning of the sub-space, given Contiguity between (conceptual) definition and (empirical) data. To an extent, the property of LSB has similar characteristics of a typical reliability construct, as intended by the domain sampling model of test construction (e.g. Ghiselli, *et al.*, 1981). LSB therefore lends itself to reliable plots. Since the use of regional interpretation and local monotonicity is geared towards the creation of LSB, it is also therefore geared towards the creation of reliable constructs or sub-constructs.

The usual measure of reliability in measurement theory for items that form a scale - the role the region essentially plays - is Cronbach's α, or its dichotomous equivalents the Kudar-Richardson formulae 20 and 21. These act as the average of all possible corrected split-half reliability coefficients (Ghiselli, *et al.*, 1981). Cronbach's α can act as a useful check on LSB and the reliability of the region, since it is theoretically possible to find low α reliability even where LSB is strong. This would occur if the items in a region were of low, zero or even negative correlation, but were nevertheless placed close in a region. This would be detected by the use of reliability testing with Cronbach's α. If an item were to be placed in a region when it is in fact not a partial measure of the region, then the Cronbach's α calculation with the item deleted would reveal this. On discovering this, there is a trade-off between on the one hand the substantive weight (if any) added to the region by the point, and on the other its poor intercorrelation.

However, the value of Cronbach's α can only be used as a clue to LSB rather than as a firm index for two reasons. Firstly, LSB is more of a substantive concern to be argued qualitatively given the nature of the representation. This may mean that items with a low empirical closeness (i.e. correlation) may nevertheless be conceptually close under a particular structural hypothesis. In such an instance, the partitioning according to LSB rather than Cronbach's α would be justified if the nature of the data allowed it, if the research were highly exploratory, etc.

Secondly, in most cases the similarity measure between points is not the same as that used by Cronbach's α, namely Pearson's r, the Product Moment Correlation Coefficient (PMCC) or ϕ, its dichotomous equivalent. This problem arises where an association coefficient such as the Jaccard's index was used because this coefficient may not satisfy the distributional assumptions of Pearson's for the Cronbach's α calculation (see chapter 8). This may also occur where there is a direct judgement of similarity.

Empirical Study 7.6 examined the relation between LSB, reliability and local monotonicity.

Empirical Study 7.6: Local Monotonicity, LSB and Reliability

The Cronbach's α reliabilities for the FBI data were examined with FSSA under three conditions. These were:

1. +5 local monotonicity weighting with substantive interpretation using LSB;
2. +5 local monotonicity weighting with heuristic interpretation, and
3. 0 local monotonicity weighting with heuristic interpretation.

The substantive partitioning method for one +5 weighting condition was the regional interpretation outlined in Empirical Study 7.5. The other +5 weighting and 0 weighting conditions used an alternative the method of finding whichever motive was physically closest to each non-motive item. This method was somewhat analogous to the single linkage or nearest neighbour technique in cluster analysis (e.g. Everitt, 1993).

It would be hypothesised that the +5 weighting regional interpretation using considerations of LSB and the local monotonicity would create the most robust partitions, given that Empirical Study 7.5 demonstrated Contiguity between definition and data. Consequently, it should give the highest Cronbach's α scores. Furthermore, it would be expected that the +5 weighting heuristic regions should be better than the

0 weighting heuristic since the local monotonicity criterion should mean that most empirically similar items are more accurately represented i.e. closer together. In other words, there should be better LSB even if there were no theory guiding it. However, it would not be expected to be as good as the substantive regional interpretation with the same +5 local monotonicity weighting.

The Cronbach's α reliabilities produced by these three conditions are reproduced in Table 7.7.

Partition Method	Local Weighting	Motive				Average
		Commercial	Elimination	Revenge	Sexual	
Regional interpretation	+5	532	735	352	656	**543**
Shortest Distance	+5	533	652	182	636	**500**
Shortest Distance	0	533	020	604	636	**448**

Table 7.7. Regional reliabilities for FBI motives according to different partitioning rationales
Values for Cronbach's α coefficients are without decimal points

Caution must be attached to these Cronbach's α results since the similarity measure was Guttman μ not PMCC, though these are similar functionally (see Empirical Study 8.1). However, the averages of the results do suggest that the regional interpretation with +5 local monotonicity weighting was indeed the best, followed by +5 local weighting with heuristic regions and then 0 weighting (global monotonicity). Thus for these data any sort of weighting for or against small distances should be sought. Given that the faceted approach should adopt LSB where possible, the conclusion is that the +5 solution is preferable even if it leads to higher alienation, as was shown earlier. In other words, locally accurate representation is more important than a globally accurate representation even if the local representation is less accurate overall.

Looking at the individual scores, it can be easily seen that the cause of most of difference is the relative adequacy of the solutions to find reliable regions with the Elimination and Revenge motives. It will be remembered from Empirical Study 7.3 that the configuration of the plots were very similar except for the Elimination and

Revenge motives changing place. In other words, the non-motive items space itself stayed unchanged but the two motive items swapped.

To confirm that the local weighting is creating the correct placing of these two motives, the second part of this Empirical Study calculated the correlation between the Elimination and Revenge motives and the two non-motive groups of items. Figure 7.8 shows the correlations between the groups of non-motive items in the themes and motives that may be used to describe them.

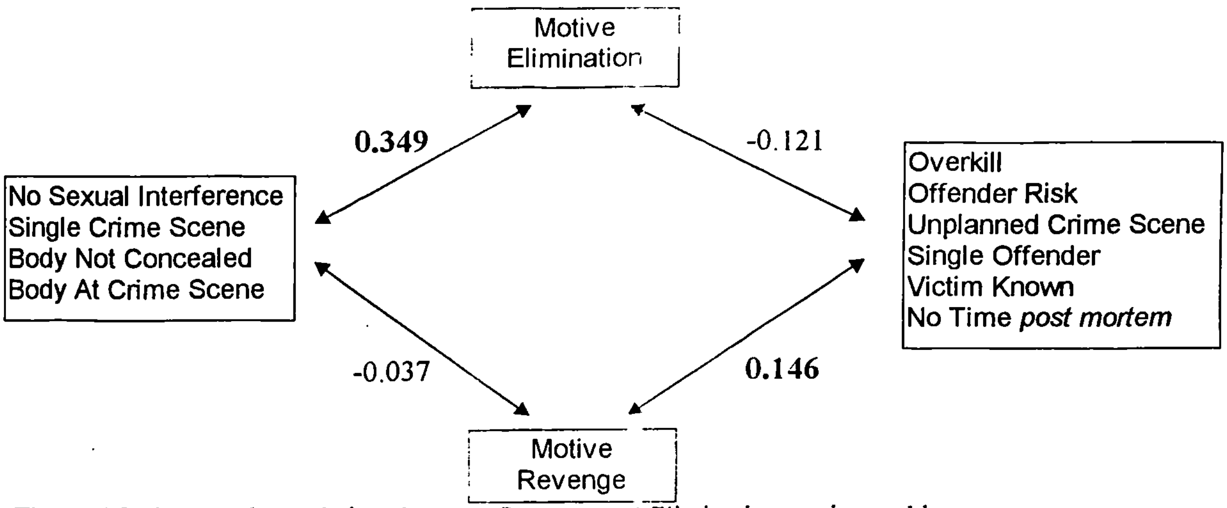


Figure 7.8. Corrected correlations between Revenge and Elimination motivesand item groups

Clearly the correct 'match' in terms of highest motive-group correlation is also the best substantively; namely that Revenge describes best the emotional group headed Overkill and that Elimination describes best the non-emotional group headed No Sexual Interference.

Rejecting Structural Hypotheses with Local Monotonicity

In the same way that local monotonicity can improve the power of representation to test the existence of faceted structure, it can also be used to reject structural hypotheses. Thus the example with the FBI data set showed how local monotonicity could prevent a False Negative or type II error. A re-examination of the Levelt, van de Geer and Plomp (1966) data by Shepard (1974) can show how locality can be used to reject structural hypotheses in a non-criminal domain.

Levelt, *et al.* (1966) gathered data on the perceived similarity of musical scale intervals. Levelt *et al.* attempted to explain the non-metric representation of these data in a two-dimensional plot, using a dimensional interpretation of the plot. Levelt *et al.* fitted a parabola to the representation, an attempt which would seem to be extrinsic (Shye and Elizur, 1994) to the design and analysis of the data since the fit to the parabola introduced error of its own.

The reanalysis of these data by Shepard (1974) was driven by the structural hypothesis that the data were fundamentally one-dimensional. Shepard claimed to have recognised that a higher dimensional solution was diagnosed for the data than was necessary, going on to show that both the Levelt data and an independent replication of it could be understood by placing a semi-circular horseshoe on the data.

Shepard pointed out that the similarity matrix was similar to the simplex of Guttman (1954b). Because a semi-circle in two dimensions is monotonically identical to a line in one dimension, the representation was indeed a simplex, argued Shepard. Though it was not specifically stated in that section of the text, Shepard is clearly referring to an adoption of a *local* monotonicity transformation, since he then went on to argue how points on the surface of a steep parabolic dish could be decomposed to the idealised radex, as mentioned before.

The configuration suggested by Shepard (1974) is represented schematically in Figure 7.9.

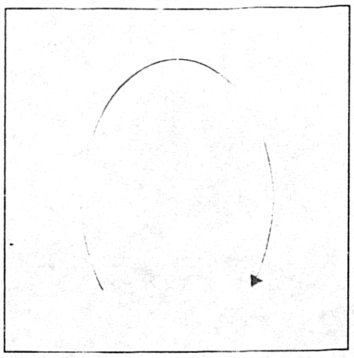


Figure 7.9. Schematic representation taken from Shepard (1974)

Lingoes and Borg (1977, p. 129) followed Shepard's line of reasoning up to a point, agreeing that approximately 'C' shaped curves in two dimensions could be modelled as simplexes. However, they did point out that this was only the case when the configuration did not bend back on itself, as Shepard's clearly did. Only if data were perfect could such a horseshoe as Shepard's be a simplex; these data were far from perfect.

Lingoes and Borg (1977) also stated that this trend would be found irrespective of the proportion of tied values were in the similarity matrix. This implied that bending back on the seriation of Kendall (1971) was not due to the limited range, contradicting Shepard (1974). Lingoes and Borg instead cited their own evidence to suggest that the initial configuration was the cause of the bending, as was the case with the Levelt *et al.* data.

The simple way to decide if the data truly are a simplex, as Shepard (1974) hypothesised, is to try scaling in one dimension with strong local monotonicity.

Empirical Study 7.7: Locality of Monotonicity and the Dimensionality in the Levelt et al. Data

By scaling the Levelt *et al.* data in one dimension with local monotonicity, the true structure of the data can be judged. If the horseshoe was a two-dimensional representation of a simplex, then a one-dimensional SSA-I with local monotonicity should recover the order hypothesised by regression of the points onto Shepard's horseshoe, as was shown by Borg and Lingoes (1977, p. 129) with their curved two-dimensional approximation to a simplex. (Unfortunately it was not possible to scale these data in one dimension with Shye's FSSA and a +5 local weighting since the minimum dimensionality possible in FSSA is two.)

The one-dimensional hypothesised sequence of Shepard's horseshoe goes from a 2:5 interval to a minor 2nd for the 15 intervals used by Levelt *et al.*, according to Shepard (1974). For simplicity, these are numbered 1 to 15.

The results of the SSA-I with local monotonicity are shown in Table 7.8.

Hypothesised structure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Empirical structure	4	3	6	5	2	8	1	7	9	10	11	12	13	14	15

Table 7.8. Hypothesised and empirical structures for one dimensional solution of Levelt *et al.* data

The one-dimensional SSA-I with local monotonicity produced a result with an alienation of 0.4055. This was high, even with the low dimensionality and high number of variables taken into consideration. Looking at the sequence produced, it clearly did not follow Shepard's horseshoe-derived simplex.

There is however one way to make it possible to recover a simplex from the Levelt *et al.* data. That would be done by only looking at what is the single highest association between points, and weighting that several magnitudes above all other associations. This is because all the Levelt *et al.* musical intervals except one have their highest or joint highest similarities with intervals adjacent on Shepard's horseshoe. In terms of the association matrix, this corresponds to the value immediately below the diagonal. Doing this would be equivalent to weighting local monotonicity extremely highly, such that any similarity other than the very highest would be ignored. Such an exponential relation is illustrated in Figure 7.10.

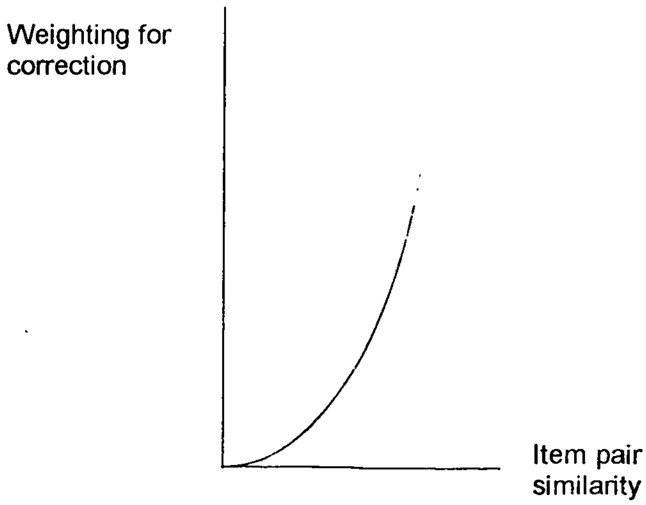


Figure 7.10. Hypothetical weighting required for MDS to ignore low similarities

However, creating such extreme weightings for some values and not others - i.e. virtually excluding low similarities could lead to the *construction* of a simplex, rather than *testing* for a simplex, which would be unacceptable. This weighting method would be similar to the linking done in single linkage or nearest neighbour cluster analysis (e.g. Everitt, 1993), since only the highest association value is considered at a time for joining the objects into clusters.

As was admitted by Borg and Lingoes:

> It is not possible, of course, to give here a simple answer as to how much violation of the simplicial structure in the data matrix and its geometric representation should lead to a rejection of the simplex as a model understanding the data. (Borg and Lingoes, 1977, p. 130)

The question still remains as to why the data curved back on itself. Borg and Lingoes (1977) suggested this was not due to the ceiling of dissimilarities. But the case of the Levelt *et al.* data there was a distinct 'flooring effect' of similarities - the same phenomenon but in the opposite direction to the 'ceiling effect'. Figure 7.11 shows that this was the case for the Levelt *et al.* data.

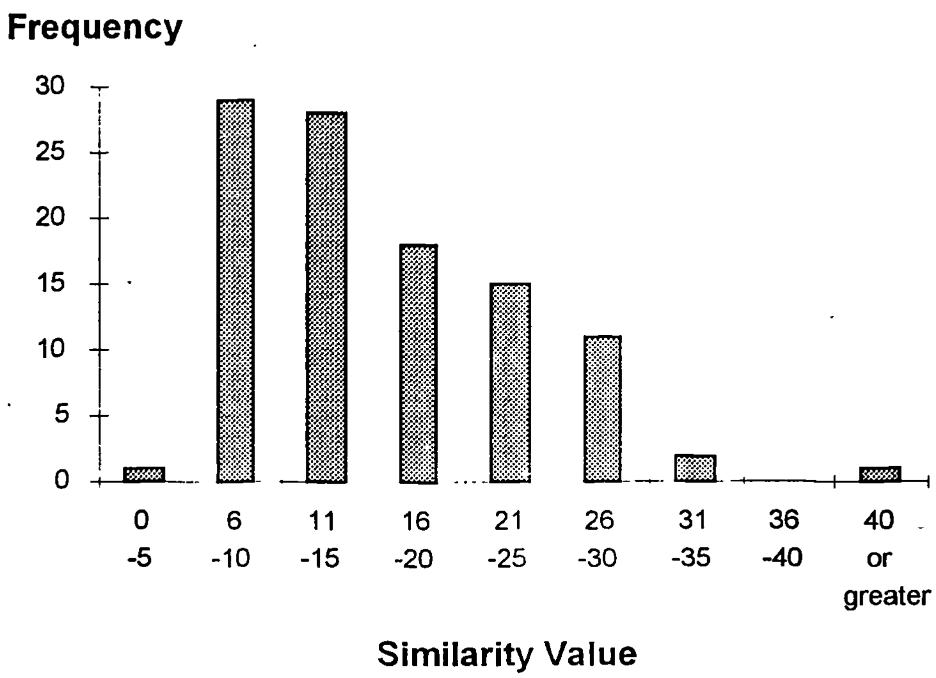


Figure 7.11. Histogram of values in similarity matrix of Levelt *et al.* data

Other than one value at 73, the vast majority of values were skewed towards the lower end of the scale, with for example the value 10 appearing nine times, 9 appearing seven times, and 7 and 8 appearing six times.

It is still possible that Levelt *et al.* data were curved due to the effects of the starting configuration: this cannot be decided from these results. However, a two-dimensional analysis of the data showed that a global monotonicity solution produced a more rounded circle than a horseshoe, though still with a segment of the circle missing. The alienation for this solution was lower than the local monotonicity two-dimensional solution. This might even suggest that the data were in fact a circumplex with some key intervals missing to complete the circle, a not implausible structural hypothesis given that the colour circumplex of Shepard (1978) on the Ekman data also 'filled in' part of the circle. The proof of this would be to find a musical chord which completes the circle.

Summary of Chapter 7

This chapter assessed the influence of local monotonicity on representation and structure, introducing the concept of Localised Spatial Bonding (LSB). The FBI data set was taken to illustrate a substantive data set in need of such a concept. Empirical Study 7.1 showed that a particular representation did not support a hypothesis of structure, which was taken to be a false negative. Suggestions were made about the need for local monotonicity, though Empirical Study 7.2 showed this would be at the expense of higher alienation. Empirical Study 7.3 further suggested that local monotonicity could have implications for structural shapes as well as alienation, and that LSB could illuminate the correct structure in a badly fitting representation using the original correlation matrix, according to Empirical Study 7.4. A meaningful regional interpretation of the FBI data was found with local monotonicity and due consideration of LSB in Empirical Study 7.5. Suggestions were made about the relation between LSB and reliability, measured by Cronbach's α. Accurate local sub-spaces with good LSB have higher Cronbach's α reliability, though this depends on good Contiguity. Empirical Study 7.6 proved that where a regional structure is found with good LSB and strong structural hypotheses using local monotonicity then the Cronbach's α reliability was higher than with no structural hypotheses or local monotonicity. LSB and local monotonicity was used in Empirical Study 7.7 to explore an ongoing debate over the structure of musical scales.

Chapter 8
association coefficients and structural hypotheses

Coefficients of Association and Similarity

In the previous chapter, local monotonicity was shown to be important for geometric representations of criminal action information, creating the Local Spatial Bonding (LSB) property. This chapter examines an equally important structural parameter which determines and measures the 'interrelatedness' of the constituents of a domain - the association or correlation coefficient. A substantive data set termed the 'Child Abuse Data Set' is used to demonstrate this. The need for this particular data set to achieve these tests is related to the large number of items on which the material was originally coded and the unusual and potentially biased nature of the material. The significance of local monotonicity is suggested furthermore to be accentuated by the choice of certain association coefficients.

The monarch of similarity measures is widely recognised as r, the Pearson Product Moment Correlation Coefficient (e.g. Carroll, 1961; Liebetrau, 1983; Gibbons, 1993). It is the only suitable coefficient for factor analysis, according to Gorusch (1988), though its ordinal equivalent Spearman's ρ and the nominal against ordinal level point biserial coefficient. Von Eye (1988) stated that the PMCC is a parametric coefficient requiring - like all other parametric statistics - that the following key conditions must hold:

- the shape of the distribution must be known, such as a bivariate correlation having a binormal distribution,
- the dependent variable is at the interval level,
- the sample is random. (also necessary for non-parametric tests), and
- the sample size must be greater than the number of predictors and number of criteria.

Where these conditions cannot be met then non-parametric alternatives should be used. Parametric techniques are preferable where the conditions can be met, since the non-parametric tests are less efficient than parametric ones - a diminished efficiency that increases as sample increases (von Eye, 1988).

But how well are these strictures followed in real research? Macdonald was pessimistic:

> Random samples are the exception rather than the rule in psychological research and the distributional assumptions made by parametric tests are almost certainly wrong. (Macdonald, 1997, p. 340)

This assertion was backed up by evidence from Micceri (1989; cited by Macdonald, 1997) that in 440 distributions of raw data obtained from published large sample psychometric research, not one claim of normality in these distributions was justified, with p values for non-normality all < 0.01; yet all still made parametric assumptions.

There is a variety of alternatives to the PMCC, and some have been specifically developed with certain data types or qualities in mind. For example, the dichotomous Jaccard association index was proposed originally by Jaccard (1900, 1908; cited by Snijders, Dormaar, van Schurr, Dijkman-Caes and Driessen, 1990) as a measure that would not count objects (e.g. ecological site) as more similar because neither contained some attribute (e.g. plants in common).

Table 8.1 shows the key to the calculations of the various formulae.

		Variable 2	
		Yes	No
Variable 1	Yes	*a*	*b*
	No	*c*	*d*

Table 8.1. Key to calculation of coefficients

Using the key in Table 8.1, the index can be simply calculated in Formula 8.1.

$$J = a / (a + b + c)$$

Formula 8.1. Jaccard's measure

The cell denoting joint absence of both variables, cell *d*, is not counted in this coefficient. By contrast, the PMCC for dichotomous data is calculated as in Formula 8.2.

$$r = \frac{ad - bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$$

Formula 8.2. Pearson's r

The most notable feature of Jaccard's is the exclusion of cell *d* from the calculation. Therefore joint non-occurrence - conjoint absence - is not part of the calculation of Jaccard's. This property was deliberately included by Jaccard so as to be useful in 'ecological studies when a series of sites are being compared and these possess a few species common to each with the remainder restricted to a few of the sites.' (Clifford and Stephenson, 1975, p. 54) Jaccard's 'reflects the proportion of events where both [variables] occur, given at least one of them occurs' (Bilsky, Borg and Wetzels, 1995, p. 43).

In the context of criminal actions, this would refer to those cases where overall there was little behaviour recorded as present, with only a few behaviours being shared in common by most cases. To reify for the sake of an analogy, Jaccard's might be seen as searching for some behaviour that was recorded as occurring for one variable and then search to see if it also was recorded as happening for another variable. If it was recorded as happening for the other variable, it is significant; if it was not recorded, then it is less significant. If the first variable was not recorded at all, then Jaccard's would not 'search' at all so then it is even less significant.

The Importance of Conjoint Absence

Generally behaviour coded dichotomously is denoted as being 'present' or 'absent'. However, it was noted above that for criminal actions data such as in Kirby (1993), this in practice means 'present' and 'not recorded as present', due to the uncertainty. More simply, this could be phrased 'recorded' and 'unrecorded'. To reduce the uncertainty, it is desirable to use the 'unrecorded' coding as little as possible.

In Jaccard's, there is a cut-off point below which the uncertainty is too great. This point is simply where both behaviours are 'unrecorded', though one 'unrecorded' is acceptable and below the 'just noticeable difference' (JND) of noise. If there is an amount u of uncertainty, $2u$ would be greater than the JND while u would be less than the JND. With reference to the key in Table 8.1, Figure 8.1 schematically represents this as a partial order. (Partial orders are explored in greater detail in chapter 15.)

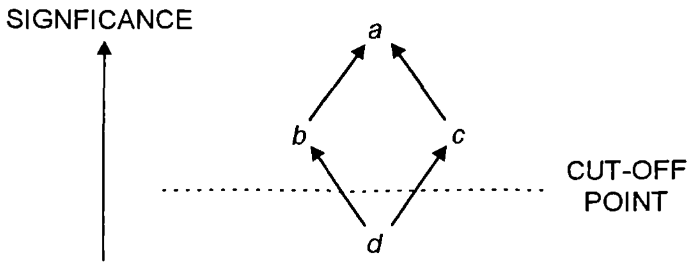


Figure 8.1. Partial order and just noticeable difference in uncertainty of associations

The significance of the d cell is therefore downplayed relative to the others, and is in fact not used in the calculation of Jaccard's. The inclusion or exclusion of this cell neatly divides many of the various measure of similarity, and the subsets created on this basis are monotonically related to each other (Gower, 1985, p. 399). The importance of the d cell is fundamental and is best illustrated by comparison of Jaccard's in Formula 8.1, excluding the d cell, with a similar coefficient including the d cell.

The addition of the d cell into the denominator of the Jaccard's equation gives a measure known as the Russell and Rao (Gower, 1985). The Russell and Rao calculation is given in Formula 8.3.

$$Russell\ \ and\ \ Rao = \frac{a}{a+b+c+d}$$

Formula 8.3. Russell and Rao measures

Since the denominator of the Russell and Rao is constant across values of the cell - namely the total number of cases or respondents - then the value is simply the proportion of all cases with conjoint presence.

The Russell and Rao is equal formulaically and numerically to Jaccard's when cell $d = 0$; otherwise the numerical value is less than that of Jaccard's. For any value of d other than 0, the Russell and Rao coefficient is lower than that of Jaccard's. Consequently conjoint absence lowers the similarity of two objects or variables. In the context of criminal behaviours, this would be appropriate if the uncertainty ratio was such that most of the 'not present' coding was because it did not happen, rather than it was not recorded.

Where the uncertainty ratio is high (i.e. poor certainty) it cannot be justified to use the stronger coefficients that include the d cell. The ratio varies across variables, but the same coefficient must be used across all data to ensure comparability meaning that the coefficient is determined by the weakest variable. Where there are no weak variables - or they have been excluded from the analysis - then stronger coefficients may be justified. The Russell and Rao is therefore very similar to Jaccard's, but slightly stronger especially where there is much conjoint absence in the d cell. If there were much uncertainty with the d cell then this would be contrary to obtaining a realistic representation.

Another commonly used and stronger coefficient than Jaccard's or the Russell and Rao is Yule's Q. This is a dichotomous measure equivalent to Guttman's coefficient of (weak) monotonicity, μ (Shye, 1985a, p. 73), which is used in its semi-strong form in the calculation of alienation in SSA-I (Guttman, 1968, p. 480; Borg, 1978, p.478). With reference to the cells of the Table 8.1, Q is explained in Formula 8.4.

$$Q = \frac{ad - bc}{ad + bc}$$

Formula 8.4. Yule's Q

The importance of the d cell in Formula 8.4 is even greater since it is a multiplicative rather than additive factor, as it is in the PMCC. This means that if the 'unrecorded' coding - and more so for the d cell - was too full of uncertainty then the measures derived would be inappropriate or inaccurate. Marsden and Laumann (1978) characterised Yule's as showing the proportion of consistent as opposed to inconsistent pairs of cases for the two variables. The relation falls down, however, when any cell value equals zero, meaning that Yule's takes an extreme value of +1 or -1 automatically. Therefore it is sensitive to sample size: in a small number of paired comparisons, there is a higher chance that one cell will equal zero.

To generalise from what Carroll (1961, p. 349) stated in his presidential address to the Psychometric Society, no assumptions are necessary to *calculate* any measure, but *valid inference* relies on adequate matching between reality and model. The information provided on the homicide types in the FBI data set were intended by Ressler *et al.* (1992) to be complete and definitional for each of the homicide types. Therefore even though the adjectival qualifiers such as 'usually found' and 'sometimes observed' were fluid, the presence or absence of the items was more certain. This was because if features were found in a homicide then they tended to be noted as being present, with those not found being either noted as absent or inferred as absent. By this rationale, the Jaccard's index may be particularly suitable where data were extremely noisy, such as the Child Abuse Data Set.

The Kirby Child Abuse Data Set

Qualitative recorded observations in police files were characterised in chapter 1 as being from a partial content universe. That is to say, under the expanded CRM of Coombs (1964) phases 0 and 1 have already been decided by someone external to the research - thus the relevant questions and answers to them have been written as statements.

Kirby (1993) made use of recorded observations about sexual offences committed against children between 5 and 12 years from the following sources·

- crime reports of the initial complaint and inquiries about that complaint,
- victims' statements taken down by suitably trained detectives,
- physical evidence of police surgeon where necessary and consented to by the victim,
- forensic evidence such as swabs, fibres or fingerprints,
- suspects' statements about the alleged offence as interviewed by detectives, and

- offender records such as previous criminal history, age and marital status.

Kirby examined records on 416 separate offences against children in the Lancashire Constabulary jurisdiction for the three year period from 1987 to 1989. These offences had been committed by 97 different offenders. The last known recorded offence for these 97 people were taken as a sub-sample of 97 on the basis that it would be a more valid sample since the offence for which an offender was arrested would be the most complete and accurate.

The research carried out by Kirby was within a partial content universe already established by the investigating officers in charge of the case. This was because Kirby was using the recorded observations that had been through phases 0 and 1 of the expanded CRM, meaning that when Kirby went through phases 0 and 1 then the content universe was not complete. Kirby designed items to code the material having already had practical experience of what differentiates offenders. Though it was the first study of its kind in this particular domain, nevertheless the items were guided by building on existing frameworks used successfully in the analysis of adult sexual assault by Heritage (1992) and Canter and Heritage (1990). Similarly to these studies, Kirby was unable to use a fully faceted design to develop the items and examine structure; but similarly, Kirby used the technique of content analysis to record observations. The full Child Abuse Data Set contains 59 variables scored dichotomously on the 97 offences, the details of which are given in Table 8.2.

Despite taking the last offence as the best quality source, there was nevertheless a great deal of error which must be considered when associating behaviour to model the structure.

No.	Description	Frequency
1	Child on own when offender first meets	19%
2	Child with others when offender first meets	49%
3	Child on own during offence	71%
4	Child with others during offence	29%
5	Offence inside	71%
6	Offence outside	30%
7	Offender previously on own with child*	61%
8	Grooming strategy used*	43%
9	Minimisation strategy used*	44%
10	Offender cons child into going elsewhere	54%
11	Offence is unplanned	45%
12	Offender was caring for the child	45%
13	Offender took drink/drugs prior to offence*	24%
14	Victim male	33%
15	Victim female	67%
16	Offender commits more than 1 offence against the child*	60%
17	Offences against child for at least 6 months*	29%
18	Offences become progressively worse during series*	32%
19	Offender shows affection towards the child*	37%
20	Offender reassures/talks to the child to minimise*	29%
21	Victim asked to participate*	40%
22	Offender deterred by adverse reaction or indication of no consent	37%
23	Offender not deterred by adverse reaction or indication of no consent*	27%
24	Offence facilitated through use of initial force*	39%
25	Force or threat of force used gratuitously*	11%
26	Offender shows remorse to the victim	4%
27	Offender threatens violence if victim reports*	9%
28	Offender uses non-violent threat if victim reports (e.g. plays on conscience)*	38%
29	Offender promises or gives a gift or money to victim*	26%
30	Offender naked at time of offence*	16%
31	Victim naked at time of offence*	24%
32	Sexually explicit language used by offender *	23%
33	Offender kisses victim on lips*	22%
34	Offender's penis erect*	47%
35	Offender places victim hand on penis*	39%
36	Victim required to masturbate offender*	33%
37	Victim required to fellate offender*	20%
39	Offender rubs victim's genitalia outside victim's clothing*	7%
40	Offender rubs victim's vagina inside clothing but does not penetrate digitally*	16%
41	Offender masturbates male victim*	20%
42	Digital penetration of vagina*	25%
43	Digital penetration of anus*	10%

Table 8.2. Details and frequencies of Kirby full data set
*indicates variable was used in main analysis of Kirby (1993)

No.	Description	Frequency
44	Actual or attempted penile penetration of vagina*	13%
45	Actual or attempted penile penetration of anus*	14%
46	Offender ejaculates*	31%
47	Victim stroked or touched in other than in genital area*	49%
48	Victim receives physical injury	3%
49	Offence disturbed	8%
50	*Offender is stranger to victim**	20%
51	Offender has spoken to victim several times prior to offence	32%
52	Victim knows offender well	44%
53	Offender claims common interest with victim*	18%
54	Offender gives victim an alias	5%
55	Victim less than 9 years when offence(s) start	54%
56	Victim goes to special school	7%
57	Victim has behavioural difficulties	15%
58	Victim has absent parent	44%
59	Offender has no previous convictions	42%

Table 8.2 (cont). Details and frequencies of Kirby full data set (cont.)
*indicates variable was used in main analysis of Kirby (1993)

Uncertainty and the Association Coefficient

Of the 59 variables in the Child Abuse Data Set, several pairs of items could not both be recorded as present, such as 'Child on own when offender first meets him or her' and 'Child with others when offender first meets him or her'. A crosstabulation of the two variables is given in Table 8.3.

$N = 97$		Child on own	
		Yes	No
Child with others	Yes	0	48
	No	18	32

Table 8.3. Crosstabulation between 'Child on own' and 'Child with others' from Kirby data set

As can be seen from this table, there were no cases of presence of both 'Child on own' and 'Child with others'. However, there were 32 cases where the presence of neither variables was recorded. Such a pattern was found in several of these pairs of exclusive variables. For there to be 32 conjoint absences, either the coder missed these cases or the data were incomplete to explain this pattern.

Turning to the first possibility that the coder missed these cases, Kirby acknowledged a small number of errors or disagreements in an inter-rater reliability check. There were 10 coding discrepancies in a sample from the 11,400 decisions made in the 416 offences matrix, and 64 discrepancies out of 5432 decisions in the 97 offence/offender matrix. These figures are admirably small, suggesting the second possibility is correct, namely that the data were incomplete.

The incompleteness of data can readily be understood given the nature of the police interview with the victim of the assault. During these interviews, as little prompting of the child as possible is done to minimise trauma but enough to elicit sufficient information on which to investigate and charge someone (Kirby, personal communication). Furthermore, given the vulnerability of children to leading questions and inadmissibility of such questions under the Police and Criminal Evidence Act of 1984, it is important that the child be in control of the interview. More recently, police forces have implemented policies to interview children only once so as to minimise even more the trauma of relating the incident. This would mean that missing evidence or information remains that way.

Returning to the example above, clearly in reality the child was either alone or with others when first approached. There may be cases where the child was unsure - perhaps he or she was in playground with other children present though not actually with them - but a well defined variable should make coding clear. Yet if the child was not asked or did not mention anything about this during the interview, the information would have been left out. In terms of the coding framework, information on being either alone or with others when approached is therefore not so much missing as *unrecorded*. But if the child did mention being alone when approached by the offender then coding the variable for being with others is marked as absent; there would be no way to be otherwise.

Another example illustrates this point further. In 46 out of 97 cases the child stated that the offender's penis was erect throughout or at some time during the offence. But it cannot be assumed that in the other 51 cases there was no erection - that the variable should be coded as absent. As suggested above, perhaps this question was not asked or was not mentioned by the child - the information was unrecorded. Furthermore, it is possible that the offender's penis was erect but the child did not notice. Is it possible to tell how many were coded as 'offender erect penis is not present' when this was unrecorded rather than absent?

There is some theoretical *uncertainty ratio* of 'unrecorded' to 'absent' information which acts as an index of noise in data. In this context of analysis from a partial content universe the ratio can act as an index of the nature of the 'not present' coding decision. While there are certain robust items that are low in uncertainty - such as recording the gender of the victim - there are some questions that may not be specifically asked, such as whether the offender had an erect penis during the assault. Assessing the uncertainty ratio in a single 'not present' decision requires an understanding of the way the observations were recorded by the person who had access to the full content universe, namely the investigating officer in complaints of sexual assault against children.

Yet all this information need not be treated as singular and unrelated to wider range of behaviours done during the assault. All the information was combined and associated with other information by Kirby (1993) to create quadrant IV similarities data (Coombs, 1964), phase 2 of the CRM. What is important to note is that the quality of the information varies when combined with different pieces of information. If the information on the offender's penis being erect were combined with information on whether or not the offender was naked at the time then the certainty with the offender's erection changes. Table 8.4 illustrates this point.

N = 97		Penis erect	
		Yes	No
Offender naked	Yes	13	3
	No	33	48

Table 8.4. Crosstabulation between 'Offender naked' and 'Penis erect' from Kirby data set

In the 3 cases where the offender was naked, we can be more certain that his penis was not erect than in the 48 cases where he was neither naked nor with an erect penis. The uncertainty ratio as to whether the offender had an erect penis will be lower (i.e. the researcher is more certain) given the offender was naked rather than if it were given that the offender was not naked. In other words, where the offender was not naked, more uncertainty exists.

But if instead of the crosstabulating with the variable 'Offender naked' the alternative variable 'Victim performed oral sex on the offender' were to be used, then the uncertainty ratio changes even more markedly, Table 8.5 illustrates this point.

N = 97		Penis erect	
		Yes	No
Victim fellated offender	Yes	19	0
	No	27	51

Table 8.5. Crosstabulation between 'Victim fellated offender' and 'Penis erect' from Kirby data set

For oral sex on the offender to occur and to be recorded as such, the offender's penis must be erect. According to the data, in all of the 19 cases where oral sex occurred the offender's penis was recorded as being erect. The uncertainty ratio is zero since it would be difficult for a child reporting having fellated the offender without reference to the offender's erect penis.

By contrast, 27 offenders were reported as having erect penises without any reference being made to fellation. Now it would seem likely that the child would report fully or be questioned specifically about the sexual content of the assault. It is therefore more likely that fellatio was absent and did not occur for these 27 offenders, rather than it went unreported. In other words, fellatio was absent rather than unrecorded for 27 cases given that the offender's penis was erect; the uncertainty ratio is very low.

Where the original case files cannot be accessed - as was the case with the present analysis of the Kirby data - such crosstabulations and the reduction of uncertainty can be used to calculate the quality of the original information. If there were cases where fellatio was present but erect penis was absent then clearly the absence was a case of 'not recorded' rather than 'not present'.

This leads back to the original question about the number of offenders for whom the 'offender erect penis' variable should be coded 'absent' rather than 'not recorded'. Information from a variety of crosstabulation would indicate the extent of 'not recorded' as opposed to 'not present'. For example of the 51 'erect penis not present' cases, then at least 3 of them were 'absent' rather than 'not recorded' since these three were recorded as offender being naked. The amalgamation of these would reduce the uncertainty ratio, but would add an entirely new extrinsic source of unwanted error.

The implications of the uncertainty ratio is that certain association values are more unreliable than others. Items which are identically used by police investigators for evidence and by researchers for behavioural information - such as fellatio on the offender - are likely to have low uncertainty. The converse is true of good indicators of the intent of the criminal action - such as language - but which are of poor evidential quality are likely to have high uncertainty. Regional hypotheses made on the basis of LSB between an unreliable item pair are less robust than those regional hypotheses made on low uncertainty. Yet the simple exclusion of items with high uncertainty may hide the structure of the domain. Instead, the association of these behaviour must take this uncertainty into consideration and not bring error in another form.

Roskam (1981) stated that finding a similarity measure suited to the data analytical task at hand was one of the most fundamental stages in MDS research. Roskam warned that 'At least some theory is require to choose a particular similarity index' (Roskam, 1981, pp. 214-5). Maimon concurred with this sentiment, suggesting that 'contents considerations should always have priority over other considerations' (Maimon, 1978, p. 262). Nowhere is this more important than in the context of criminal actions, the uncertainties of which were outlined above. The effects of choosing different coefficients for the Kirby data set on child abuse was considered in Empirical Study 8.1.

Empirical Study 8.1: The Relative Impact of Similarity Measure on Scaling Substantive Data

It has been suggested that the choice of similarity measure is an important structural issue. However, the relative impact of the similarity measure compared to representational issues of dimensionality and local monotonicity using criminal actions data is unknown. Therefore the Child Abuse Data Set was taken and considered in relation to each of four measures: Jaccard's, Russell and Rao, Yule's and Pearson's. These coefficients were selected to represent a number of possible interpretations of the data, ranging from weak and accepting of uncertainty (Jaccard's) to strong and assuming bivariate normality (ϕ, dichotomous PMCC). Each of these four coefficients was used to created four different triangular one-mode two-way input matrices using the same 36 variables.

Each coefficient gave (36 x (36 - 1)) / 2 = 630 values. Descriptive statistics for the four similarity measures on the 630 values are given in Table 8.6.

Coefficient	Mean	Std Dev	Minimum	Maximum
Jaccard's	18	13	00	67
Russell and Rao	09	08	00	38
Yule's Q	05	51	-100	100
Pearson's	04	21	-98	71

Table 8.6. Descriptive statistics for associations and correlations in Kirby data
Values are multiplied by 100 for clarity

As can be seen from this table, the Jaccard's and Russell and Rao coefficients that diminish the importance of the *d* cell were of lower variability than the others. However, the standard deviation scores must be taken into consideration with the overall distribution and range of values. The Jaccard's and Russell and Rao coefficients are bound between 0 and 1, though notably neither reaches near the upper bound at all and Russell and Rao is distinctly skewed towards the lower bound.

Yule's Q (i.e. dichotomous Guttman's weak μ) was the only coefficient to achieve its full distribution. It had an appropriately high standard deviation that suggests a normal distribution of scores, with 95% of scores being within two deviations of the mean of nearly 0. As Levy and Guttman (1975; cited by Maimon, 1978, p. 263) stated, this is because marginal distributions vary among items and Pearson's cannot cope with this, unlike Guttman's weak monotonicity. Pearson's is more strict in terms of the general formulae for monotonicity given by Guttman (1986), as is the monotonicity coefficient used in SSA-I namely semi-strong monotonicity. However, since SSA-I employs a non-metric transformation then unequal marginal distributions are equalised (Guttman, 1968, p. 481).

To provide a comparison for assessing the relative impact and importance of the similarity measure, two and three dimensional solutions were sought with local and global monotonicity. Each coefficient therefore was used on four different runs, namely 2D local, 2D global, 3D local and 3D global. In total, then, there were 16 SSA-I solutions. Table 8.7 shows the alienation values for the 16 SSA-I runs.

	2D		3D	
Coefficient	**Local**	**Global**	**Local**	**Global**
Jaccard's	299	212	192	147
Yule's *Q*	259	239	178	158
Russell and Rao	242	196	155	134
Pearson's	237	228	166	152

Table 8.7. Alienation values for SSA-I on Kirby data
Values for Coefficient of Alienation are without decimal point for clarity, i.e. multiplied by 1000.

As can be seen from Table 8.7, the lowest alienation solution tended to be the Russell and Rao and the solutions of highest alienation were the Jaccard's with local monotonicity and Yule's *Q* with global monotonicity.

For the second part of Empirical Study 8.1, the triangular association or correlation (i.e. input) matrices for the four coefficients were taken and listed as one column each in a rectangular two-mode two-way matrix. This contained four columns and had (36 x (36 - 1)) / 2 = 630 rows, since there were 36 items included in this analysis.

To this matrix were also added the triangular derived distance matrices (i.e. output) for each of the four coefficients, using both two-dimensional and three-dimensional solutions with local and global monotonicity. There were therefore 4x2x2=16 distance matrices added as columns to the original similarity values for the four coefficient association matrices. This made a new rectangular matrix with 20 column variables and 630 rows, with the rows corresponding to the similarity and distance information for each profile pair. Table 8.8 illustrates some of this matrix.

Point pair	Jaccard simil'ies	Jac'd 2D local distance	Jac'd 2D global distance	Jac'd 3D local distance	.	etc.	.	P'son 3D local distance	P'son 3D global distance
2,1	55	31	25	24	.	.	.	30	14
3,1	24	98	49	93	.	.	.	157	130
3,2	21	95	59	81	.	.	.	149	135
4,1	36	65	31	73	.	.	.	85	104
4,2	35	37	42	52	.	.	.	86	96
4,3	13	84	79	99	.	.	.	191	188
.	.	.	.	.	.	.	.	.	.
etc.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
36,34	08	33	58	82	.	.	.	173	179
36,35	11	41	65	77	.	.	.	151	156

Table 8.8. Associations and distances in SSA-I runs on Kirby data
Values are without decimal points for clarity

The 20 column variables of this matrix were correlated together using Pearson's PMCC. This was an appropriate coefficient given the approximately the normal distribution for the four similarity measures. Therefore a new triangular correlation matrix was created containing information on all the original input coefficients and the output distances for the solutions in the above table. The absolute values of the this triangular matrix were taken, namely removing any negative correlations. This was necessary since any similarity column would naturally correlate highly but negatively with any distance. In other words, the new triangular matrix was a summary of the similarity among the various coefficients and the distance matrices derived from them. The distance matrices further varied in dimensionality and local or global montonicity.

The values obtained were all positive and high, with no value less than 0.400. This meant that no association matrix or distance matrix was independent or uncorrelated with any other. The smallest value was 0.401, which was the correlation between distances in two dimensions using Yule's Q with local monotonicity, and distances in two dimensions using Russell and Rao with local monotonicity. The largest was 0.989, the correlation between distances in two dimensions using Pearson's with local monotonicity, and distances in two dimensions using Pearson's with global monotonicity.

All these correlations between and among the similarities and distances were statistically significant at the $p = 0.001$ level. However, the r values were still lower than those obtained by Maimon (1978), who correlated the association matrices from attitude questionnaire data using 5 different coefficients on 3 different response ranges, obtaining the lowest correlation of 0.88. The values of the correlation obtained in the present study were typically larger though than those of Marsden and Laumann (1978) on account of two reasons. Firstly, two of the four similarity measures of Marsden and Laumann were of 'social distance', with the other two being Yule's and Pearson's. Secondly, one-dimensional solutions were also sought by Marsden and Laumann, but these would have been misleading due to the inadequacy of scaling in one dimension the Kirby data with its high number of variables, namely 36.

The triangular matrix of correlations was analysed using SSA-I with both local and global monotonicity, and +5 and -5 weightings with FSSA. The same regional interpretation was possible with all solutions, indicating that local monotonicity was not an issue in these data because the alienation of the solutions were all roughly equal to or less than 0.05. Therefore the global monotonicity SSA-I solution with alienation of 0.05 in two dimensions used to display the results in Figure 8.2.

In Figure 8.2, the distance between two points refers to how highly correlated the points are. But in the case of distance from the original association or correlation matrix to any solution derived from *that original matrix* is also a measure of the alienation of that particular solution. This is because the formula for the Coefficient of Alienation is a squared correlation between the association matrix and the distance matrix which is subtracted from one to create a function to minimise. Therefore a perfectly fitting empirical SSA-I solution with zero alienation would be represented in the SSA in Figure 8.2 as being on top of the original similarity measure. Equally, a completely imperfect solution would be on the opposite side of the SSA-I in Figure 8.2 to the original similarity measure.

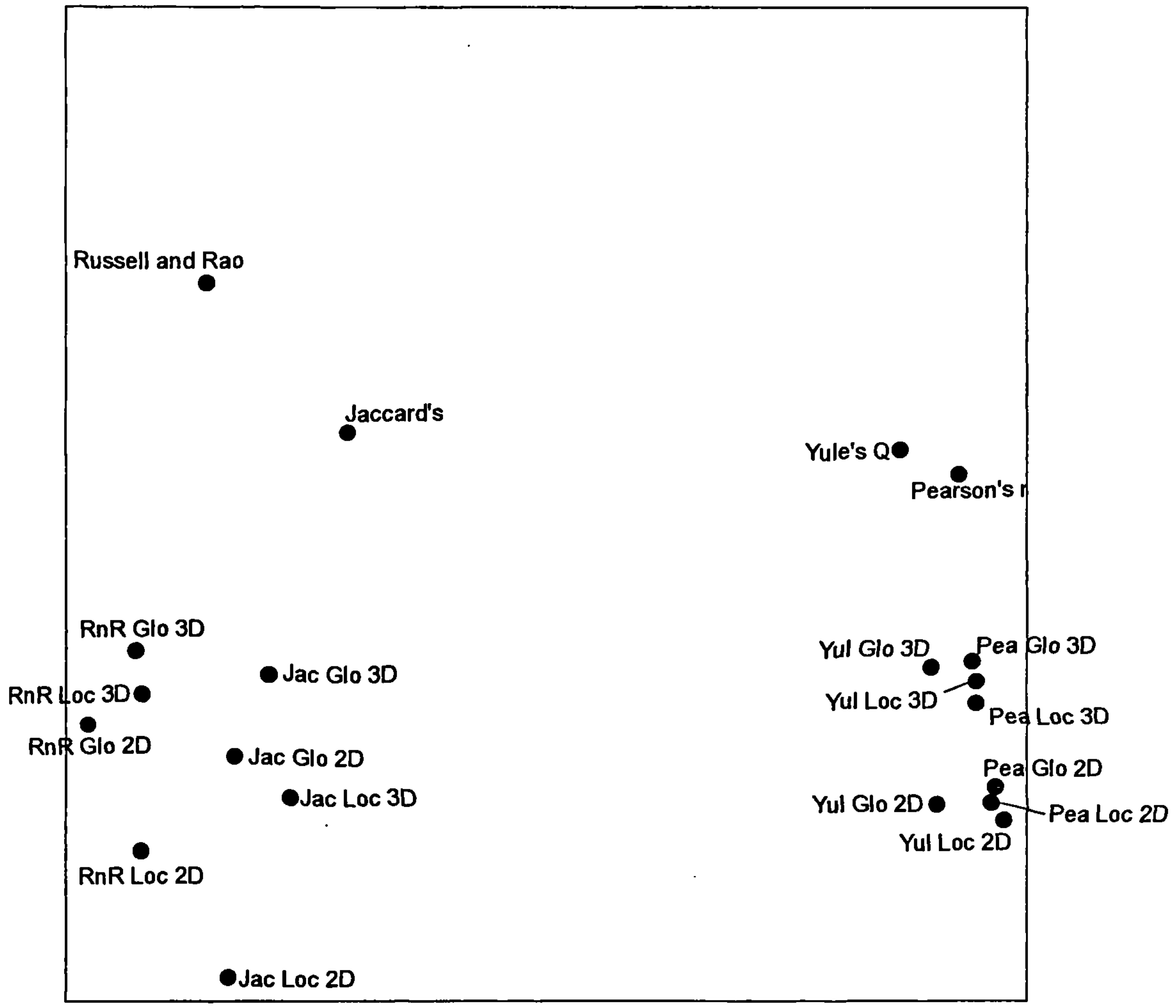


Figure 8.2. SSA-I of association, correlation and distance matrices for Kirby data

The first regional structure noticeable in this plot is that the three-dimensional solutions are closer to the original matrix than the two-dimensional solutions for all four of the coefficients. These were partitioned in the space as in Figure 8.3, where exceptions to the regional interpretation are marked with an arrow.

The three dimensional solutions are closer to the original coefficients than the two dimensional, as would be expected from the better alienation values shown in Table 8.7. However, there is some mixing of dimensionalities in Jaccard's and Russell and Rao plots, with the three dimensional Jaccard's with local monotonicity being too far away, as indicated by the arrow in the plot. However, what is important to note is that the two and three dimensional solutions are closer to each other than they are to the original similarities themselves. Thus there is are greater correspondence between the output distance matrices in the two dimensionalities than the original input matrix.

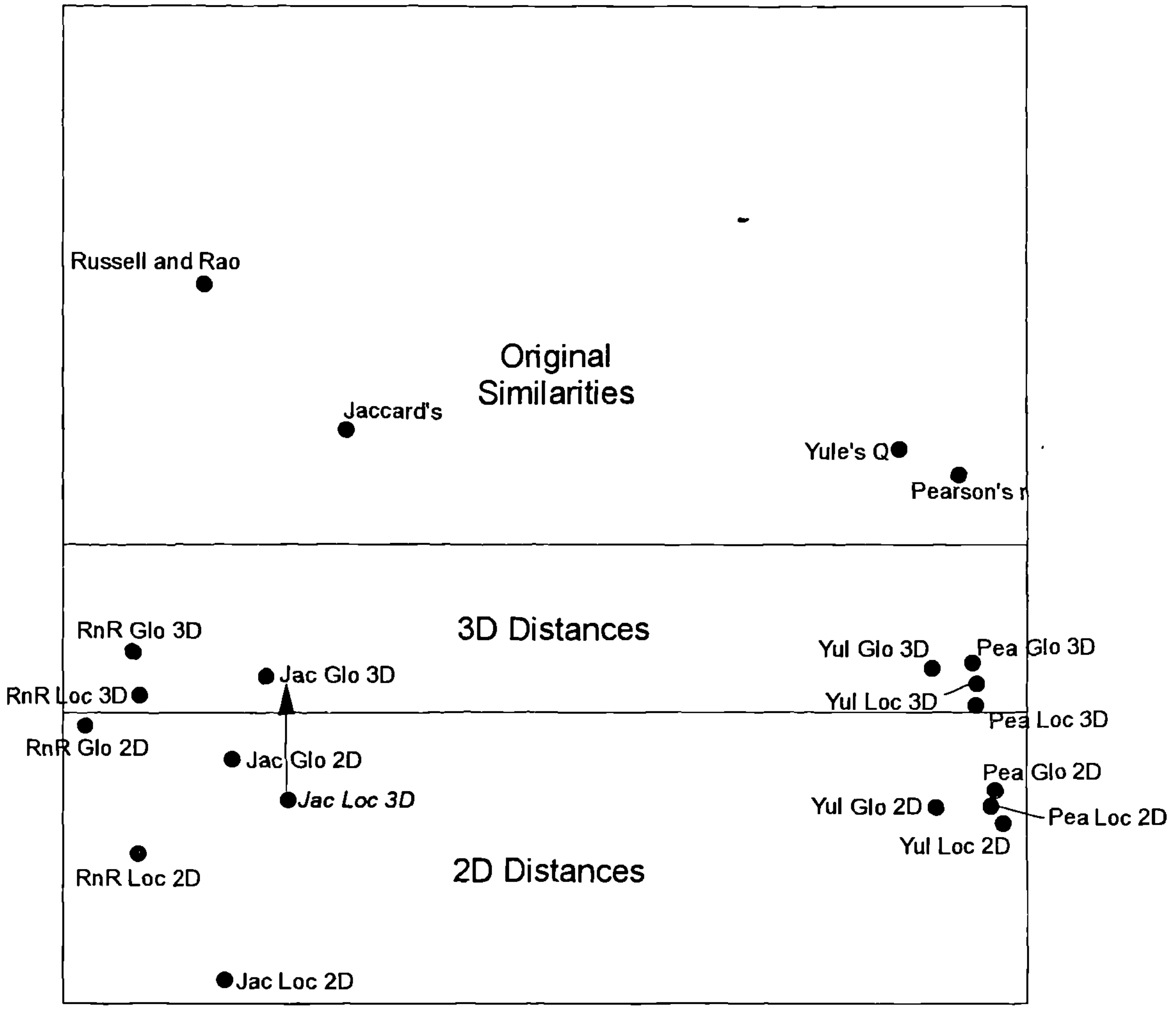


Figure 8.3. Partition of SSA-1 into similarities and distances

The regional structure in Figure 8.3 can be further refined by partitioning on the type of monotonicity, namely global or local. This is shown in Figure 8.4. As can be seen from this plot, the partitions show that generally the global monotonicity distance matrices are closer to the original than the local monotonicity distance matrices. This is a further demonstration that local weightings for the sake of substantive or methodological reasons create higher alienation values, as was discussed in the previous chapter.

Figure 8.4 also proves what was asserted in chapter 7, that the impact of local monotonicity will be greater on those coefficients that are bounded between 0 and 1, namely the Jaccard's and Russell and Rao. This is shown by the greater distance between the local and global solutions in the same dimensionality for these two

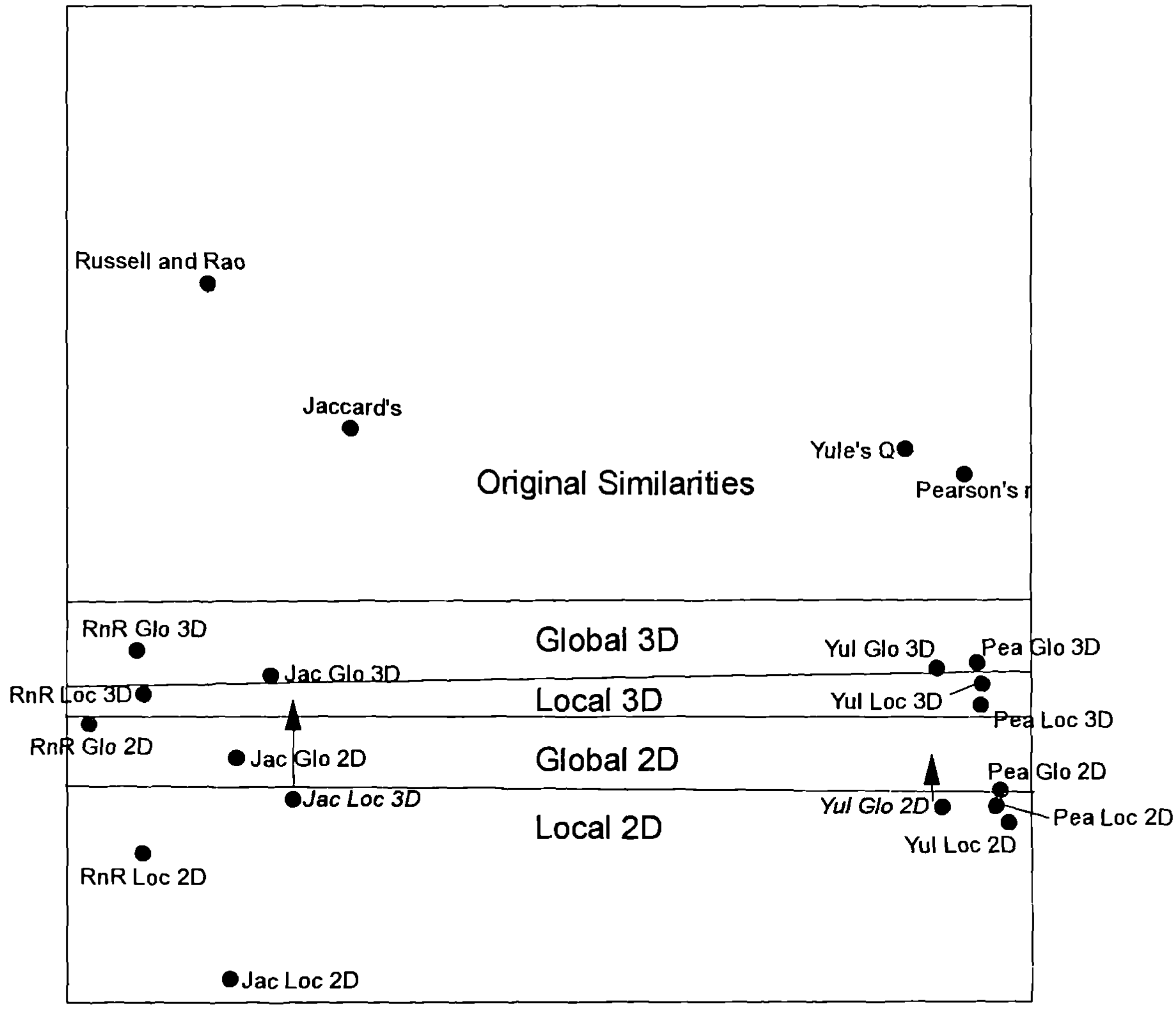


Figure 8.4. Partition of SSA-I into similarities and global and local dimensionalities

coefficients, more so than for Yule's Q and Pearson's. Table 8.9 calculates the difference between local and global values within the same dimensionality.

Coefficient	Local minus Global 2D	Local minus Global 3D	Average
Jaccard's	087	045	066
Russell and Rao	046	021	034
Yule's Q	020	020	020
Pearson's	090	014	012

Table 8.9. Difference between local and global alienation values in SSA-I of Kirby data
Values for Coefficient of Alienation are without decimal point for clarity, i.e. multiplied by 1000.

The SSA-I of the solutions in Figure 8.2 can be used to further indicate that the different coefficients and their corresponding outputs occupy distinct regions of the space in a facet of 'Coefficient type'. The representation can therefore be partitioned according to the coefficient used also, which is displayed in Figure 8.5 again with the exceptions marked by arrows.

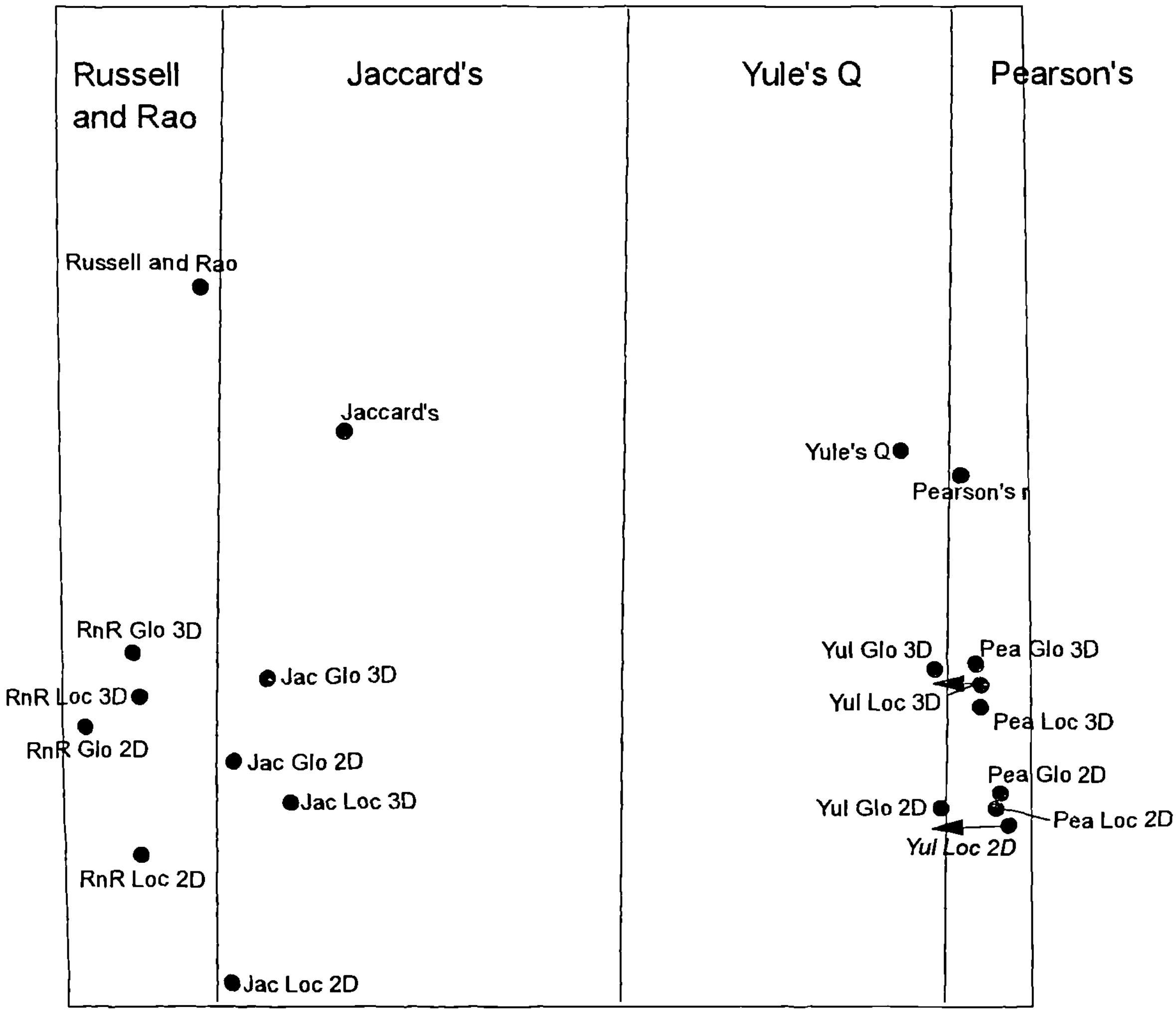


Figure 8.5. Partition of SSA-I into coefficient types

However, it is interesting to note the close proximity between Yule's Q and Pearson's correlation matrices and distance matrices. This would be expected since the formulae are almost equivalent, though the Pearson's measure is limited in its distribution.

What is more important is that there is a clear partition vertically down the middle of the plot into two 'clusters', namely Jaccard's and Russell and Rao on one side and Yule's Q and Pearson's on the other. It is hypothesised that this split is characterised by the role of the *d* value (i.e. joint non-occurrence) from Table 8.1 in the formulae for the coefficients. Yule's *Q* and Pearson's *d* values have a highly significant multiplicative role, whereas the Russell and Rao *d* value has a lower significance in its additive role and Jaccard's does not use the *d* value at all.

It was noted earlier that the two and three dimensional solutions are closer to each other than the original similarities themselves. This means that the choice of the coefficient used in the association or correlation matrix is more important than the choice of dimensionality. Most of the existing literature on scaling has focused mainly on the impact of deciding the 'correct' dimensionality, as was shown in chapter 6 with the Monte Carlo studies (e.g. Wagenaar and Padmos, 1971; Spence and Graef, 1974). However, this present Empirical Study shows that the choice of association or correlation measure is at least as important if not more so.

Integrating all this information, it can be seen that the relative impact of the similarity measure exceeds that of dimensionality and monotonicity with these data in terms of computational correspondence between association or correlation matrix and the distance matrix. This is because the coefficient used to measure similarity is a structural issue, assessing the nature and relationship between the items in the content universe before they are even put into a geometric representation, where the issues of dimensionality and local monotonicity are important.

The importance of conjoint absence is such that where there is uncertainty in the data, the choice of a measure of low emphasis on the *d* cell is fundamental. The amount of variability and difference is surprisingly small between Jaccard's and Russell and Rao, and between Yule's and Pearson's, both in terms of similarities the solutions that the two groups produce. The difference between Yule's and Pearson's is especially small.

Of all the coefficients, the importance of local of monotonicity is notable especially for the Jaccard's index, and also the Russell and Rao measure. The reason for this is that in a Jaccard's matrix derived from criminal actions information - such as this data set - there tends to be only one possibility for a tied value. This is the value of 0, the lower bound of the Jaccard's range. The primary untying approach means that these tied 0 values may be optimally broken to improve fit. However, in local monotonicity these values are weighted against since higher values are improved first. This means that the optimal breaking of these ties for a better is diminished and hence the resulting alienation score is markedly higher. As was seen in Table 8.8, this is less notable for Yule's and Pearson's which have a range going to -1 i.e. they are not truncated at 0. Therefore the choice of local monotonicity for the faceted approach with regional interpretation using the Jaccard's index is especially significant.

Summary of Chapter 8

Various dichotomous measures to associate or correlate items were considered. Their treatment of conjoint absence was examined in detail. The Child Abuse Data Set was introduced as an example of an extremely noisy data set. The error or 'uncertainty ratio' in the data were shown to vary according to the items. It was suggested that the Jaccard's index was the best for the Kirby data, especially when compared to strict parametric measures. The effects of choice of coefficient, dimensionality and local weighting of monotonicity were investigated in Empirical Study 8.1. Contrary to the focus of the literature, dimensionality was found to be less important than the choice of coefficient for this data set. The interaction of local weighting of monotonicity and the Jaccard's index was reiterated.

Chapter 9
exclusivities and testing for structure

The Similarity Measure and the Spatial Configuration of Frequencies

The previous chapter suggested that one of the most important decisions for scaling criminal actions is the choice of coefficient. It was suggested that the Jaccard's index may be particularly suitable where data were extremely noisy, such as the Kirby Child Abuse Data Set. These were contrasted in uncertainty from the FBI data set, which were intended by Ressler *et al.* (1992) to be definitional for the homicide types they were describing. This meant that the presence or absence of the items was more certain, even if there was repeated use of qualifiers such as 'sometimes' or 'usually'. This chapter examines how the nature of the items themselves influences the applicability and suitability of the Jaccard's index for the Child Abuse Data Set. Different types of 'exclusivities' are suggested to influence the representation and hence the testing of structure.

There were many similarities between the Child Abuse Data Set of Kirby (1993) and Canter and Heritage (1990) on sexual assaults against adult women. The source of the data in Kirby (1993) and Canter and Heritage was noted in Kirby as similar in uncertainty and proneness to error. Both sets of data were derived from partial content universes, namely information recorded in police case files containing victim statements and offender records. The structural hypotheses for these adult sexual offence data in Canter and Heritage was also that offences could be empirically classified into distinct regional themes, each of which could be supported by the literature.

Since Kirby (1993), Canter and Heritage (1990) and Canter and Fritzon (1998) used behaviours that were dichotomously coded, the frequencies of each behaviour shows the proportion of offences where the behaviour occurred. In Kirby the frequencies associated with the offence behaviours were found to have distinct contours, as represented schematically in Figure 9.1.

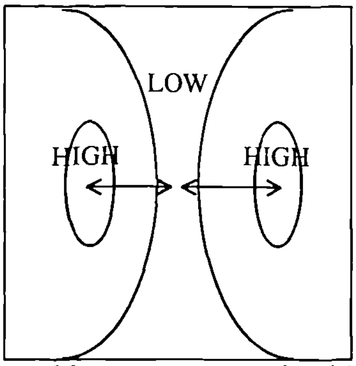


Figure 9.1. Schematic representation of frequency contours in Kirby (1993)

By contrast, in Canter and Heritage (1990) and Canter and Fritzon (1998) there was one single central area of high frequency behaviours which are associated with the *sine non qua* or 'core' of the content universe. This is represented schematically in Figure 9.2.

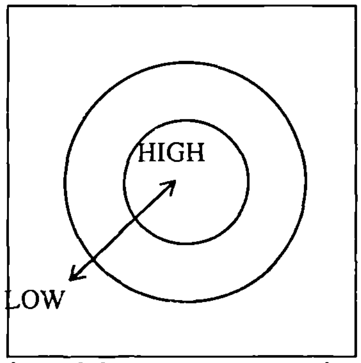


Figure 9.2. Schematic representation of frequency contours in Canter and Heritage (1990) and Canter and Fritzon (1998)

Figure 9.2 shows the shape of a single circular representation, as if in a modular facet, rather than a partial double circular representation. Kirby commented on the lack of the Canter and Heritage (1990) spatial representation of frequencies in offences against children, saying that:

> The fact that there are no variables conceptually central to child molestation as there are with rape [i.e. Canter and Heritage, 1990] illustrate the complexity of child molestation ... [which] is not as easily defined and includes a number of acts on both genders. (Kirby, 1993, p. 184)

The analysis performed on these data was cited by Kirby (1993, p. 183) as being the Guttman-Lingoes SSA-I. However, according to the results print out (Kirby, 1993, appendix C, pp. 318-321) the analysis actually used was Shye's FSSA. This use of FSSA was later confirmed by Kirby (personal communication). This in itself does not present any difficulties. What does become problematic is the fact that the coefficient used to create similarity matrices from two-mode two-way matrices in FSSA is Guttman's coefficient of weak monotonicity, equivalent dichotomously to Yule's Q. From the argument in chapter 8 about the uncertainty of 'unrecorded' data and the importance of the d cell of conjoint absence (see Table 8.1; also Formulae 8.1 and 8.4), it is argued that a more appropriate coefficient for the analysis of the Kirby data would be Jaccard's.

It is therefore hypothesised that the crucial difference in the type of coefficient will change the representation of the frequencies with the Kirby data from that of Figure 9.1 to that of Figure 9.2. This was investigated in Empirical Study 9.1.

Empirical Study 9.1: Frequency Contours in the Kirby Data with Jaccard's

The test of this hypothesis was performed in two parts. In the first part, the same subset of variables were taken as were used by Kirby (1993) in the original study. This gave a two-mode two-way matrix of 36 variables on 97 offences. The variables were associated with the Jaccard's index and then put into FSSA using a default local monotonicity weighting of +2 in two dimensions. In other words, the original study of Kirby (1993, appendix C, pp. 318-321) was replicated except for the association coefficient being Jaccard's rather than Yule's Q (dichotomous Guttman μ), the default association coefficient in FSSA.

Figure 9.3 shows the plot produced in this analysis. This solution produced an alienation of 0.23, which was higher than with the alienation of 0.17 using Yule's Q as found by Kirby. This was consistent with Table 8.7 in the previous chapter which showed that the use of Jaccard's gave higher alienation than Yule's Q, though with a different subset of variables.

In order to test the hypothesis of concentric frequency contours around a single point, as schematically represented in Figure 9.2, the frequencies associated with each behaviour were plotted as the points in Figure 9.4.

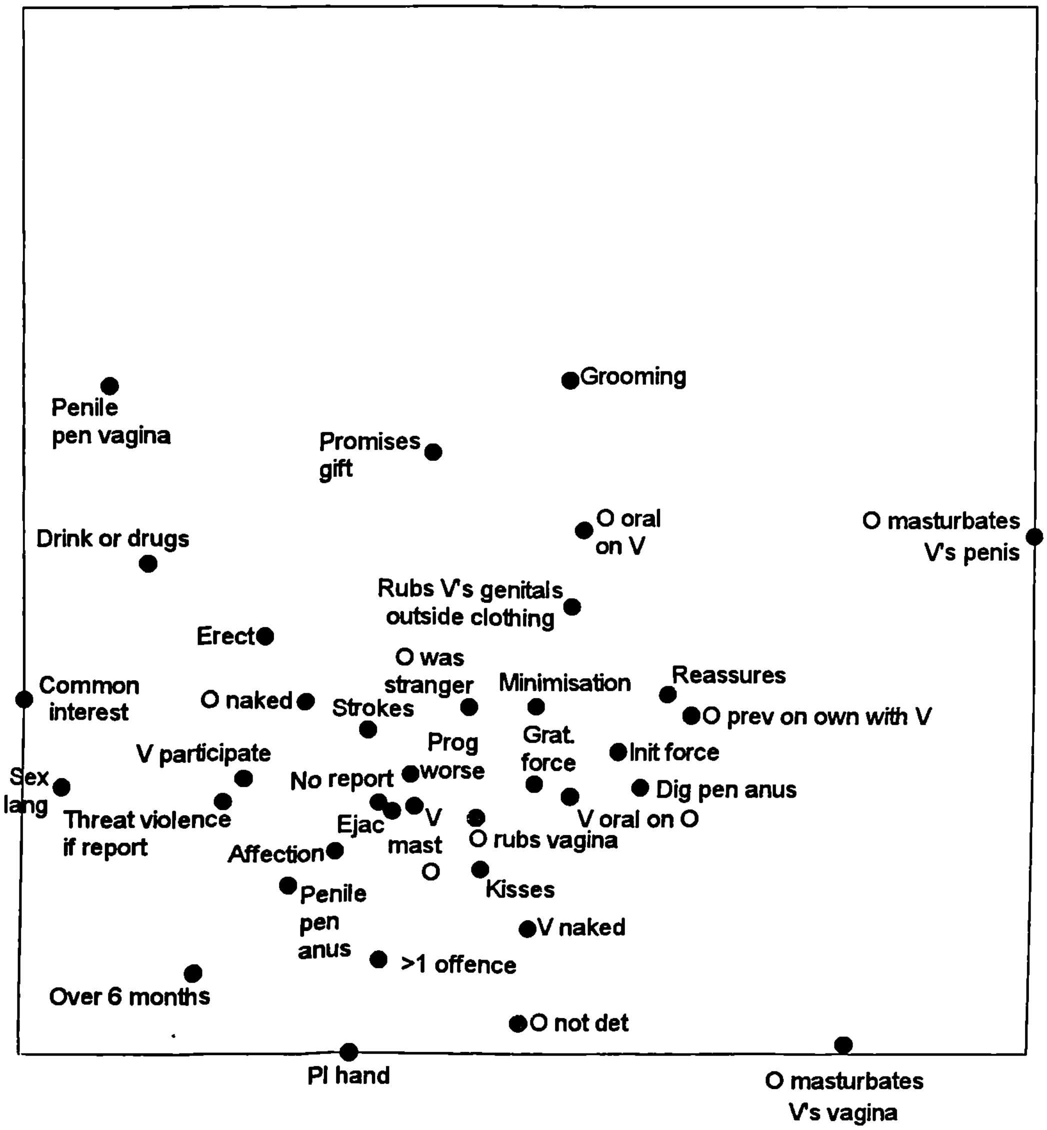


Figure 9.3. FSSA of Kirby (1993) variables using Jaccard's rather than Yule's Q (dichotomous Guttman μ)

The spatial configuration of these frequencies cannot be partitioned in any way which is similar to the contours in Canter and Heritage (1990) and Canter and Fritzon (1998) as hypothesised and represented schematically in Figure 9.2. Any contours that were to be plotted in Figure 9.4 would be far less regular than was hypothesised. In fact, several frequency values were found in places lacking contiguity with the surrounding space, i.e. the sub-space contained frequencies of both high and low values. This was also found in the three dimensional solution.

The second part of the Empirical Study was to use the same variables as in the original Kirby (1993) study with the Jaccard's measure but using the SSA-I program instead of FSSA. In other words, this was a replication of the first part of the

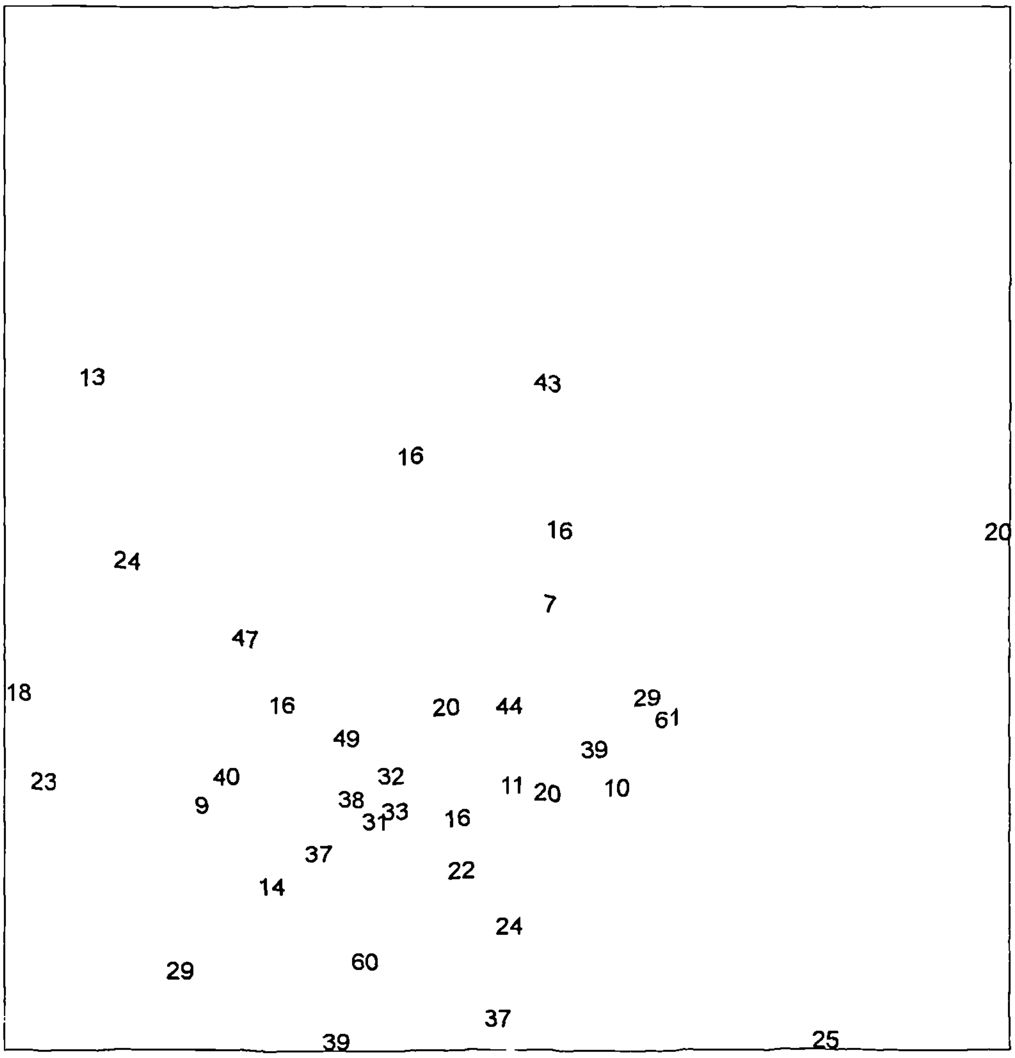


Figure 9.4. Frequencies of Kirby (1993) variables from FSSA replication

Empirical Study but for the different SSA program. The rationale behind this was that Canter and Heritage (1990) and Canter and Fritzon (1998) had used SSA-I rather than FSSA, so it was necessary to eliminate the possibly confounding factor of the different SSA program.

As would be expected, the SSA-I plot was extremely similar to the FSSA plot though with a slightly lower Coefficient of Alienation of 0.22. This is commensurate with the finding in Empirical Study 7.2 in chapter 7 that the default FSSA local monotonicity weighting of +2 gives slightly higher alienation than the local monotonicity weighting in SSA-I. In the SSA-I replication with Jaccard's, the same discontinuities in frequency pattern were found again which ran counter to the schematic shape in Figure 9.2. Again the three dimensional solution was also unclear, with alienation 0.14.

From this Empirical Study, it seems that the use of the Jaccard's coefficient does not automatically imply that the frequencies will found in a concentric circular pattern, as if in a modular facet. This unexpected result is returned to later in this chapter and repeated in Empirical Study 9.5, after an alternative explanation is proposed in Empirical Studies 9.3 and 9.4 that several 'exclusivities' caused this effect.

However, there is another facet in the Kirby (1993) study that must also be investigated, which is addressed in Empirical Study 9.2.

Empirical Study 9.2: Axial Theme Facet in the Kirby Data with Jaccard's

Kirby (1993) examined the literature on sexual offences against children and concluded that three themes were repeatedly referred to in the literature. These were:
1. overt aggression and violence towards children,
2. attempts at intimacy or emotional gratification from children, and
3. obtaining sexual gratification from children.

Kirby sought and found evidence to support these structural hypotheses in his original data of 37 variables on 97 offences. Consequently, Kirby proposed regional hypotheses in an FSSA plot which is represented schematically in Figure 9.5.

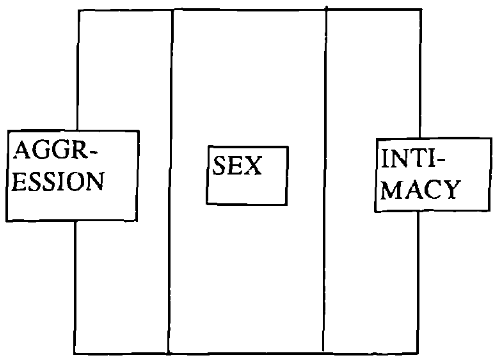


Figure 9.5. Schematic representation of axial theme facet in FSSA of Kirby (1993, p. 192)

Represented this way, this regional interpretation suggests that there is an axial facet consisting of behaviours in three elements of aggression, sex and intimacy. Thus each item in the plot was classified by Kirby into one of these three themes.

Using these themes, the revised FSSA in the present Empirical Study using Jaccard's (i.e. Figure 9.3) was also investigated. The items were classified into the same themes as were proposed by Kirby (1993). When the regions of the different items were partitioned according this original classification, it was found that the partitions were not as regular as would be expected given the clarity and strength of the structural hypotheses made by Kirby. The partitioning of Figure 9.3 according to Kirby's original theme classification is given in Figure 9.6.

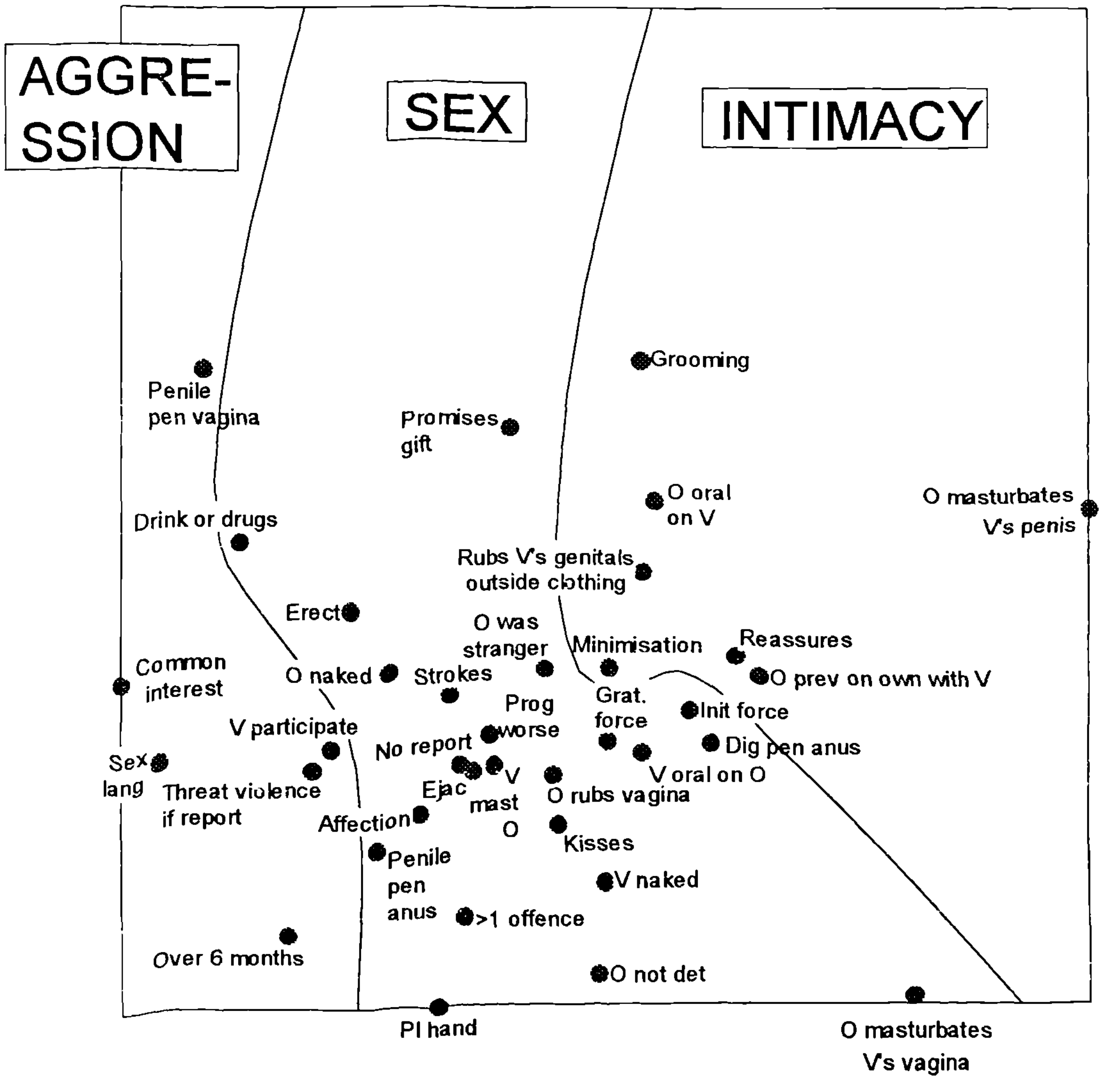


Figure 9.6. Partitioning of revised FSSA with Jaccard's according to Kirby (1993) theme classification

The partitioning of the revised FSSA with Jaccard's on the axial facet of offence theme was found to be irregular, just as the pattern of frequency contours had been in Empirical Study 9.1.

In summary, then, the Jaccard's reanalysis of the Kirby variables revealed two things. Firstly, the frequency contours were not as regular as those found in Canter and Heritage (1990), a data set which shared much in common with the Child Abuse Data Set, or Canter and Fritzon (1998). Secondly, Kirby's clear structural hypotheses of themes of aggression, sex and intimacy were not represented in the same way. An alternative explanation is proposed in the next section and Empirical Study 9.3 which follows it.

Logical and Substantive Exclusivities

The alternative explanation requires a close examination of the precise wording of the items used in the 36 variables chosen by Kirby (1993) and used above in the reanalyses with Jaccard's. The FSSA plot from Figure 9.3 in two dimensions using Jaccard's with the same items as Kirby is reproduced in Figure 9.7 with two pairs of items marked on it. The pairs indicated on this plot indicate variables which are in some sense *exclusive* to each other. For example, the test of whether or not the offender masturbated the victim is done with two variables:

> 40 Offender rubs outside of the victims' vagina but does not digitally penetrate.
>
> 41 Offender masturbates male victim. (Kirby, 1993, p. 311)

Clearly it is not logically possible for conjoint occurrence, since the sample of data contained offences against both males and females, and a child can only be of one sex. The crosstabulation in Table 9.1 illustrates this point.

N = 97		Variable 41	
		Yes	No
Variable 40	Yes	0	24
	No	19	54

Table 9.1. Crosstabulation between two variables taken from Kirby (1993, p. 311)

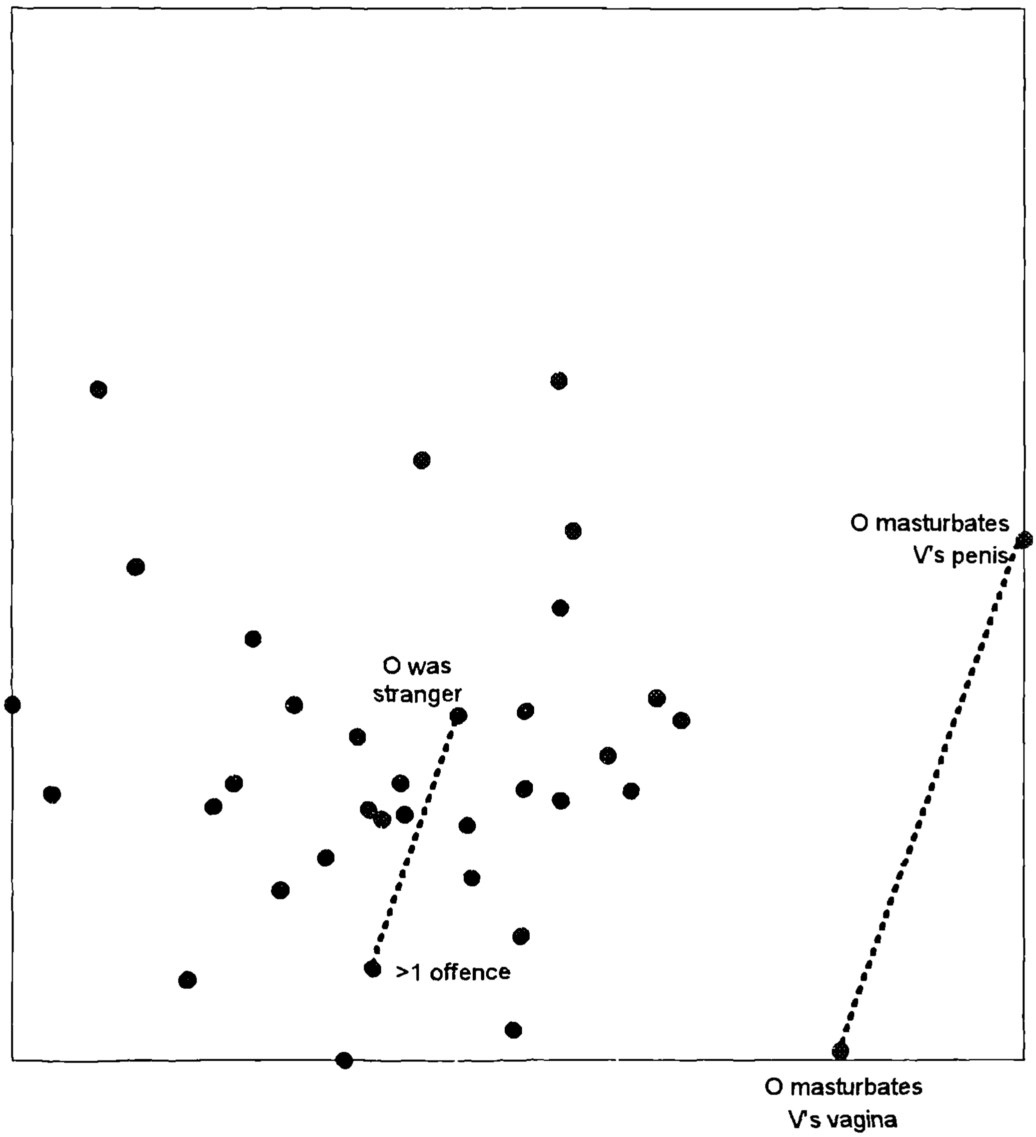


Figure 9.7. Two item pairs on revised FSSA with Jaccard's on Kirby data

The intention of the action was the same for both sexes in that the offender is sexually manipulating the child's genitals, though the behaviours are different on account of the gender of the victim. The value of Yule's Q (see Formula 8.4) for these two 'masturbation of victim' variables is -1 since one of the cells is empty, and Jaccard's equals 0 since the a cell (see Table 8.5) equals zero.

By itself, such an exclusivity and the consequent association or correlation value is not disastrous. However, there are several other gender-specific variables such as 'offender digitally penetrates victim's vagina' and 'offender performs fellatio on victim'. Where the number of exclusivities increases, the SSA program must try to reproduce order in the plot which is logically untenable, threatening the continuous space imagery of the solution. These variables are *logically exclusive*.

For example, masturbation of the victim may be highly associated with a strategy of minimisation or desensitisation to lesser sexual contact before moving onto more explicit and stronger sexual contact. Suppose this were true irrespective of the victim gender, yet the masturbation variable was split into the two genders. An SSA would therefore attempt put the minimisation variable close to male masturbation and also close to female masturbation; yet at the same time, it would also attempt to separate male masturbation and female masturbation. To achieve this it would have to put minimisation at the centre of a line between male masturbation and female masturbation, a line artefacutally created and stressed in its local bonding (as it did with the local two-dimensional local SSA-I). If the SSA were global, this stress would be even more significant.

Another sort of exclusivity exists in the data, as shown in the plot above. This concerns what can be called *substantively exclusive*. Here, though it is logically possible for conjoint presence to occur, the meaning of the variables makes this impossible. An example of this is the crosstabulation of the variables 'Offender was a stranger' and 'Offender committed more than one offence against the same victim', as shown in Table 9.2.

N = 97		O is stranger	
		Yes	No
More than 1 offence	Yes	0	58
	No	19	20

Table 9.2. Crosstabulation between 'Offender is a stranger' and 'More than 1 offence by offender on victim' from Kirby data

As with the logical exclusivity, the value of Yule's Q for these two variables is -1 since one of the cells is empty, and Jaccard's equals 0 since the a cell equals zero. In this example *possible* that the offender will be a stranger to the child and he will offend against that child on more than one occasion, but substantively this does not seem feasible since the child will surely became acquainted with the offender over the course of a series of offences such that he is no longer a 'stranger'. Again, these exclusivities will influence the plot adversely in a similar way to that described above.

The reduction of logical and substantive exclusivities is a more straightforward task with fully faceted designs with Mapping Sentences, and with secondary analysis of data using Mapping Sentences. However, when looking at partial content universe - as with Kirby (1993) - it is sometimes harder to achieve this and consideration should therefore be given to exclusivity reduction. Where work is exploratory or the first of its kind, this becomes a particular challenge as can be seen with the Kirby data.

Therefore it is suggested that there is an alternative reason for not finding Contiguity between the structural hypotheses of Kirby and the representations. This is that error was unnecessarily added into the analysis by the choice of mutually and substantively exclusive items which partially concealed the structure. Empirical Study 9.3 investigates this line of reasoning.

Empirical Study 9.3: Regional Interpretation and the Violation of Content Analytical Categories

To test this assertion to its extreme, a subset of items was needed from the Kirby Child Abuse Data Set that *deliberately* violated the content analysis and fully faceted item construction guidelines. Therefore each of the full set of variables in the original data set was examined for any possible exclusivities with any other variable. The three type of exclusivity were:

1. logical exclusivity: both items cannot be true (present) for logical reasons; e.g. 'Offender masturbates female victim' and 'Offender masturbates male victim' - victim cannot be male and female.
2. substantive exclusivity: both items cannot be true (present) for logical reasons; e.g. 'Offence spontaneous' and 'Con approach' - an offence cannot be spontaneous if the child has been lured into some situation.
3. mutual exclusivity: if one item is true (present) the other must be false (absent) either one or the other, e.g. 'Offence inside' and 'Offence outside' - offence took place in one or the other.

Those variables that had any of these conditions of exclusivity were then placed into a subset of the original data set to create a new matrix consisting of 24 variables by the 97 offences. The details of this data set are given in Table 9.3.

SSA Label	%	Full Description
>1 offence	60	Offender commits more than 1 offence against child*
Con appr	54	Offender cons child into going elsewhere
Digital pen vagina	25	Digital penetration of vagina*
First on own	19	Child on own when offender first meets
First with others	49	Child with others when offender first meets
O deterred	37	Offender deterred by adverse reaction
O gives alias	5	Offender gives victim an alias
O masturbates female V	16	Offender masturbates female but digital penetration*
O masturbates male V	20	Offender masturbates male victim*
O not deterred	27	Offender not deterred by adverse reaction or *
O prev spoken to V	32	Offender spoke to victim several times before offence
O stranger	20	Offender is stranger to victim*
Offence inside	70	Offence inside
Offence on own	71	Child on own during offence
Offence outside	30	Offence outside
Offence spontaneous	45	Offence is unplanned
Offence with others	29	Child with others during offence
Offender carer	45	Offender was caring for the child
Over 6 months	29	Offences against child for at least 6 months*
Penile pen vagina	13	Actual or attempted penile penetration of vagina*
Prog worse	32	Offences become progressively worse during series*
V female	67	Victim female
V knows O well	44	Victim knows offender well
V male	33	Victim male

Table 9.3. Details of exclusivities subset of Kirby data
*indicates variable was originally included by Kirby (1993);'O' refers to offender and 'V' refers to victim

The 24 variables were put into SSA-I with Jaccard's using local monotonicity in two dimensions to investigate how the deliberate violation of item design would influence the structure. The plot is reproduced in Figure 9.8. This solution gave an alienation of 0.27, higher than in Empirical Study 9.1 with either FSSA (0.23) and SSA-I (0.22) which had 36 variables. Even though this particular subset of data would have been inappropriate for theory building purposes, the alienation was not so high as to instantly indicate that the data were no different from random data.

In the solution represented in Figure 9.8 the structural hypotheses of Intimacy, Aggression and their associated sexual behaviours clearly cannot readily be determined. Towards the bottom and bottom left of the plot there is some intention of

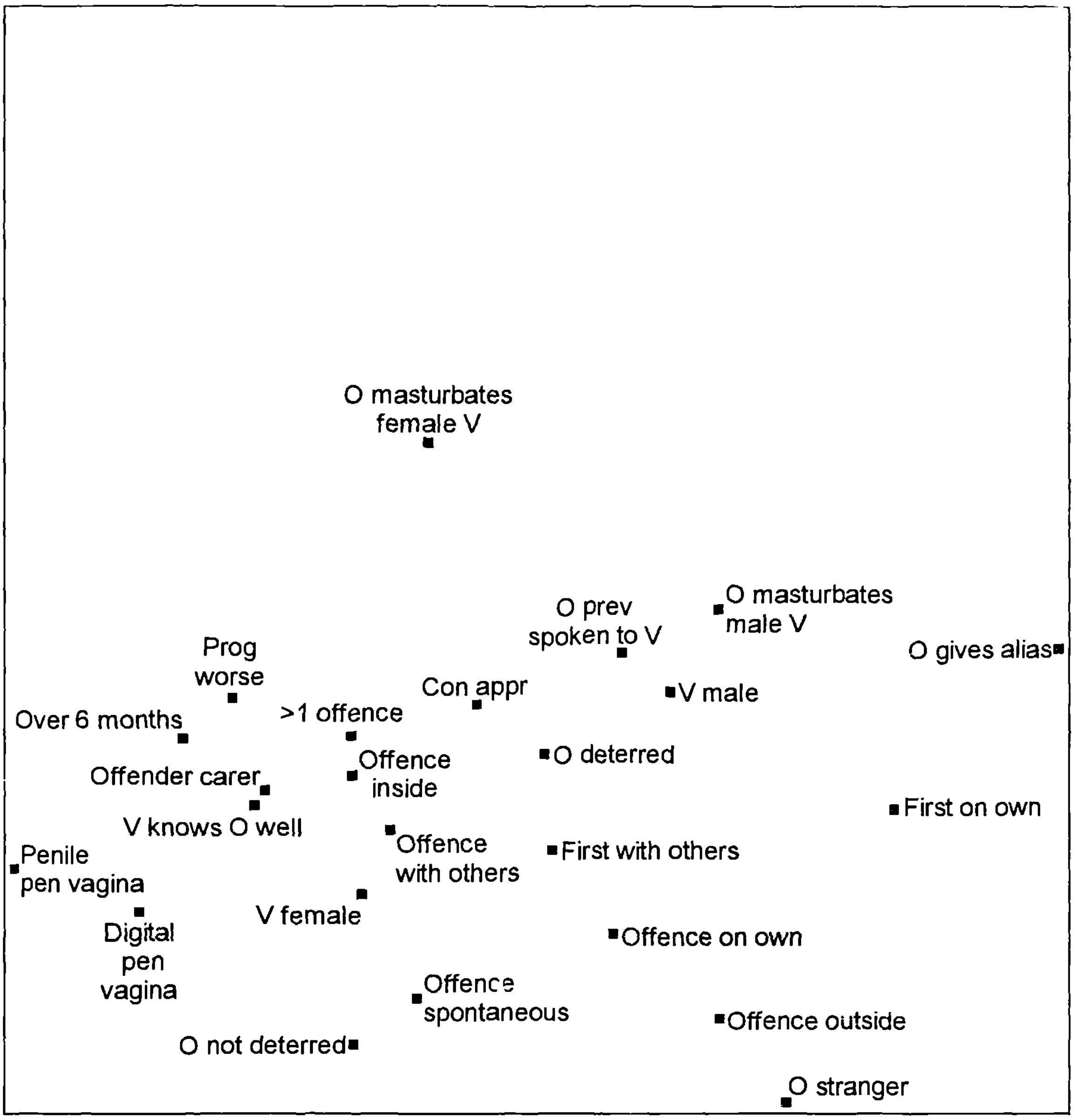


Figure 9.8. SSA-I of exclusivities subset of Kirby data

Aggression with aggressive sexual behaviours. In the left-centre of the plot implies some ongoing abuse in a relationship with the child, such as father or step-father with some element of 'giving' gratification to the child towards the centre and centre-right of the plot.

However, the exclusivities in the plot make this interpretation much harder. In particular, there are certain relationships in the plot which are very badly represented. For example, the masturbation of a female victim by the offender (i.e. intimate) is a long distance from the other female variables, including the digital and penile penetration of the female victim (i.e. aggressive), even though these are hypothesised to be a continuum. It is just as close to the male variables with which it has a lower association. Clearly, this variable has low local spatial bonding, and Figure 9.9 shows the highest three associations with the 'masturbation of a female victim by the offender' variable. In Figure 9.9, the solid line indicates the highest association, the dashed line indicating second highest and the dotted line indicating third highest, as in Figures 7.5 and 7.6 in chapter 7.

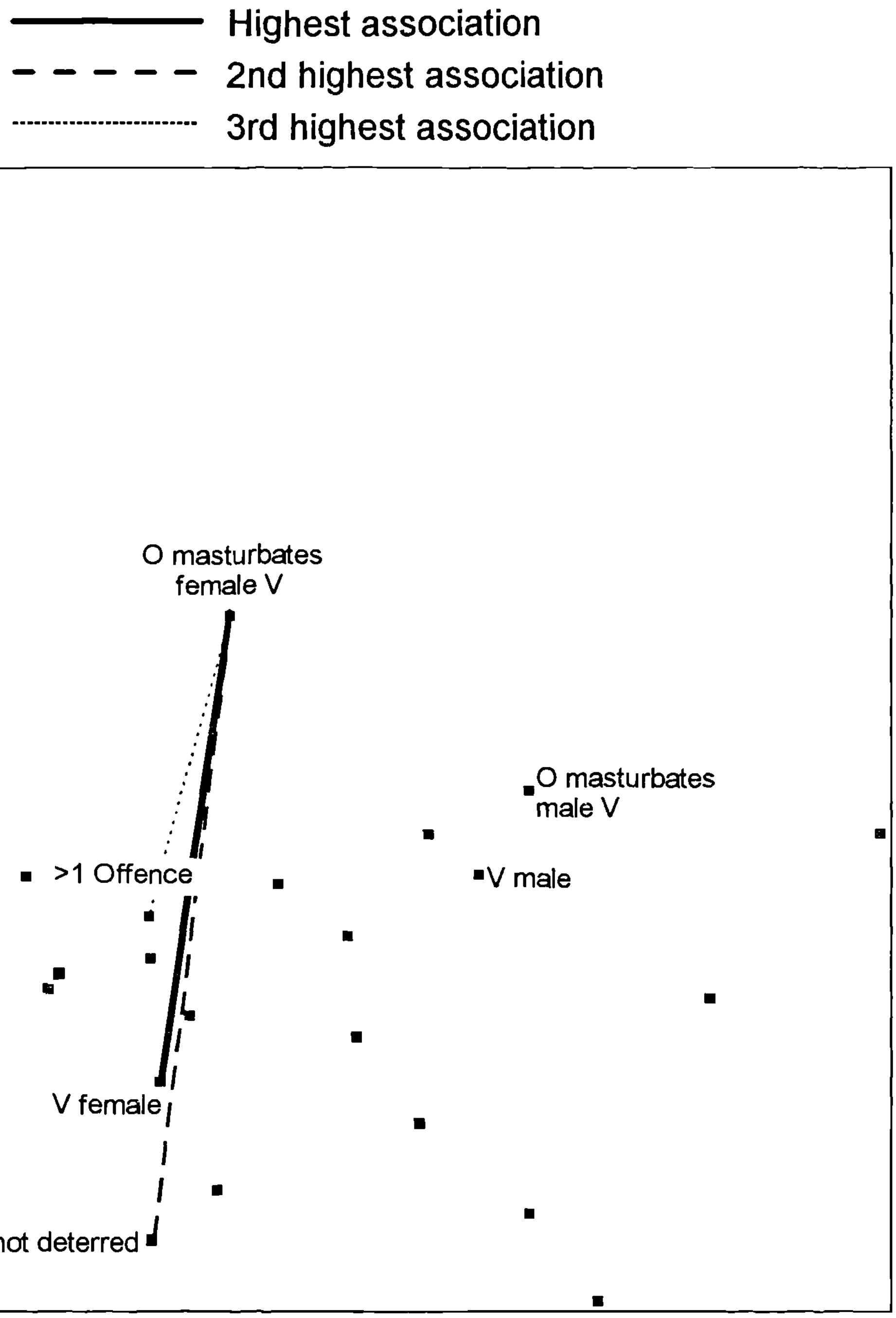


Figure 9.9. LSB in 'Offender masturbates female victim' in exclusivities subset of Kirby data

There were several other such poorly bonded variables, even with local monotonicity in the SSA-I. In fact, even the mutually exclusive variables 'victim male' and 'victim female' were quite close together in this plot.

It can be concluded therefore that the quality of the data collection about a content universe can be crucial to the success of testing for structural hypotheses. The use of deliberately poor items showed that structure can be concealed, which implies that the choice of good items should be able to reveal structure. This conclusion of no structure from this (deliberately) inadequate representation is a type II error i.e. False Negative (see chapter 6).

Empirical Study 9.4: Reanalysis of Kirby Data with Exclusivity Reduction

Having examined the set of items with exclusivities in the Kirby data on child abuse, it is possible to remove combinations of variables are logically or substantively exclusive. These can then be removed from the sample of possible variables from which to draw; in effect, the partial content universe becomes even more valid. The pool of variables that were not logically or substantively exclusive was one less than Kirby's original analysis, namely 35 variables. It was hypothesised that the structure would be stronger using these more strict variables with what was argued to be a more appropriate coefficient, the Jaccard's coefficient.

It was discovered that most of the logical exclusivities were to be found in the sexual gratification theme, since many of these variables were gender specific. If all or most of these were to be removed, there would be no way of testing Kirby's original structural hypotheses of aggression, intimacy and sex. Therefore it was necessary to devise new variables from existing ones. Three new variables were constructed in this way: 'digital penetration', 'penile penetration' and 'masturbation of victim'.

'Digital Penetration' was defined as present by seeing if either of the existing variables of 'Digital Penetration of Vagina' or 'Digital Penetration of Anus' was coded as present, using a Boolean OR function. Similarly, 'Penile Penetration' referred to either vagina or anus or both, and 'Masturbation of Victim' was of either male or female victim. Therefore, the sexual variables were diminished slightly in number, but strengthened in meaning.

The full details and frequencies of the variables are given in Table 9.4.

SSA Label	%	Full Description
Affection	37	Offender shows affection towards the child*
Common interest	18	Offender suggested to victim he had a common interest*
Con appr	54	Offender conned victim into going elsewhere
Con rpt	38	Offender uses non-violent threat if victim reports*
Dig pen	31	Offender digitally penetrates male or female victim
Drink or drugs	24	Offender was drunk or on drugs*
Ejac	31	Offender ejaculates*
Erect	47	Offender's penis was erect*
First with others	50	Victim approached by the offender in presence of others
Gratuitous force	11	Force or threat of force used gratuitously*
Groom	43	Grooming strategy used*
Init force	39	Offence facilitated through use of initial force*
Kisses	22	Offender kisses victim on lips*
Minimise	44	Minimisation strategy used*
O masturbates V	36	Offender masturbates male or female victim
O naked	16	Offender naked at time of offence*
O not det	27	Offender not deterred by adverse reaction *
O oral on V	16	Offender performs oral sex on victim*
Offence outside	30	Offence outside
Offence spontaneous	45	Offence is unplanned
Offence with others	29	Child was with others during commission of offence
Penile pen	25	Offender penetrates male or female victim with penis
Pl hand	39	Offender places victim hand on his own penis*
Promises gift	26	Offender promises or gives a gift or money to victim*
Reassures	29	Offender reassures/talks to the child*
Rubs V's genitals outside	7	Offender rubs victim's genitalia outside victim's clothing*
Sex lang	23	Sexually explicit language used by offender *
Stranger	20	Offender is stranger to victim*
Strokes	49	Victim stroked or touched in area other than genitals*
Threat violence if report	9	Offender threatens violence if victim reports*
V injured	3	Victim receives physical injury
V mast O	33	Victim masturbates offender
V naked	24	Victim naked at time of offence*
V oral on O	20	Victim required to fellate offender*
V participate	40	Victim asked to participate in the offence*

Table 9.4. Details of optimal subset of 35 variables from Kirby data
*indicates variable was originally included by Kirby (1993); 'O' refers to offender and 'V' refers to victim.

The 35 variables were put into SSA-I with Jaccard's using local monotonicity in two dimensions. This gave an alienation of 0.29, less than with the analysis on 36 variables in Empirical Study 9.1. The plot is shown in Figure 9.10. The structural hypotheses of Kirby of Intimacy, Aggression and Sex were sought in the plot. As Figure 9.11 shows, these themes were readily apparent. The partitioning through regional interpretation was regular in this Jaccard's solution on the non-exclusivity sample of variables, and showed stronger support for the Kirby (1993) themes of Intimacy, Sex and Aggression. However, it is suggested that the reanalysis of the Kirby data with a

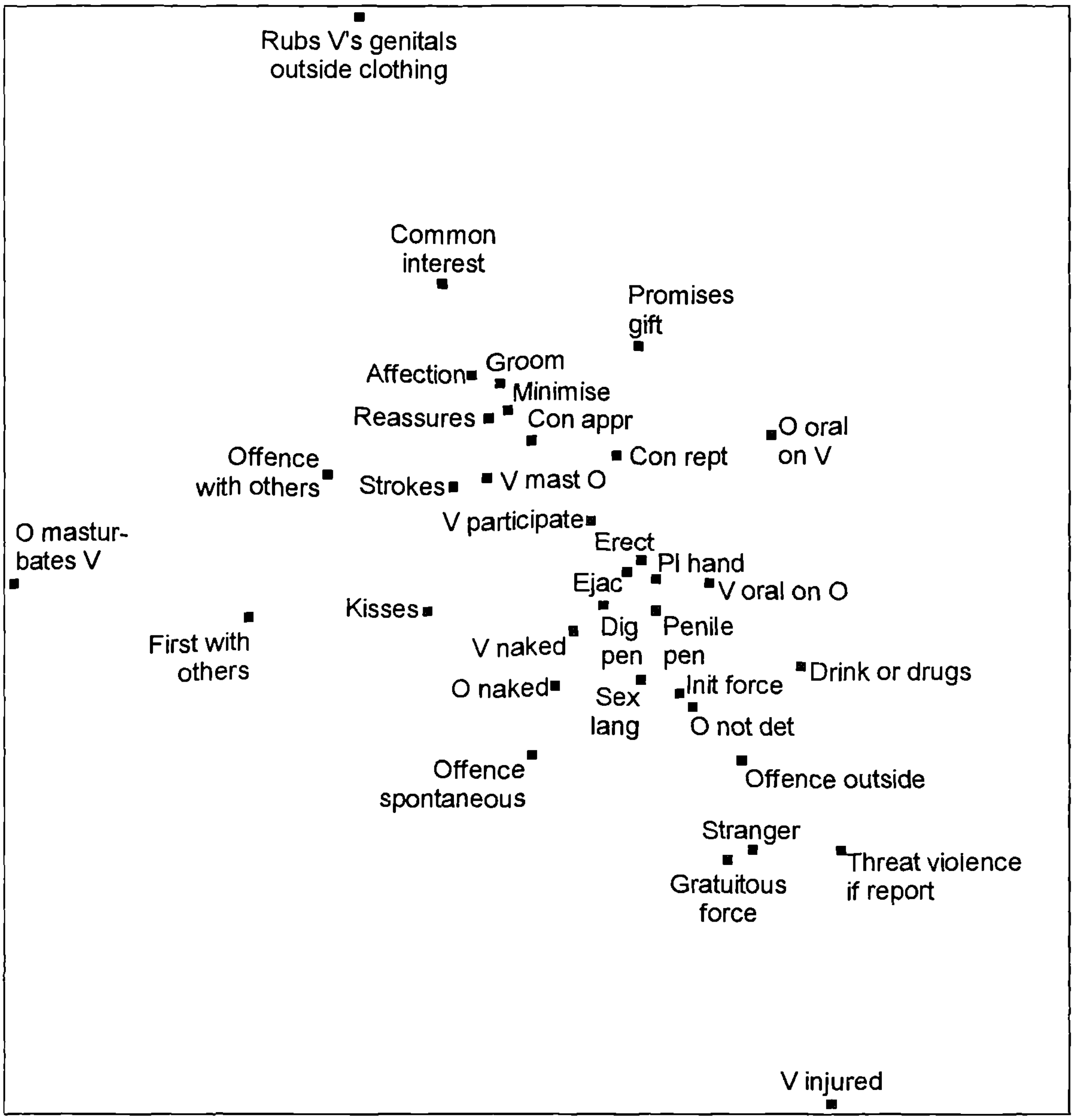


Figure 9.10. SSA-I of optimal subset of Kirby data

more appropriate coefficient and set of variables means that additional meaning can be found in the plot within Kirby's original structural themes.

One of the fundamental weaknesses of Kirby's regional interpretation of his plot in Kirby (1993) was that the Aggression and Intimacy themes did not contain any sexual variables. Core Aggression behaviours were those of using and threatening violence, and not being deterred by the child's adverse reactions. The intention of these behaviours was Aggression. Similarly, core Intimacy behaviours were using confidence strategies to get closer to the child in order to perform abuse, showing

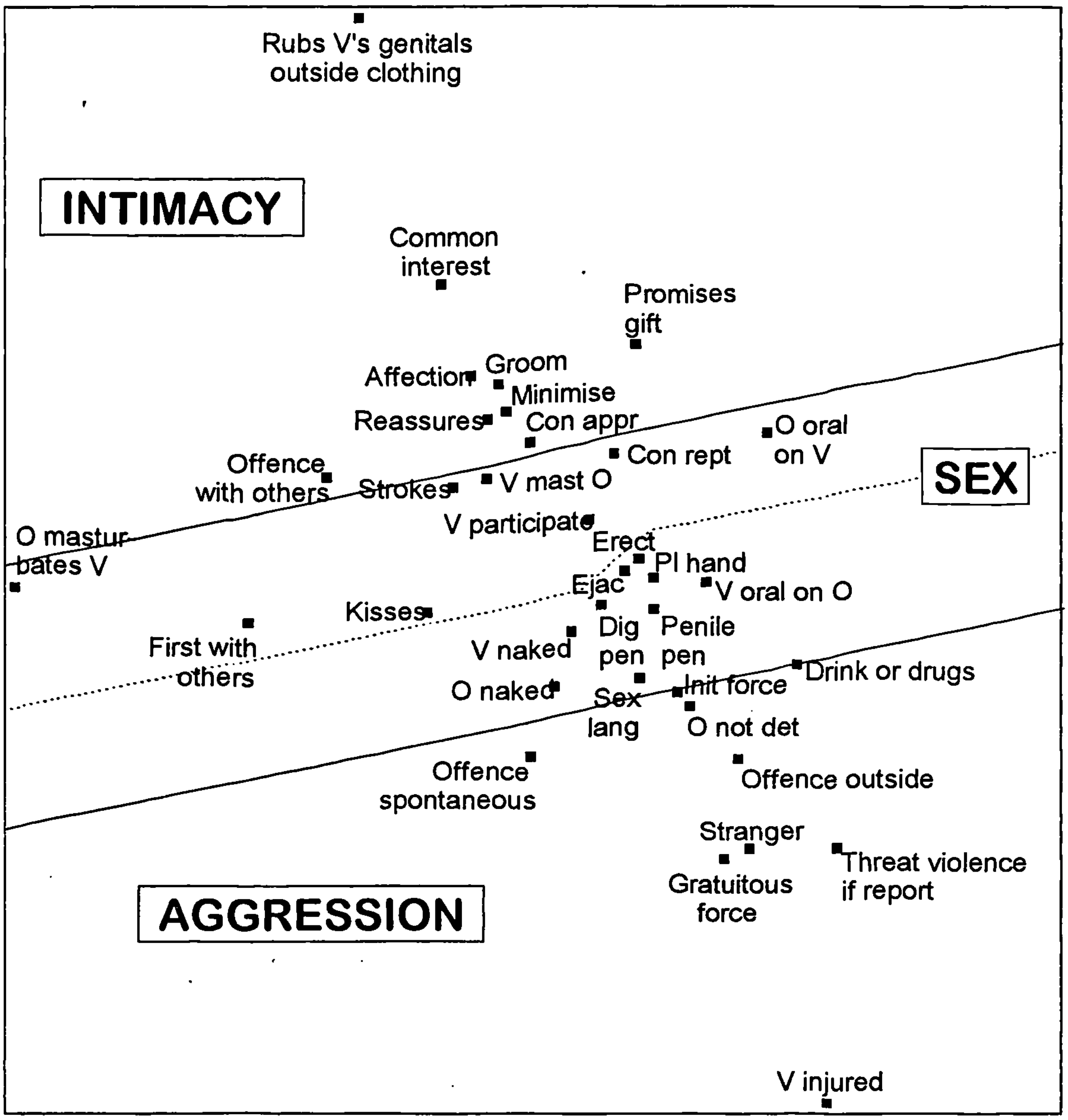


Figure 9.11. Partitioning of optimal subset of Kirby data into themes

affection, reassuring the child, and stroking the child in non-genital areas. The intention of these was an attempt at Intimacy.

However, it is argued here that Kirby's emphasis on sexual gratification as a separate Sex theme on a parallel with Aggression and Intimacy, was misinformed due to the poor representation of the plot. One of the fundamental differences between this plot and the plot of Kirby (1993) is in the spatial configuration of the sexually-related offence variables. Several of these for Kirby were to be found mixed with the non-sexually related interaction themes of Intimacy and Aggression. Thus the rubbing and stroking behaviours were not actually in the Sex theme at all but in the Intimacy theme.

In the present plot, the behaviours in the middle between Aggression and Intimacy are purely sexual, with none referring to approach leading to the offence or *sequelae* after the offence. The approach and *sequelae* are suggested by the non-sexual offence behaviour, in the styles of Intimacy and Aggression. These styles of *non-sexual* interaction with the victim then link with styles of *sexual* interaction, such that the sexual behaviour can also be differentiated quite cleanly into Intimate and Aggressive Sex. The hatched line in the original plot showed this, and this is schematically shown in Figure 9.12.

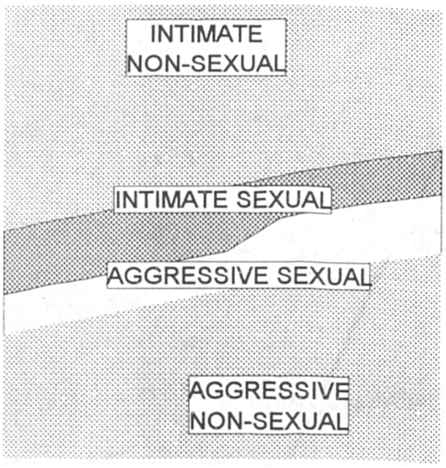


Figure 9.12. Schematic representation of themes from optimal subset of Kirby data

Support for this new structural hypothesis is actually given by Kirby (1993) who cited the explanation of Berkowitz (1993) that violence is on a continuum. The reanalysis shown above shows the continuum of increasing aggression is: Intimacy Non-Sexual, Intimacy Sexual, Aggression Sexual, Aggression Non-Sexual.

An offender who interacts with children in an aggressive and exploitative way - or is venting his frustration onto the child - will have the different probabilities of performing behaviours. This is shown in Figure 9.13 with the darker shading indicates higher probability. On the other hand an offender who interacts with children in an intimate way or in an attempt at pseudo-intimacy will have different probabilities of performing a behaviour in the different themes. This is shown by the darker shading indicates higher probability in Figure 9.14.

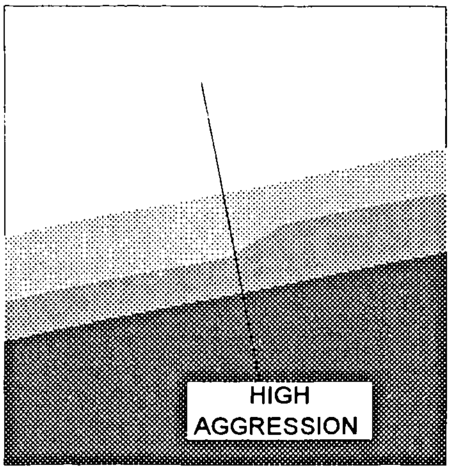


Figure 9.13. Schematic representation of aggression from optimal subset of Kirby data

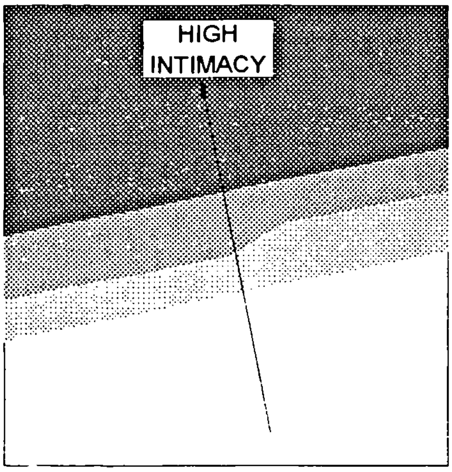


Figure 9.14. Schematic representation of intimacy from optimal subset of Kirby data

In a way, the Intimate and Aggressive Sexual behaviours act as Guttman scales such that in order to have intimate sexual behaviours it is implied that Intimacy theme had occurred, i.e. an intimate approach and post-offence release of the child. Equally, aggressive sexual behaviour need an aggressive and violent approach and are accompanied by physical threats. The Intimate Sexual actions have an intention of involving the child *with* sex; in some sense, the offender perceives he is giving gratification to the child. The Aggressive Sexual actions have an intention of using the child *for* sex; the offender is taking gratification from the child. Most importantly, another reanalysis of the original Kirby data have suggested that there is very little overlap between doing both of these styles of sexual interaction (Canter, Hughes and

Kirby, in press). This could be validated by examining the demographic lifestyles and previous convictions of the offenders, which were unavailable for the present thesis.

Therefore it is hypothesised that the Sex theme is actually an extension of the Aggression and Intimacy themes, and the sexual behaviour can be differentiated as indicated on the above plot by the dotted line in Figure 9.10. Support for this regional interpretation comes from consideration of the behaviours that are in the Intimacy and Aggression themes.

This additional interpretation to the Kirby data was facilitated by the reanalysis using Jaccard's, the use of non-gender specific variables and the minimisation of mutual, substantive and logical exclusivities. This allowed the true nature of the interrelationship of the variables to be modelled in the SSA-I and for a stronger structure to be found. Nevertheless, the power of the regional interpretation in terms of diminishing type II error was derived from the strong structural hypotheses in the original Kirby (1993) analysis.

The original Kirby study did not distinguish between the genders and investigated the structure of the domain of all molestations against children. It is possible that the aggression mode of interaction with the victim was disproportionately used against girls while the intimate mode was more characteristic of attacks against boys, and this interaction was shown by the characteristic plots in Figures 9.1 and 9.5. The removal of gender specific variables may have reduced this bias and caused the plot to become less 'clustered' around the two modes in separation from each other, instead allowing them to mix in the central sexual boundary shown in Figure 9.11.

In conclusion, Kirby (1993) hypothesised three senses to the domain of child abuse: intimacy, aggression and sexuality. However, the exclusivities in the set of variables 'pushed' some items apart, giving the impression that sexuality was quite distinct. However, the reanalysis using an optimal set of variables with exclusivity reduction showed that sexuality was better understood as being merged into intimacy and aggression, creating a continuum.

Finally, the frequencies of the optimal reanalysis were examined in Empirical Study 9.5.

Empirical Study 9.5: Frequencies of Reanalysis of Kirby Data with Exclusivity Reduction

The items in Figure 9.10 were plotted as frequencies of occurrence of that behaviour. This is shown in Figure 9.15.

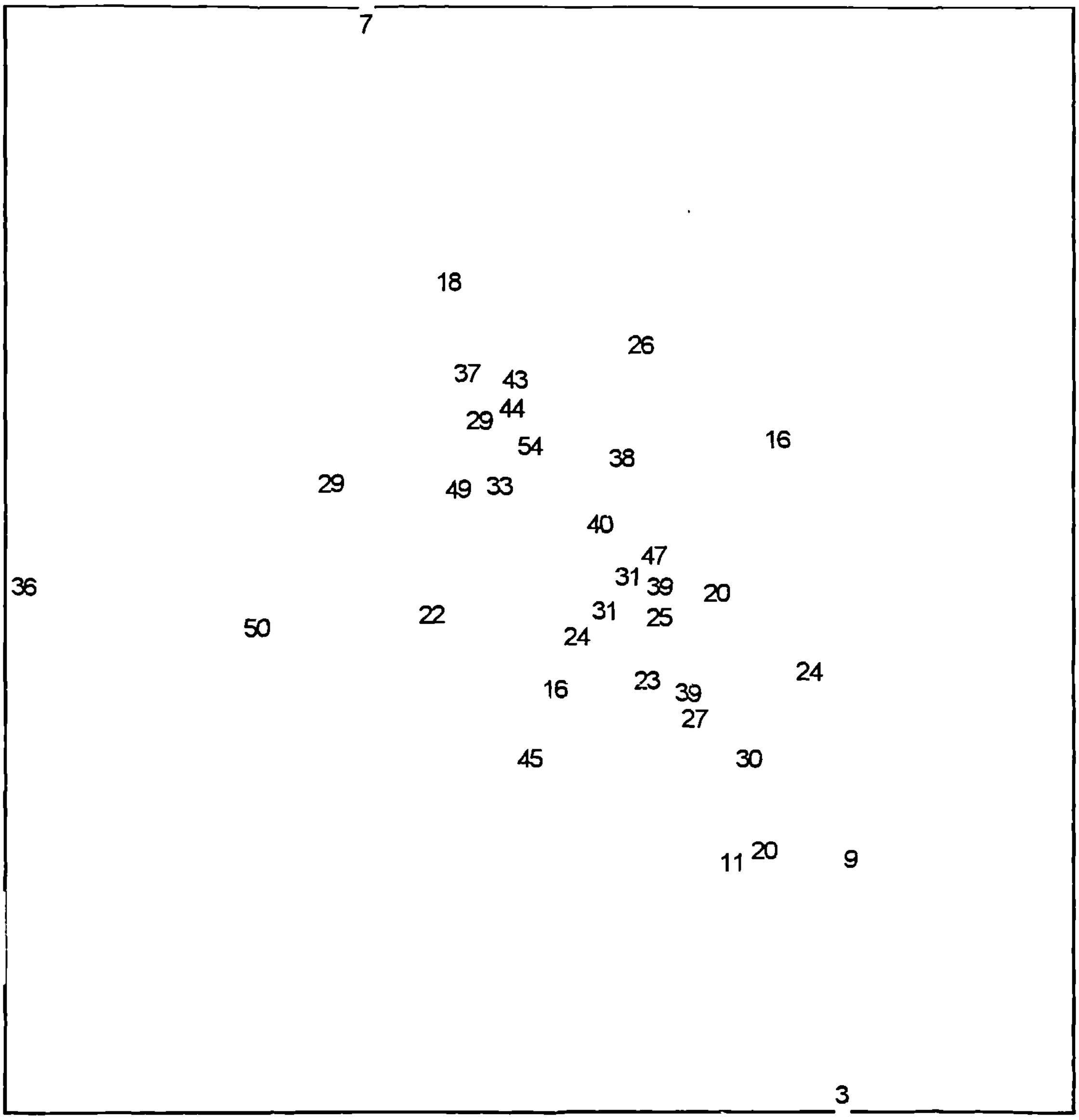


Figure 9.15. Item frequencies in SSA-I of optimal subset from Kirby data

Frequency contours were then sought around these item frequencies, which is shown in Figure 9.16. As can be seen from this figure, the contours are quite regular, with exceptions marked as arrows on the plot. More importantly, they are far more regular than those shown by Figure 9.4 in Empirical Study 9.1. Contrary to Kirby's assertion

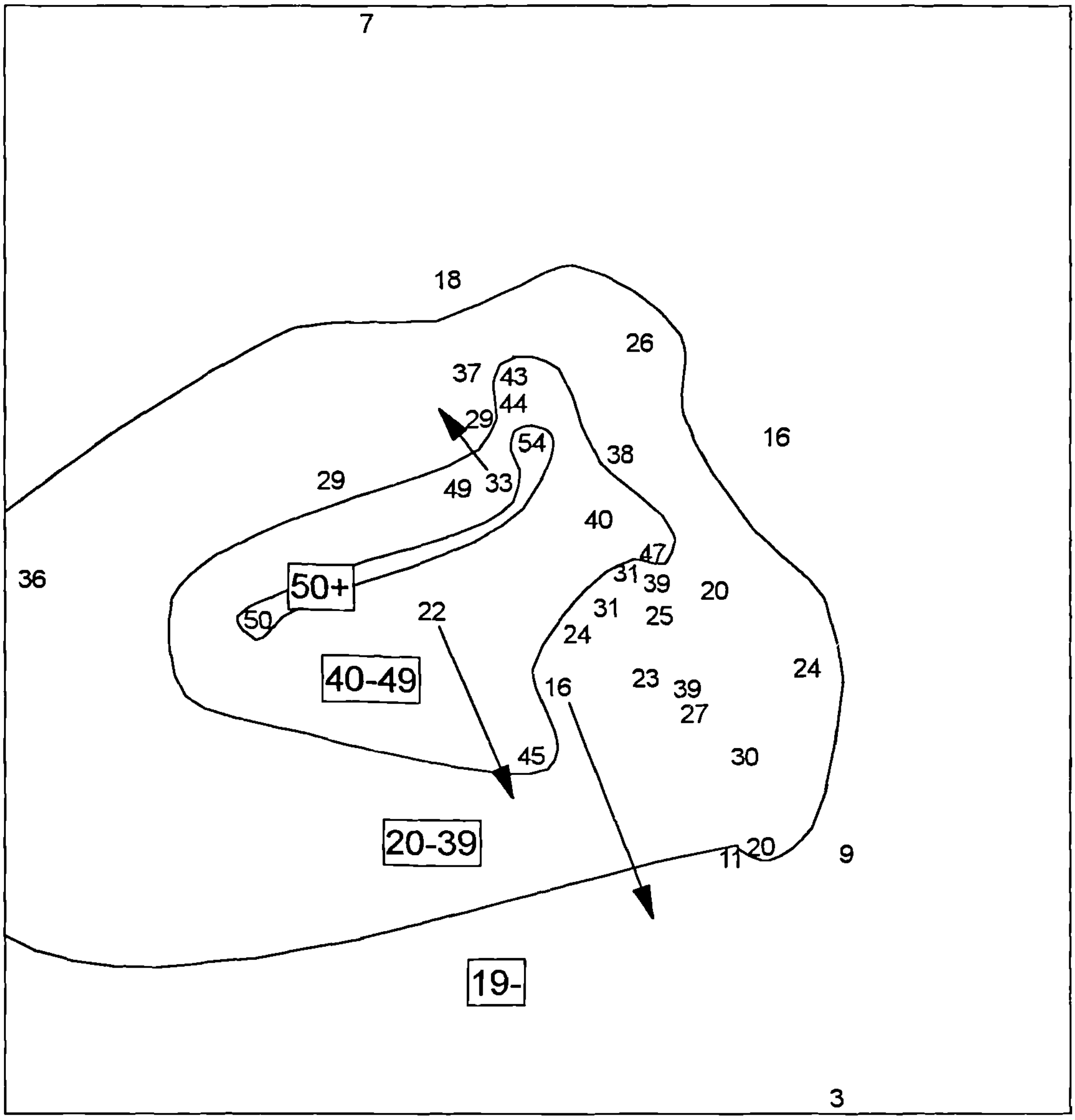


Figure 9.16. Frequency contours in SSA-I of optimal subset from Kirby data

that 'there were no variables conceptually central to child molestation' (Kirby, 1993, p. 184), this plot show that the revised analysis with optimal items has a central area on account of both frequency and meaning, mainly sexual and offence approach items. The removal of the gender-specificity means that the confusion and blurring in the domain of child abuse is removed, confusion created by including behaviours on two genders. This could be further investigated by comparing regional structures from male and female victims separately to see what - if any - differences emerge. The age of the victim may also act as a confounding variable.

Summary of Chapter 9

The Kirby Child Abuse Data Set was reanalysed using the more suitable Jaccard's index with local monotonicity, as compared to Yule's *Q*. Empirical Study 9.1 showed that the results did not replicate the frequency pattern in Canter and Heritage (1990) or Canter and Fritzon (1998). Empirical Study 9.2 showed that the results did not replicate the offence theme facet in Kirby (1993) either. Logical, substantive and mutual exclusivities in pairs of items in the analysis were proposed to be responsible for these results. Empirical Study 9.3 showed that such exclusivities could lead to poor recovery of structure and poor Local Spatial Bonding. Therefore Empirical Study 9.4 was run noting problematic combinations and using an optimal set of items with the Jaccard's index. This revealed a stronger understanding of the structure of offences against children than was shown by Kirby in terms of suggesting both intimate and aggressive sex, which was alluded to but not confirmed by Kirby. Empirical Study 9.5 showed that the frequency pattern in Canter and Heritage could be replicated better with the optimal set of items.

Chapter 10
structure and representation in factor analysis

Structural hypotheses about a noisy data set on Child Abuse from Kirby (1993) were shown to be highly susceptible to the association measure (chapter 8) and exclusivities in items (chapter 9). These structural issues were suggested to interact with representational issues, such as dimensionality and locality of monotonicity to influence the structural hypothesis testing procedure. The remaining chapters in the thesis are concerned with the structural analysis of the self-reported criminal actions of young males. Such a survey is presented in chapter 12 - the Youngs data on juvenile criminal actions.

It had been suggested in chapter 1 that crime surveys taken from the offenders themselves were especially useful source of information. The analysis of such surveys have generally employed the method of factor analysis, though some earlier studies of self-reported delinquency had used Guttman scaling and MDS to a lesser extent. Chapters 11, 12 and 13 offer an extended comparison between factor analysis and MDS on the Youngs data in terms of structure and representation. Chapters 14 and 15 revisit and revise Guttman scaling as applied to the Youngs data, and put forward some interesting links between the representation of SSA-I and the structural hypothesis of Guttman scales.

Firstly, then, it is necessary to introduce the key differences between factor analysis and non-metric analyses. This chapter explains factor analysis, highlights its shortcomings compared to MDS and shows how even in quantitative studies on personality - traditionally dominated by factor analysis - structure may be concealed by the factor analytical representation.

The Stages of Factor Analysis

Factor analysis has 'probably generated more empirical research than any other mathematical model in psychology' (Schönemann and Borg, 1981, p. 381). The method dates back to Spearman's investigations of mental abilities in 1904 (Hearnshaw, 1979), which was based on Galton's ideas about variation in ability amongst individuals and the measurement of that variation (Child, 1990). Originally, Spearman conceived of factor analysis as a structural hypothesis for the order of values in the correlation matrix to establish a hierarchy of scores (Guttman, 1966), termed 'order-factor analysis' by Guttman (1954b).

Modern factor analysis however is not so much a single technique but the general name given to a family of procedures with different means to a common end. This is systematised reduction in the rank of a two-way one-mode matrix. In fact, the most popularly used variant of factor analysis - principal components analysis - does not strictly use 'factors' at all (Kline, 1994). The designation 'factor analysis' or 'factor' in this thesis refers to the generic technique, unless otherwise specified.

Though factor analysis has similarities with MDS, it cannot be explained using the same terms of representation, transformation, data input and algorithm put forward in chapter 5. The broad stages in factor analysis are as follows:

- Stage 1: The Two-Mode Rectangular Input Matrix

Similar to SSA-I, the starting point for factor analysis in terms of input data tends to be a two-way two-mode rectangular matrix of cases (or subjects) by variables (or tests), but can be a two-way one-mode triangular matrix of cases or variables.

- Stage 2: Creating The One-Mode Triangular Correlation Matrix

If not using a triangular matrix already, the next stage in factor analysis is the creation of the matrix of correlations between variables, or with the less common *Q*-factor analysis this is the correlations between objects. It is also possible to perform a second-order factor analysis, where the factor structure obtained from a previous factor analysis is inputted again.

The measure of similarity between each pair of variables in factor analysis is normally the Pearson's Product-Moment Correlation Coefficient or its dichotomous equivalent (Gorusch, 1988).

- Stage 3: Transforming the Correlation Matrix

In factor analysis, the correlation matrix is transformed by reducing its rank through the extraction of components or factors. In principal components analysis, one component at a time is successively extracted from the correlation matrix, such that the extracted component accounts for the maximum amount of variance in the matrix. The amount is termed the eigenvalue, which when squared gives the variance accounted for. The correlation matrix then has that component partialled out of it and then another component accounting for the next most variance is extracted. Components thus successively account for a diminished proportion of the variance in the correlations. This process continues as desired, with later components extracting very little variance since most has already been extracted. In principal components analysis, the ideal situation is where most variance is accounted for in far fewer components than the number of original tests. In this way, the components act as a summary of the correlation matrix.

The common factors (rather than components) of a correlation matrix are similar except that unwanted unique variance in the test items is estimated and removed. This may add extrinsic error to the factors. Since components do not do this, they are more suitable where test reliability is high and the tests are not idiosyncratic. However, Harman (1976; cited by Kline, 1994) suggested that with large matrices the difference between principal components and common factors is minimal. However, it was noted that even this fundamental difference 'would lead to very similar solutions' (Velicer and Jackson, 1990, p. 10) if the correct number of components were extracted.

- Stage 4: Optimising the Transformed Matrix

The variables or tests in the original correlation matrix have been summarised such that each original test item now is a correlation (or 'loading') with the factors or components extracted. This could be an end in itself, giving a direct solution (Child, 1990). However, the distribution of factor loadings can be improved by rotation of factors to a derived solution (Child, 1990). The combinations of variables in the factors stay the same in derived solutions though the loadings change. There is an infinite number of rotated solutions which are all mathematically equal, but factor analysts seek easier interpretability in solutions - a heuristic requirement. This is the 'simple structure' of Thurstone which attempts to make rotated factor loadings either very high or very low.

There are two distinct style of rotation which make different presuppositions about the nature of the factorial structure of the content universe. Orthogonal rotations make the factors independent of each other (i.e. uncorrelated), whereas oblique rotations allow the factors to correlate with each other. This nomenclature was inspired by the original geometric approach to factor analysis such that the factors were at right angles (i.e. orthogonal) to each other or the angle between them was allowed to vary (i.e. oblique).

- Stage 5: Presentation and Interpretation of the Rotated Factors

The presentation of the factors (or components) is usually a list of the tests or variables and their loadings on all rotated factors. The loadings for all the tests or variables against a particular factor are examined. The summation of squared loadings of a variable on all the factors extracted is termed the communality, which gives the amount of variance in the variable explained by the factors extracted. Since the highest loading of a variable on a factor accounts for most of its variance, it is assumed that the item adds most to the meaning of that particular factor. Conversely, those variables loadings on factors below a certain cut-off point - normally $\pm$ 0.3 - are ignored (e.g. Kline, 1994), though a cut-off can be calculated from the Burt-Banks calculation (Child, 1990). Where a variable loads on more than one factor - the occurrence of which simple structure attempts to diminish - then the highest loading tends to be taken. The meaning of the factors is denoted by examining those tests that load most highly on it, with a high negative loading indicating what the factor does not contain.

Where factors are oblique and intercorrelate, the relations between factors may also be examined. If there are many meaningful intercorrelations, then the oblique factor matrix may also be factor analysed to achieve 'higher order' factors. A hierarchy of factors and variables may then be built up.

A Critique of Factor Analysis

Despite the stages in the previous section seeming straightforward, there is much controversy about factor analysis. This section summarises some of the main objections following the same order of stages as above, starting with stage 1, the creation of the input two-way two-mode matrix.

It was stated by Kline that in the 'in any field there must be a good rationale for sampling variables' (Kline, 1994, p. 72), by which Kline intended at least three variables per factor. This would eliminate the possibility that factors may be bloated specifics, namely completely idiosyncratic unrepresentative items. Additionally, as Kline pointed out, if the sampling of the content universe is not complete and exhaustive then the factor analysis will not show all the universe. This is not such a threat to MDS solutions since missing items (or questions not asked) appear as 'gaps' in the geometric space. Furthermore, the explicit consideration of what might constitute the content universe *a priori* in the faceted approach may reduce this possibility even more. In fact, it is possible to use faceted design with factor analysis (Shye and Elizur, 1995), though this would undermine the benefit of the overall faceted approach.

Similarly, the criterion on the number of respondents is more important to factor analysis than MDS, with recommended ratios ranging from 10:1 to 2:1 respondents to items, with a minimum of 200 respondents (from Kline, 1994). Child 1990, p. 115) suggested a more rational criterion by Baggaley calculated using the number of items and average inter-item correlation. This is important for factor analysis comes due to the requirements of the PMCC coefficient used at stage 2 of the analysis. In non-metric SSA the respondent:variable ratio is less important since firstly the coefficients are not necessarily correlations, and secondly the ordinal transformation from similarities to distances diminishes the effects of systematic error in the distribution of the association or correlation matrix.

Guttman was critical of the '[g]reatly exaggerated use .. of linear equations' of Pearson's PMCC which is extrinsic to analyses and would 'unnecessarily introduc[e] error of approximation of its own' (Guttman, 1986, p. 82). As was seen in Empirical Study 8.1, the actual difference between the dichotomous PMCC and Yule's Q was only slight, meaning that the adoption of less restrictive coefficients has a larger substantive impact than statistical effect.

At stage 3, the transformation of the correlation matrix into a new form, variance is lost by both factor analysis and MDS, since errors of approximation are found where the inherent dimensionality of the data is decreased in real-world data. However, in terms of proportion of variance lost, what would be held as an acceptable loss for factor analysis would be disastrous for a SSA-I solution.

For example, consider a factor analysis that recovers 47.3 percent of the variance in the first 6 factors (as in the factor analysis of the Youngs data in Empirical Study 12.1). It can be calculated what sort of SSA-I representation would be found with this amount of variance since in SSA-I the amount of variance from the correlation matrix explained by the distance matrix is simply the squared correlation between distances and correlations (Borg and Lingoes, 1987, p. 46). Let us assume that the correlations were approximately equal to Guttman's coefficient of monotonicity. The alienation of such a plot using Formula 5.5 i.e. $K = \sqrt{(1 - \mu^2)}$ would be $\sqrt{(1 - 0.473)} \approx 0.73$. This alienation would be far from as acceptable, meaning that the variance explained would be far from adequate. Yet there is nothing in the factor structure matrix or the indicators of 'factorability' in Empirical Study 12.1 that would highlight this - the factors may still be interpreted as valid and useful.

By the same token Shepard (1978) showed that 99% of the variance of the Ekman data on visual perception could be explained by using two spatial dimensions, which was better than going to five orthogonal dimensions in a factor analytical solution. Schlesinger and Guttman stated 'Smallest Space Analysis makes it possible to arrive at a smaller space than does factor analysis' (Schlesinger and Guttman, 1969, p. 99).

In fact, the initial approximation of SSA-I and MINISSA-I (N) is a principal components extraction on the ranks of the similarity matrix. Inevitably the variance explained in the MDS solution must increase, unless the initial approximation was extremely good.

The optimisation of the transformed matrix in factor analysis attempts to rectify this selective and hence poorly distributed variance in stage 4, whereas the optimisation with SSA-I is for purely computational reasons. The rotation of axes in non-metric MDS does not change the plot in any way, since the distance information is invariant under this geometric transformation and they are not 'weighted' in any way. However, metric MDS with individual differences weightings are closer to the factor analytical rotation of extracted factors since they do rely on the orientation of axes.

Rotation to simple structure is a convenient solution to the problem of factor invariance (Child, 1990), where subsequent replications on different samples or different tests do not produce the same factor structure. This phenomenon severely limited the scientific usefulness of factor analyses, so the heuristic usefulness of simple was seized upon by factor analysts and is now taken by convention. Kline (1994, p. 44) suggested that principal axis factoring improved invariance in that it removed non-systematic variance, though the disinction between principal axis factoring and principal components extraction was low where communalities were high. Similarly, Kline stated that MLE methods of factor extraction are close to principal components extraction. However, MLE models have the restriction that respondents be sampled randomly from the population, though the advantage of offering a statistical test for the number of factors to be extracted. Nevertheless Velicer and Jackson (1990) criticised the chi square tests in MLE models on the grounds of indicating too many factors.

The issue of which is the 'best' of these infinite solutions is still not decided (Child, 1990, p. 48), though simple structure is by far the most popular. At its very worst, rotation to simple structure could mean that everybody is rotating to the same inadequate solution. The use of Thurstonian simple structure is not the only set of rotational criteria. For example, an early factor analytical study of delinquency by Lander specifically rejected the use of 'arbitrary rotational criteria set up by Thurstone for the isolation of simple structure' (Lander, 1954, p. 51) It employed the 'rotation as hypothesis' argument (e.g. Kline, 1994) and extracted factors by the centroid method. In this way Lander identified and named two factors supported by external criteria and anticipated by research hypotheses. Lander then hypothesised that the 'socio-economic' and 'anomic' factors could be rotated obliquely to produce certain loadings for certain variables in the study; this hypothesis was supported.

Guilford (1959) argued that under the meaning of simple structure, an orthogonal rotation of factors is essential since an inter-correlated set of (oblique) factors is less 'simple' than an uncorrelated set of (orthogonal) factors. Eysenck reported that for him the Thurstonian 'demands of simple structure and those of orthogonality of factor structure were found to be irreconcilable' (Eysenck, 1970, p. 16). This led to his adoption of oblique rotations and second-order factors.

Furthermore, Kline (1991) suggested that as a means to an end oblique rotation is more favourable since the rotated structure obtained will be more simple in terms of maximising the number of factors with loadings of near zero with a few having high loadings. Assuming this to be the case, the widespread use of oblique rotations with principal components extractions could be viewed as somewhat contradictory, since the components extracted are orthogonal to each other due to preceding components having been removed from the residual correlation matrix by partial correlation.

Not only is variance lost during the factor extraction stage, but further variance is plainly ignored during the interpretation stage. This problem is addressed in greater detail in the next section.

The Presentation and Interpretation of Factors: An Incomplete Representation

However, perhaps the key difference between factor analysis and MDS is in the nature of the dimensionality of the solution and its interpretation, at stage 5. Jacoby asserted that dimensionality is 'simply defined as the number of separate and interesting sources of variation among the objects' (Jacoby, 1991, p. 27). In factor analysis, each dimension is indeed literally taken as separate and *assumed* to be interesting - an assumption that each and every dimension extracted by the program is conceptually unique and distinct.

Each dimension or factor in factor analysis is identified as a source of variation, contravening the admonition of early factor analysts such as Burt (1940) against reifying factors. MDS on the other hand uses dimensions as a framework onto which to map variation. With factor analysis, the axes of the solution give the form its meaning by the co-ordinates on the items on those axes; with MDS, however, the axes are simply a tool for defining the distances between objects, from which meaning is derived.

The interpretation of rotated factors from factor analytical solutions was critiqued powerfully by Maraun (1997). It was pointed out by Maraun that in the (geometric) factor space, the coordinates - i.e. factor loadings of the variables - function purely as locators for the points or variables. The factors and the factor loadings are the representation of the structure extracted from the correlation matrix, which was hypothesised to sample the domain. The only essential requirement for the axes (i.e. the factor-dimensions), argued Maraun, was that they located the variables as reliably as possible. This is analogous to the requirement that an MDS representation be accurate as possible by achieving a global minimum.

Given that the factor-axes only serve as representation of the underlying structure, Maraun argued that orthogonality or obliqueness or type of rotation of the factor-axes was irrelevant. The representation of the structure in the space was indicated by the configuration of the vectors from centroid to variables. Yet regularly the interpretation of the representation is made deliberately unreliable.

Maraun (1997) argued that this commonplace exclusion of factor loadings less than 0.3 or 0.4 in value makes the use of the factors as axes meaningless. The representation of the structure is unreliable since the factor-reference axes range from +1 to +0.3 then -0.3 to -1. How can a coordinate representation be reliable when the axes are missing? Maraun analysed published studies showing that up to a third of variance is ignored by a cut-off of 0.4. Small factor loading such as 0.05 were just as important as a large loading of 0.75 since values are required to locate the variable in space, Maraun argued. The parallel for excluding items in SSA-I would be simply to physically cut out the middle of the plot, since variables in that part of the space are equally related to the points at the periphery and hence are not useful.

In rotation to Thurstonian simple structure, items are supposed to load highly only on one factor. This is not always the case where the number of factors is low. Items may load on more one than factor, meaning that the variance is spread across the factors rather than restricted to one. Since items tend to be interpreted as indicating the meaning of the one factor with which it has the highest factor loading, then the extra variance is lost. The naming of dimensions by identification of content relies on regression of the variables onto dimensions one at a time, ignoring the loadings other than the highest.

In the instance of the colour circle, Shepard (1978) showed that Ekman's data on perceived colour similarity were readily and plausibly interpretable as a five orthogonal factor model, as indeed Ekman had claimed. Furthermore, Maraun (1997) pointed out that the (arbitrary) cut-off of 0.4 on the two non-factor reference axes in the Shepard (1978) colour circle would readily hide the data circumplex structure. Certainly a faceted radex structure was hypothesised and found in a low alienation two-dimensional representation by Maraun (1997), using the scales of the NEO-PI and Goldberg-40 personality inventories.

Since the dimensions of factor analysis are taken to be the meaning of the content universe, the 'correct dimensionality' problem encountered in chapter 6 for MDS gains critical importance for factor analysis.

Representation, Dimensionality and the 'Correct' Number of Factors

Given the importance of choosing the 'correct number' of factors in factor analysis, the solution to this problem is far from definite. As Block commented:

> Although the method of factor analysis has been used for almost a century, there is still not a clear, unequivocal basis for deciding on the number of "factors" to extract in a factor analysis or how to obtain an "optimum" rotation of the particular set of factors settled upon. (Block, 1995, pp. 189-90)

Cattell (1978) stated that the exact number of factors to be extracted was a fundamental issue: underextraction leads to vague, higher-order factors; overextraction leads to splitting of factors during rotation (Kline, 1993). Given the two evils of overextraction and under extraction, Kline (1994) commented on the tendency of researchers to overextract factors rather than risk losing structure, and Comrey (1978; cited by Velicer and Jackson, 1990) confirmed that underextraction was the greater of the two evils. Factor analysts need instructions on the 'correct' number of factors to extract. However, unlike the guidelines on MDS fit values presented in chapter 6, which can be modified and understood with the nature of the data and the research, factor analytical guidelines are far more crucial to the whole analysis.

When not using maximum likelihood (MLE) tests for the number of factors, two methods are generally used to calculate the number of factors to be extracted: the Cattell 'scree test' and the Kaiser-Guttman criterion (Child, 1990). The scree test (Cattell, 1966) assumes that the first few factors in extraction contain the highest proportion of common to unique variance, a proportion which decreases exponentially such that the last few factors contain only unique variance. The plot of eigenvalues against factors shows an 'elbow' where no more common variance is included in the factors - viz. where the scree begins. (This has a parallel with the elbowing test of Kruskal (1964a) plot of stress against dimensionality in MDS for deciding the correct dimensionality.) Kline (1994) suggested that gauging the elbow on the scree test was possible with a high inter-rater reliability.

The 'Kaiser-Guttman root of unity' test - also known as 'eigenvalue-one criterion' or 'truncated principal components' - refers to retaining unities in the diagonal of the correlation matrix as communalities and extracting all factors with eigenvalues greater than unity. Hakstian and Muller (1973) suggested that the Kaiser-Guttman criterion arose from the marriage of Kaiser's use of 'psychometric logic' with the misapplication of Guttman's algebraic work on the lower bounds of the principal components of a correlation matrix. As H.B. Lee and Comrey (1979) explained, the criterion is actually based on components rather than factor analysis. In large data sets, it was found that this method included too much error variance in too many factors. Velicer and Jackson (1990) claim that the use of the Kaiser criterion, the default option in many statistics packages, causes most of this overextraction by typically retaining $n / 3$ components with n variables in the analysis. Too much error is included since both error and specific variance is included in the principal components solution with communalities of one. With empirical data on orthogonal rotations, H.B. Lee and Comrey suggested that it produced distortions that threatened the validity of the factor interpretations. The eigenvalue-one criterion and varimax rotation together should be used with 'extreme caution' (H.B. Lee and Comrey, 1979, p. 319) - despite being the default option in many statistics packages such as SPSS!

Ekman (1954) reported that the structure of colour was of five orthogonal factors since they accounted for the variance of colour (cited in Shepard, 1974; 1978). The reason for five factors was because a five-dimensional space was used to rotate factors and represent the items; the 'correct' number of factors was five. A reanalysis of these data by Shepard was based around the structural hypothesis of colours as a circle that had been traced back to Newton by Herrnstein and Boring (1966; cited by Shepard, 1978, p. 39). Using dimensions only as part of the spatial representation rather than being used to decide the structure, Shepard (1978) confirmed the circular order of the colours. A five factor dimensional interpretation may have been 'correct', but a two-dimensional spatial interpretation was more scientifically useful.

Factor Analysis and its Inadequacies: A Summary

In factor analyses, it is assumed that structure in any content universes can be sampled adequately in terms of items and respondents which conform to the requirements of Pearson's PMCC. The extraction of factors or components with or without extrinsic error estimates is sometimes guided by opposing principles. By this point, much variance may have been lost due to poor sampling of the content universe, poor items design and the underextraction of factors. Factors are then improved on 'interpretability' according to arbitrary criteria. These are placed into the geometric representation, when more variance is then ignored when items are regressed onto one factor.

As Guttman pointed out: 'The purpose of factor analysis *has* basically been to study configurations of points' (Guttman, 1964, p. 35, emphasis in original). Yet factors are now reified as being *the* structure, ignoring the advice of Burt - a founding father of factor analysis - that factors should be 'thought of in the first instance as lines or terms of reference only, not as concrete psychological entities.' Burt (1940, p. 18)

McGrath (cited by Brown, 1985) pointed out that factor analysis essentially provides a neat *a posteriori* summary of what went into the analysis, when the real problem involves knowing what to put in *a priori* and what the representation tells us about the true structure of the domain. The factor analytical structure is *created* by factor analytical representation; the faceted structure is *suggested* by the faceted representation.

The Practical Implications of the Controversies in Factor Analysis

The 'correct number of factors' and other factor analytical parameters in the debate above are not purely abstract issues. Instead, they have wide reaching substantive implications which have polarised the debate over studies of individual differences in intelligence and personality.

Scores on intelligence test items are normally positively correlated with scores on other intelligence test items. For example, Guilford found that of 50,000 correlations of test items, 'fewer than five per cent were negative, and most of these not significantly so.' (cited by Hearnshaw, 1979, p. 50) Spearman suggested this could be explained by the 'Two-factor' on account of each test containing some saturation of general ability ('*g*') and some specific test variance (Thomson, 1948). However, Kelley and Burt refuted this extremely parsimonious explanation and suggested that additional 'group' factors were required to fully account for the variance (Hearnshaw, 1979). Thurstone (1935) replaced *g* with eight factors of primary mental ability, and Guilford (1959; cited by Hearnshaw, 1979) posited 120.

Guttman (1966) pointed out that the refutation of *g* and the popularisation of factor analysis into Thurstonian 'common-factor analysis' (Guttman, 1954b) led to researchers neglecting patterns of order in the correlation matrix. Instead, Guttman offered a definition of intelligence from which structural hypotheses can readily be derived (cited by Gratch, 1973; also Guttman and Levy, 1991). Guttman was attempting to develop a 'newer theory [that] unifies what are otherwise opposing schools of thought' (Guttman, 1954b, p. 260). In doing this, Guttman attempted to move emphasis away from the representational issue of the number of factors and back to structural concerns regarding the nature of intelligence tests.

Equally, factor analytical research into personality has suffered from similar difficulties. All trait approaches to personality assume that behaviour is fundamentally consistent across time and situation; thus the manner in which people act and the way they react does not change (Eysenck, 1970). Those enduring and stable characteristics are operationally defined as traits. Eysenck (1970, p. 9) cited Allport (1937) as suggesting that these traits may not be measured directly or observed but only inferred. The most usual method for discovering traits is by self-report questionnaires asking items which purport to measure different constructs, though sorting tasks are also used.

In the example of the 16PF personality inventory of Cattell (1957), these constructs were derived from factor analyses of the Allport set of English language adjectives describing people. Cattell derived lower order factors with oblique rotations to find inter-related dimensions of personality. By contrast, Eysenck (1970) attempted to find the fewest number of independent personality dimensions, thus using higher order factor analysis with orthogonal rotation to find two independent dimensions of Extraversion - Introversion (E), and Neuroticism and Stability (N). Indeed one of Eysenck's strongest criticisms of Cattell's work was that a second order factor analysis of Cattell's oblique factor correlation matrix did in fact produce the same two dimensions that Eysenck had previously found. To the dimensions of E and N were added two more dimensions of Psychoticism (P) and Social Desirability (L) by Eysenck and Eysenck (1975). Previously, L had been the 'Lie scale' (e.g. Eysenck, 1957, p. 203) to detect untruthful responses, but was here given full dimensional status. This development took the Eysenck personality inventory closer to the Big-Five explanation of personality dimensions (e.g. Costa and McCrae, 1992), which offers another answer to the 'number of factors' question.

Maraun (1997) observed that the Big-Five is not so much a statement about the structure of trait descriptors as a statement about the dimensionality of the space in which they are placed. In other words, the debate over the Big-Five or three or two is about representation, as is the present thesis. By contrast, Maraun went onto claim that 'the structure of the trait descriptors is still very much an issue', echoed by Wiggins' earlier plea for the 'use of explicit structural models in research in personality and social psychology' (Wiggins, 1980, p. 266).

The Wiggins Circumplex and the Structural Analysis of Personality

An alternative structural approach to personality is the interpersonal wheel of Wiggins (1979), though this is not without flaws. Wiggins (1980) characterised much personality research as being based around separate, distinct dimensions of personality such as authoritarianism or achievement, with little work on the interrelationships of these constructs. In an attempt to integrate these similar sub-domains, Wiggins proposed the circumplex model of interpersonal behaviour as providing 'a coherent rationale for expecting a definite pattern of relationships to exist among indicants of personological constructs' (Wiggins, 1980, p. 287). The eight constructs of the Wiggins (1979) circumplex deliberately overlapped sixteen of the most common personality scales around the wheel.

Wiggins (1979) worked from the same principles as the factor analytical trait theorists Cattell and Eysenck in that a large set of natural English language adjectives was condensed into a more explanatory few, though with greater reference to Leary (1957). The circumplex structure was advantageous since

> Strictly empirical procedures of variable selection are likely to deemphasize the importance of certain variables that are implied by the logic of the circumplex system but are underrepresented in the English language. (Wiggins, 1979, p. 400)

In other words, the spaces or discontinuities in the circumplex of traits imply items that have not been included in the study, given that the content universe were to be ordered as a circle.

Within scale variance was far less than the across scale variance, with all scale Cronbach's α reliabilities high and at least above 0.8 (Wiggins, 1979). The results of several cross-validations by Wiggins showed that quasi-circumplexes highly close to perfect circumplexes were found, which was later backed up by Wiggins, Steiger and Gaelick (1981).

In a later revision to the interpersonal wheel, Wiggins, Trapnell and Phillips (1988) performed an item analysis on the projection of the original Wiggins (1979) items on the first two principal components. Items were culled if they did not fall on the expected part of the wheel, reducing the number of items by half to 64. Each of the eight variables ('octants') thus contained eight items. Again, the representation used by Wiggins to examined the fit to the wheel was the loadings on the first two principal components extracted from the correlation matrix. Again, the final configuration was an amalgamation of the eight items into the eight variables, creating an average of the within variables variance and thereby increasing the across variable variance. And again, the structure was circumplicial. Thus the Wiggins circumplex is not without difficulties, with Wiggins succumbing to item analysis in the *construction* of the circumplex.

Within the eight item scales there was a great variation in mean and standard deviation. For example, the average rating on the Cold-Quarrelsome variable was 2.66, standard deviation 0.84. This means that few if any of 610 people in the sample rated themselves worse than 5 out of 8 on the scale. Conversely, its bipolar opposite Warm-Agreeable was mean 6.91, standard deviation 0.77, meaning that few rated themselves worse than 5 out of 8 on this scale. (These calculations assume a normal distribution - as does Wiggins.)

Consideration of the *content* of these scales is illuminating: there is a strong element of social desirability. Furthermore, a quick calculation shows that there is a correlation of 0.9788 between how 610 people score themselves and what a sample of 100 gave as socially desirable scores (calculated from Wiggins, 1979, p. 407, table 3). This casts doubt about the Wiggins circumplex as a representation of the structure of the original items due to the confounding effects of social desirability, which as will be seen later was measured specifically by Eysenck in the Lie scale.

More importantly, criticism of Wiggins' interpersonal wheel must also focus on the fact that it uses loading plot for the first two principal components - only a first approximation in SSA-I (see chapter 5). Nevertheless, this is preferable to the dimensional interpretation with simple structure and factor loading cut-offs, as criticised forcefully by Maraun (1997), since the principal component plots retain the geometric nature of the representation.

It is the continuous nature of Wiggins' circumplex representation that allows this flexibility of alternative structural hypotheses to be proposed and found. This continuity can be appreciated by examining the geometric space of the items that make up these scales, which can be calculated from the angular coordinates i.e. the angle of the items and the items distances from the centroid given by Wiggins, Trapnell and Phillips (1988). The distance from the centroid was also the communality of the item, indicating the amount of variance in the variable accounted for by the two principal components, so that items close to the centre are poorly modelled. These polar coordinates were converted into Cartesian coordinates, and gave the factor loading plot shown in Figure 10.1.

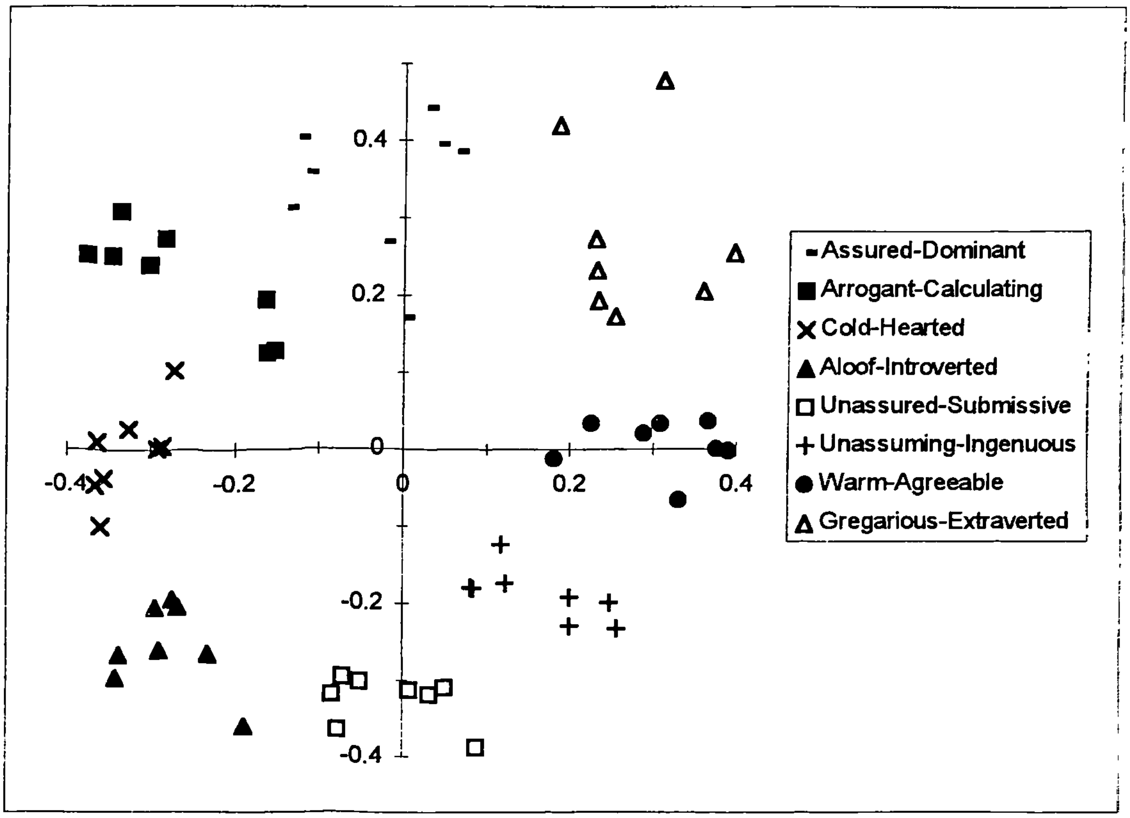


Figure 10.1. Factor loading plot of Wiggins personality scale, calculated from polar coordinates in Wiggins, Trapnell and Phillips (1988)

As can be seen from this plot, the points are not arranged as a geometric circumplex, and the 'quasiness' is too great to justify a quasi-circumplex. The circle found for example by Wiggins (1979) is due to the amalgamation of items into scales. When these scales are decomposed back into their items only a regional interpretation of the eight octant variables as a polar facet in the above plot is justified, even if this is a principal components plot. This is shown in Figure 10.2.

Furthermore, there does seem to be some extra component related to distance from the centroid. This becomes more evident in the plot above than in the table of communalities for the items. For example, the Warm-Agreeable items are ordered in increasing distance like this: Accommodating, Kind, Charitable, Sympathetic, Softhearted, Tender, Tenderhearted and Gentlehearted. The spread of items in this octant is suggestive of a linear progression - a simplex (e.g. Lingoes, 1977b). The angular separation between these items in the scale and the centroid is low. This means that they vary mainly in degree and not type. This can be contrasted with its

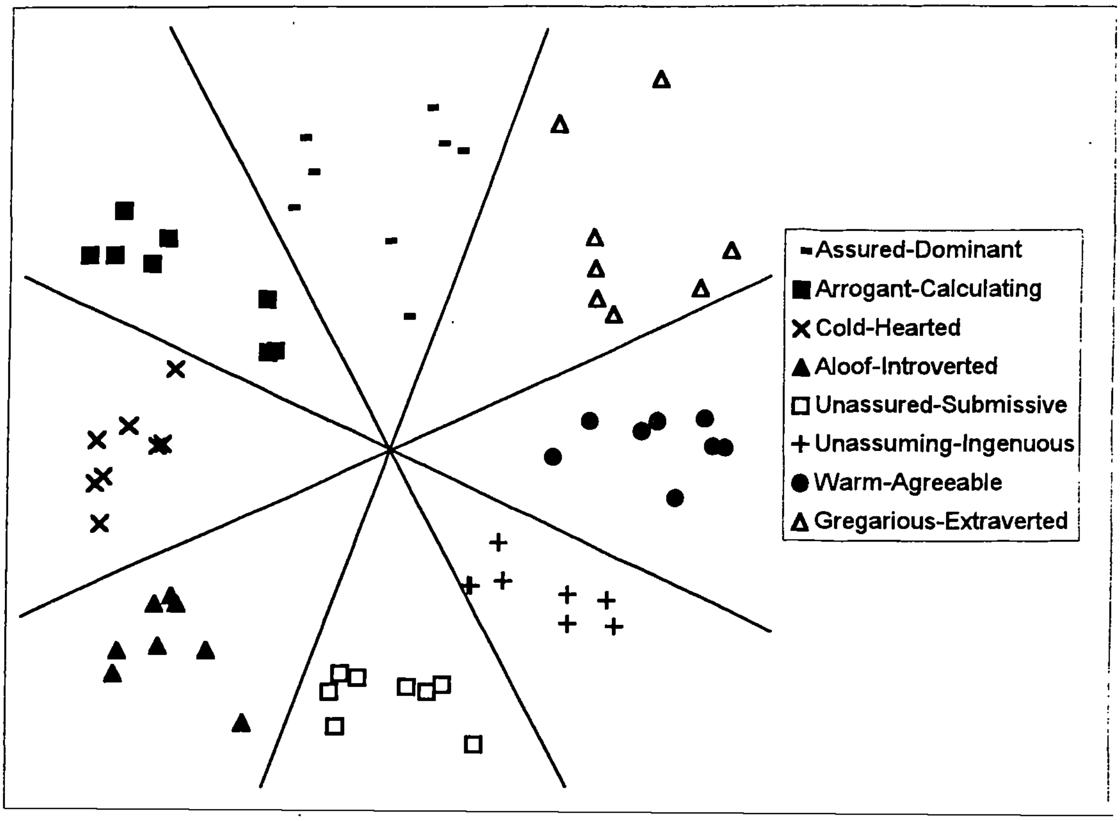


Figure 10.2. Regional interpretation of polar facet in factor loading plot of Wiggins personality scale

bipolar opposite Cold-Hearted which has half the spread in distance to the centroid, as measured by the standard deviation, but more spread in angular separation. The extra component in the Warm-Agreeable scale is suggested to be intensity, indicating involvement with the theme or type which is not evident from the factor matrix. Figure 10.3 shows the hypothesised dimension in the factor loading plot.

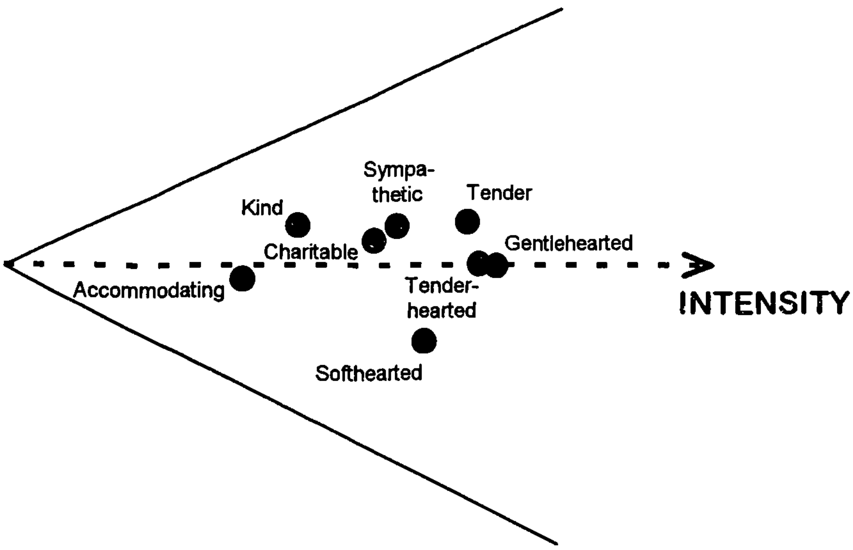


Figure 10.3. Detail of Warm-Agreeable scale and hypothesised Intensity component from factor loading plot of Wiggins personality scale

It was not possible to test this hypothesis using SSA-I, since the original correlation matrix was unavailable. However, this extra component of structure is shown to be hidden by factor analytical research in the context of criminal actions in chapters 12 and 13.

Discontinuities in non-metric MDS analyses of the structure of personality questionnaires and inventories can also be observed elsewhere, again contradicting the clear-cut boundaries implied by the exclusive categories of factor analyses. These discontinuities can be seen even though the data for the non-metric analyses were gathered on instruments constructed using factor analysis.

For example, Green and Walkey (1980) performed MDS on the Eysenck Personality Inventory (EPI), created by Eysenck with second-order factor analysis. This consists of 57 items which were then classified as two dimensions of Extraversion and Neuroticism, and the lie scale. Green and Walkey commented that these dimensions only emerged after second or third order factor analyses had been carried out, but that in the MDS space the three types of items 'would be indicated by the presence of three independent clusters of items' (Green and Walkey, 1980, p. 157). Showing their factor analytical inclinations, Green and Walkey also presented the coordinates of the 57 items. The two-dimensional MDS space is reproduced in Figure 10.4.

Green and Walkey concluded that the dispersal of the scales show what much of the research they presented had suggested, namely that Extraversion is less tightly bonded than Neuroticism, and the Lie Scale is a 'separate cluster'. In other words, two spatial dimensions represent structure just as clearly as a two factors from a third order factor analysis; i.e. a factor analysis done three times over on the same data. On the principle of parsimony alone, the MDS solution must surely be preferred. Additionally, the MDS plot should be preferred in terms of adequacy of representation. This is because the successive approximations and ignoring of variance necessary to get to a third order factor analysis reduces the accuracy of the residual correlation matrix carried over.

Furthermore, what was not explored by Green and Walkey was that these items form a virtually perfect polar facet which could be partitioned just as the Wiggins' items were in Figure 10.2. Green and Walkey did mention in passing the

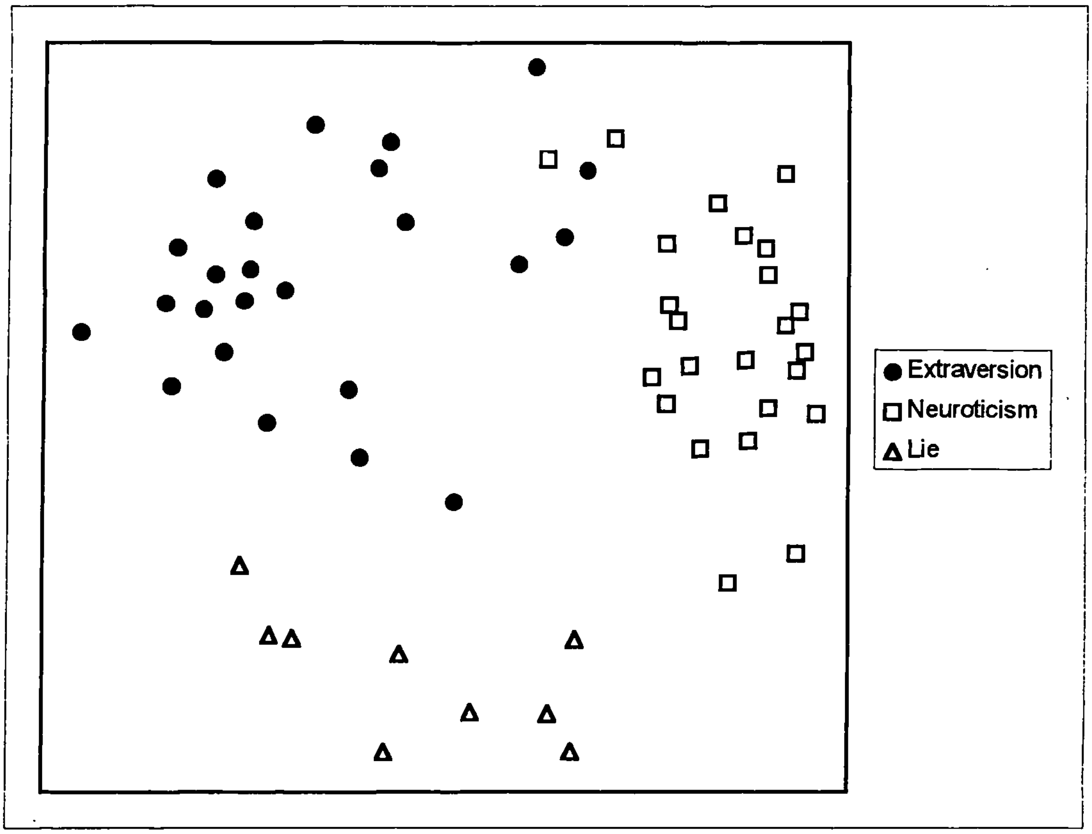


Figure 10.4. MDS of plot of EPI personality scale, reproduced from Green and Walkey (1980)

'essentially triangular configuration' (Green and Walkey, 1980, p. 159) as a reason for two dimensions rather than one dimension. Indeed, the variation in the items' distances from the centroid suggests again some extra component not mentioned, the intensity component explored in chapter 13.

However, having recognised the validity of the MDS solution, Walkey and Green (1981) compared this MDS plot with a factor loading plot. Walkey and Green concluded that their two factors extracted in a first order factor analysis were better representations of the 'cluster' structure of the three items types than the third order factor analysis of Howarth (1976) on the EPI. In other words, clear lines of discontinuity were being sought and found by reference to the factor loading plots rather than the more amorphous MDS plots.

As Canter stated 'It seems that orthogonal dimensions and linear principal components models are being used because they give answers, not because they necessarily give the most valid or the most psychologically meaningful answers.' (Canter, 1985, p. 3)

Summary of Chapter 10

Factor analysis was introduced as a method of multivariate analysis sharing common goals but different means with SSA-I. The most important differences were pointed out to be the creation of the triangular input matrix and the nature and interpretation of the output representation. In factor analysis, the representation is determined by the factors. Issues over the number of factors and the type of factor analysis were shown to be hard to resolve. Thurstonian 'simple structure' was implied to refer more to representation than to structure, compared with the meaning and distinction given in the present work. Factor analytical studies of personality and also intelligence were criticised, and a fully geometric alternative representation was suggested to recover more of the structure and show how personality theories disagreeing on the 'number of factors' actually only differ in representation but not necessarily structure. It was proposed that a component of intensity or involvement was hidden by the factor analytical dimensional interpretation, which is developed in later chapters.

Chapter 11
structural explanations of juvenile delinquent and criminal action

Versatility, Specialisation and Methodology in Juvenile Delinquency Research

It was suggested in the previous chapter and elsewhere (e.g. Canter, 1985; Block, 1995; Maraun, 1997) that the representation of factor analyses as linear combinations of items automatically implies structure without the possibility of falsification - namely, orthogonal or pseudo-orthogonal dimensions. In other words, conclusions about structure in factor analysis is constrained by the factor analytical representation and interpretation. If the structure of a content universe were to consist of orthogonal or pseudo-orthogonal dimensions then the factor analytical method would reveal this, given certain preconditions such as respondent to item ratio, etc. (e.g. Kline, 1994). However, if the content universe was not structured in this way then the conclusions from that representation would be misleading. The issue of restricted representation with factor analysis is also especially relevant in the analysis of the structure of delinquent and criminal actions by juveniles. In this chapter both factor analytical and non-factor analytical studies are examined, and their findings related back to the methodology.

Broadly speaking, researchers have tended to posit one of two structural explanations for the criminal actions of juveniles. These positions can be categorised as versatility theories and specialisation theories (e.g. Klein, 1984; Farrington, Snyder and Finnegan, 1988; Osgood, Johnston, O'Malley and Bachman, 1988). Versatility, or the 'cafeteria' model, implies that juveniles perform delinquent actions by picking and choosing with no fundamental pattern. Specialism implies that juveniles only commit certain types of crime and rarely commit other types.

If the structure of juvenile delinquent and criminal action truly were one of versatility and generalism, it is suggested that the factor analytical representation would not reveal this. Instead, the factor analysis would create representation that would impose an explanation of specialism. Therefore, it is argued in this chapter that structural hypotheses of generalism cannot be tested with factor analytical representations, and an alternative representation is required. In effect this would be equivalent to scaling as a technique (Coombs *et al.*, 1970) whereby there would be no reliable indication of departure from a hypothesised model, unlike the alternative scaling as a criterion.

Within the two explanations of versatility and specialism, there is a mixture of several elements that define the theory. In other words, the explanations can be 'facetised' (Borg and Shye, 1995) to illustrate comparabilities and incomparabilities.

The Facets of Criminal Action in Male Juveniles

As mentioned before, there are a number of different variants within both versatility and specialism theories. Looking at versatility firstly, the 'nullest' hypothesis of structure (cf. Cliff, 1973, p. 484) for criminal action in juveniles might be that there is no structure. Extreme versatility would be juveniles choosing crimes type in a random way and committing crimes without any pattern whatsoever. The cause of offending would be the random availability of targets and the individual not 'seeking' a target. Thus if this were true, the crime committed at time t would be independent of the crime committed at time t - 1, which would be an extreme situationist explanation. The choice to commit the crime may or may not some sort of 'rational choice' (e.g. Clarke and Cornish, 1985). In terms of the Markov chain of Wolfgang, Figlio and Sellin (1972) there would be no transition in the 'nullest' hypothesis.

The existence of this extreme versatility structure would be found in the repeated finding of zero or approximately zero correlations between different criminal actions across a sample of juveniles. Statistically significant correlations would be found where samples of crimes did not obey the theory of random distributions, i.e. 5% of the time by convention.

Alternatively, some sort of social learning process could be involved, revising the individual's rational choices by experience. This would therefore imply some conditional dependency between present and prior offending patterns. Thus experience of committing a crime at time t would be dependent of the crime committed at time t - 1. A longitudinal study would reveal patterns dependent upon time, such as a particular crime type being committed and making it more likely to be committed in the future. This was found by Farrington (1994, p. 527) as part of 'persistence' of commission of crime types, as opposed to desistance from crime as proposed above. A cross-sectional 'snapshot' of offending at any single time would reveal this correlation only if the sample was asked about when the crimes had been committed, which would add error not found in longitudinal. Farrington (1973) stated, however, frequency and seriousness of offences admitted seemed to have little to do with predictive validity, thereby undermining part of the advantage of the longitudinal studies.

To summarise, then, versatility explanations contain several parameters on which to vary, namely progression, desistance and change over time. With the exception of the nullest hypotheses of no structure, it can be seen that under the remaining non-random versatility explanations it is proposed that there is essentially one type of offending. The random versatility explanations would suggest there was no type. If there were to be a single type, the pattern of offending within this type may involve escalation and juveniles may persist with all delinquent actions committing more and more, or may desist the lesser crimes as the progression increases.

By contrast, specialism theories have constructed separate dimensions or types of delinquency and hence delinquent. These dimensions have been identified typically as independent, containing high within-type variance and low between-type variance. Within each type - under specialism - there can be any or all of the variation suggested under versatility.

These diverse elements may be integrated as facets in a general Mapping Sentence (MS), which is proposed to structure the universe of explanations of male juvenile criminal actions. The MS is given in Figure 11.1.

The extent to which an explanation of male juvenile criminal actions posits type(s)

Sense

$$\text{showing} \quad \begin{bmatrix} \text{Persistance} \\ \text{Escalation} \\ \text{Consistency} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{Great deal} \\ \dots \\ \text{Not at all} \end{bmatrix}$$

Figure 11.1. Mapping Sentence showing elements in explanations of male juvenile criminal actions

Following from Shye and Elizur (1994, p. 33), facets which are ordered from high to low with respect to a common meaning - i.e. with a common meaning range - can be more elegantly represented as items of the same facet, as opposed to separate domain and range facets. In this MS, each type of behaviour (e.g. drugs, violence) could be classified by the same properties, the 'Sense'. The common meaning for each of the structuples in the Sense facet is scalability over time, namely uniformity and regularity in response patterns replicated by individuals. Thus persistence is a scaleable property in the sense that individuals do not desist from any crimes as they progress along a scale, which might happen if there were some sort of 'interference'. Escalation is a scaleable property, such as an increase in the seriousness of the offence or in the distance travelled to commit the offence. Consistency is also scaleable since it states that the same patterns are repeated over time and are not subject to arbitrary or deliberate interventions or effects.

An explanation which proposed multidimensionality of types could have each type classified differently - whether or not the types are scaleable in the same senses is an empirical question. If there were only one type, then these senses are applicable only once. Although there is no requirement for a MS to incorporate a null hypothesis *per se*, the 'no type' explanation of no pattern whatsoever would be scaleable in no sense.

One important point to note is that the design of the study is extremely influential in the possible structuples that may be examined. That is to say, a cross-sectional design will not allow exploration of consistency or change over time. Similarly, desistance and persistence can only be inferred from respondents, since the questions will be asking typically 'Have you ever done X?' rather than 'How recently have you done X?' The latter question would give an idea of desistance or persistence, since the last commission of some crimes would be a long time ago while the last commission of others would be more recent. A time frame would need to be specified to give boundaries to desistance or persistence, such as a week.

However, the latter question brings with it a new set of extrinsic biases to the study, specifically related to the respondent's memory. The problems of 'telescoping' the date of last commission backwards or forwards by the respondent perhaps due to the personal significance of the event (Hough and Mayhew, 1983) may obscure persistence patterns and mislead the research.

Other retrospective ways to measure consistency and desistance are also likely to bring biases. Although not juveniles, Peterson, Pittman and O'Neal (1962) suggested that males could be mainly classified into person or property offenders, on the basis of arrest records which furthermore showed that there was little crossover before and after 30 years of age. This was taken to show consistency, though there were biases in the fact that arrest records were taken and that the men in the sample were self-selected by being arrested at the age of 40 or more, suggesting these were long term persisters. These biases can be reduced with careful longitudinal studies, such as the Cambridge study of East End boys growing into men (e.g. Farrington, 1994). These studies also give an idea of the consistency over time.

It can be readily seen in the above MS for criminal actions in juvenile males that several facets are unexpanded, or degenerate. In particular, the research has typically focused around on juvenile or young adult males, though there are background gender and age to consider. Furthermore, the population is important since many studies use only students while other studies use only convicted males - two populations which would not have much overlap. The comparison of regional structures between these unexplored facets of criminal actions could present interesting similarities.

Factor Analysis and the Structure of Male Juvenile Delinquency

The previous chapter suggested that variance is systematically removed from factor analysis in two ways. Firstly, the extraction of factors or components takes the mathematically maximum variance initially from the structure sampled in the correlation matrix, inevitably ignoring some useful common variance. Secondly, the representation in factor analysis is determined by the factor dimensions of the space, which are then taken to indicate the structure of the domain. Even though items have loadings on many factors, only the highest factor loading is interpreted, which implies that factor analytical studies will tend to suggest distinct non-overlapping types. Indeed, many factor analytical studies have proposed strong orthogonal factors of delinquency.

For example, Quay and Blumen (1963) used factor analysis to propose 5 factors of delinquency, using case files from court hearings. However, the use of factor analysis and the Pearson's PMCC coefficient were inappropriate given the nature of the data and the culling used to reduce the set. Despite having rotated five factors using the orthogonal Quartimax routine, even the best factor loadings were low and collinear with other factors. In fact, 5 of the 13 variables did not load above 0.3 on any item. These 13 variables were originally drawn from 50 items, since 37 had low inter-rater reliability of finding them in the samples' court case files. (This contrasts with chapter 7 where meaningful non-metric structure was shown even with some poor reliability in some items.) Quay and Blumen's factors were also suggestive of bloated specifics, and a simple calculation of the squared summation of factor loadings given by Quay and Blumen (1963, p. 276, table 1) revealed eigenvalues of 0.65, 1.29, 0.58, 0.41 and 0.31 for factors 1 to 5; *viz.* only one eigenvalue above unity. Despite this poor design, factor analysis was used to give five factors which were each interpreted as dimensions.

Similarly, Heise (1968) used a sample of 753 students asking 30 mainly sub-legal or anti-social delinquency items, including one item of 'Eating without washing first'. Nevertheless, Heise proposed no less than 11 orthogonal factors of different types of acts. Yet most items loaded on factors poorly, with no loading greater than 0.6 and many later factors being bloated specifics with only one item. As Kline stated that 'Factors loading on only a few items (four or five) are almost certainly worthless.' (Kline, 1994, p. 175)

Gibson (1971) performed factor analysis on the Cambridge sample of East End boys at age 14-15 and extracted 12 principal components with eigenvalues greater than unity in 31 items. These 12 were rotated to simple structure using the oblique Promax method. These were then factored again to get three second order factors, which contained some very favourable loadings but some very unfavourable, such as items loading less the 0.3 on all three higher-order factors. Indeed, it was admitted that the third factor was 'difficult to interpret' (Gibson, 1971, p. 8). Gibson did point out an intensity component, namely the seriousness of the acts. It was suggested that 'the more serious offences are denied by a greater percentage of boys, and have higher loadings on the first component' (Gibson, 1971, p. 8).

Short, Tennyson and Howard (1963; cited by Nutch and Bloombaum, 1968) found five factors when 37 items were rotated to simple structure with the orthogonal Varimax procedure after a principal axis extraction. The factor loadings were far higher than with Quay and Blumen (1963), since the data were collected from direct observation of the Chicago street gangs in the sample by Short *et al.*, a technique not used elsewhere. The factor structures were not comparable between Quay and Blumen and Short, Tennyson and Howard. It should be noted though that the Short, Tennyson and Howard items contained some distinctly dated and some dubious delinquent items, such as common-law marriage (i.e. co-habitation), illegitimate children, team sports, and even work experience.

Indeed, one factor analytical study Klein (1971; cited by Klein, 1984) even indicated the nullest hypothesis of no structure (i.e. complete versatility). Here a matrix of correlations between arrests for different items in delinquent gangs showed that 5.6% values were statistically significance at the $\alpha = 0.05$ level. Klein suggested this was due to random arrest patterns by the police, a reasonable explanation and one already explored in chapter 1. Nevertheless, five factors could be extracted from this 'random' matrix. Though they were deemed uninterpretable, this is something which Thurstonian simple structure is supposed to prevent.

Of self-reported studies of juvenile crime, a qualitative meta-analysis by Klein (1984) indicated that specialism tended to be indicated by factor analytical studies. However, not all factor analytical studies showed this pattern, since factor analytical studies without self-report data showed greater evidence against specialisation. Klein (1984, p. 188) suggested 'methodology alone seems to be an unlikely candidate' in explaining the pattern. Overall, though, 21 of the 33 studies showed evidence of versatility with 8 giving mixed results, but caution was added as to the methodology of many of these studies.

Although items used in the studies varied considerably and several had distinctly American overtones with 'status offences' (e.g. parole violations), some replication was shown by these studies. Klein suggested there were five factors repeatedly, found though these were 'a bit arbitrary'. The factors were: assault, theft, auto offences, drug offences and status offences.

Guttman Scaling and the Structure of Male Juvenile Delinquency

There is some empirical evidence for Guttman scales in male juvenile criminal actions. As mentioned in chapter 3 and later expanded on in chapter 14, a Guttman scale is found where items are ordered according to a common meaning, namely a single dimension. This is a particularly strong hypothesis of structure.

A single general type of delinquency showing escalation in a Guttman scale was demonstrated by the cross-sectional study of Nye and Short (1957). This study was on seven criminal and anti-social variables of all different types. Four levels of response were used originally: 'not at all', 'once or twice', 'several times' and 'very often', but the last two were collapsed to give three levels. Unidimensionality was found on the items and individuals, creating a Guttman scale with good reproducibility which was replicated across two samples of young males. However, for males in the sample older than 16 years the scale was less reproducible and the response range was collapsed more often. Most non-delinquent juveniles reported infrequent commission of the anti-social activities at lower levels, with a exponential decrease in numbers of respondents reporting the more serious activities at higher frequencies. The opposite pattern was found in a sample of borstal males, with a large proportion having committed most of the anti-social and criminal acts many times. This would indicate that 'delinquents' and 'non-delinquents' in the eyes of the criminal justice system were different in degree but not type. However, the issue of self-selection of the 'delinquents' must not be ignored, and two of the variables included defiance against parents and [hetero-]sexual relations, things which might now be considered part of growing up in 16 and 17 year olds. It must also be noted that the Guttman scaling technique was used almost item analytically - namely the creation of rather than testing for scales.

Scott (1959) implicitly criticised Nye and Short (1957) who had gone from 21 to 9, to 7, to 11 items in the search for high reproducibility for each of their samples and for the combined sample. Scott instead suggested that two types of delinquency in two Guttman scales could account for the variation better than the one. Scott (1959, footnote to p. 240) noted that Guttman himself had suggested the necessity of two scales to prevent the item analytic practice of Nye and Short (1957). Using a sample of criminology college students, Scott asked 15 questions with a four level response range of 'never', 'once', 'sometimes' and 'often'. Two reproducible scales were obtained containing items of theft against unknown persons and theft against known persons. The first was far more commonly reported and scaleable. However, this study only used items involving theft rather than criminality more widely, *viz.* a distinct subcontent universe, and Scott had used college students in his sample.

In another cross-sectional study, Arnold (1965) found three types of juvenile criminality in Guttman scales of theft, nuisance and violence offences. Again theft was highly reproducible, then nuisance then violence. Committing more serious acts was associated with committing less serious acts less frequently, suggesting escalation with desistance in the three scales (i.e. types). However, to obtain these scales it was necessary to dichotomise responses, asking in effect 'Have you ever committed X?' This is also found to be necessary in chapter 14 later.

Hindelang (1971) reviewed the evidence and found that most delinquency and criminality items were significantly positively correlated. The types found using cluster analysis included general delinquency, violence and drugs, which were all intercorrelated. Hindelang suggested that the pattern of results was of generalised delinquency rather than specialised. It was also noted that broadly there was no difference in type of offending between male and female juveniles, but invariably males were more frequent offenders and offending more seriously.

However, the disadvantage with finding several Guttman scales of delinquent and criminal actions is that there is not much indication of how the different scales are related. The scales tend to be interpreted as if they were independent principal components extracted from the data. For example, even though Arnold (1965) found three types of juvenile criminality, the scales proved to be highly correlated, suggesting that while the offence types can be scaled separately the offenders increased their offending seemingly in synchronisation. In other studies, the structure revealed may be limited by the representation, since a structural hypothesis *interrelated* Guttman scales is not tested. This is explored later in chapter 14.

MDS and SSA-I in the Analysis of Juvenile Delinquents

A strong criticism of inappropriate coefficients in research was made by Braithwaite and Law (1978), who rejected the use of principal components analysis on self-report questionnaires on the basis of the unreliability of the data. Non-metric structural hypotheses would not require these numerical interval restrictions. Braithwaite and Law were equally critical of the way Guttman scales had been 'forced' onto delinquency data, and suggested that high reproducibility was down to uneven frequency distributions on the items.

Instead, Braithwaite and Law (1978) performed several non-metric analyses were performed on their data, including SSA-III - the 'non-metric factor analysis' - and MSA-II as well as the more familiar SSA-I. The SSA-III showed a tendency for generalisation on one factor, though this was not totally clear and four factors were rotated with the orthogonal varimax technique. The factors were vandalism, trivial delinquency, drug use and 'vehicle theft uninterpretable'. This latter factor contained a high loading with marijuana which Braithwaite and Law suggested made it hard to interpret.

The MSA-II solution of respondents in three dimensions produced three 'clusters' of individuals. Braithwaite and Law (1978) characterised these as low frequency low seriousness delinquency, high frequency high seriousness delinquents and high frequency low seriousness delinquency. However, the MSA-II space was filled continuously with individuals rather than in discrete chunks and Braithwaite and Law did also recognise there was a continuum of seriousness across the plot.

Braithwaite and Law (1978) went onto identify six extremely tightly bonded 'clusters' of items in a three-dimensional SSA-I, including four double item clusters which in factor analysis might have been termed bloated specifics. Furthermore, three items were left unclassified in the SSA-I space. Unfortunately, having rightly noted the advantages of MDS and especially SSA-I over the principal components and factor analyses, Braithwaite and Law failed to exploit these advantages to the full with a better regional interpretation. Furthermore, the 1-D SSA-I solution was degenerate with all items placed on a single point; the 2-D SSA-I had alienation 0.23 and the 3-D SSA-I had alienation 0.10, which was high given there were only 17 items and the data were from questionnaires. It is possible that the 3-D SSA-I was a local minimum, especially if the starting dimensionality was 1 and 'dimensional slurring' carried forward the degenerate 1-D solution (Lee and Canter, submitted). Furthermore, Braithwaite and Law did not fully address their own criticism of the nature of response range data, since the response range for Braithwaite and Law's 14 of the 17 items was 'not done', 'done once' and 'done two or more times', with three items being dichotomous i.e. 'not done' and 'done'.

Nutch and Bloombaum (1968) observed that in a comparison of SSA-I with the factor analysis of Short, Tennyson and Howard (1963; cited by Nutch and Bloombaum, 1968) that essentially the same structure was to be found, though the SSA-I plot put the more questionable 'delinquency' items into a more sensible 'legal' region of the space, which were more readily contrasted with the obvious legal and sub-legal items. Nutch and Bloombaum did note a 'dimension of seriousness' in the SSA-I, though without exactly specifying where it was on the plot.

Shortcomings of Previous Research

Each of the explanations of juvenile delinquency and criminality essentially argues that the structure of delinquency is different, whether factor analytical, Guttman scaling or non-metric MDS. It has been suggested that the structure concluded by some studies has been shaped by the representation used for the data. The main forms of representation have been factor analysis and Guttman scaling. However, other considerations have also played a part in the conclusion of structure.

Studies purporting to be of juvenile delinquency have used a variety of institutionalised and non-institutionalised individuals; a varying age range; items of extreme triviality and dubious illegality; different recording methods from self-report to direct observation to official records; response ranges from discrete dichotomous to continuous polychotomous; time frames from incidence ever to prevalence in the last year; and design of cross sectional to longitudinal.

The key variant that predicted whether factor analytical studies gave specialism or versatility was suggested by Loeber and Waller (1988) to be the nature of the item response range. It was pointed out that self-report factor analytical studies with few response categories produced versatility structures, while those with more response categories produced specialism structures. Evidence used to back this up came from schoolboys of ages 9, 12 and 15, thus including non-juvenile males. Up to 13 factors were extracted, though only 8 were rotated and used in the analyses. Respondents were asked to name how many times in the last year the offence had been committed. If it was committed more than 10 times in the last year, then the options were 'once a month', 'once every 2-3 weeks', 'once a week', '2-3 times a week', 'once a day' and '2-3 times a day'. However, the response range in Loeber and Waller (1988) was taken as continuous interval level data, and followed a dubious use of the Pearson's PMCC on these non-interval level data. Therefore even this study was flawed.

What is needed is a method and data set that overcomes these difficulties and which can test both specialism and versatility structures simultaneously if necessary. As Farrington, Snyder and Finnegan (1984, p. 483) suggested from their exhaustive longitudinal study, there was 'a small but significant degree of specialization superimposed on a great deal of versatility'. The method must also be able to test for evidence of escalation in the sense of increased seriousness of offending, and desistance with less serious crimes being committed less often. Farrington, Snyder and Finnegan observed that theories should incorporate both general delinquent versatility and specialisation in some offences, explaining the reasons for the split.

Chapter 12 introduces the Youngs data set on juvenile criminal actions which overcomes many of the difficulties, but also shows that factor analysis is not adequate to test these data.

Summary of Chapter 11

Two categories of explanation of the criminal and delinquent actions of juveniles were introduced: versatility and specialisation. The parameters on which these explanations can vary were described as being persistence, escalation and consistency. A tentative Mapping Sentence (MS) of juvenile criminal actions was presented. The various types implied by the MS of explanations were explored, and it was pointed out that factor analytical studies tend to describe several distinct types of crime with no mention of escalation or persistence. Studies using Guttman scales tend to find fewer types but with evidence of seriousness of crime and escalation. The need for good methods and good data was emphasised, since some studies have used student samples and dubious items.

Chapter 12
structure and representation in a factor analysis of juvenile delinquent and criminal action

Versatility, Specialisation and Methodology in Juvenile Delinquency Research

In the previous chapter, it was suggested that the factor analysis of the self-reported delinquent and criminal actions of juveniles would tend to create several pseudo-independent types of offence style. It was earlier suggested that this kind of finding would be due to restrictions deliberately placed on the factor analytical representation of the structure sampled in a correlation matrix. This chapter presents a new and valid data set with which to test whether this is the case.

The Youngs Survey of Juvenile Delinquent and Criminal Actions

The survey data of Youngs (1994) are especially useful and valid since they overcome many of the methodological problems highlighted in other data sets on juvenile delinquent and criminal actions in the previous chapter.

The advantages include using an extensive self-report questionnaire, rather than working from court case files (e.g. Quay and Blumen, 1963) and using natural language rather than legal terminology (e.g. Farrington, Snyder and Finnegan, 1984). The items in the questionnaire ranged from highly serious criminal activities through to minor incivilities, rather than just anti-social and sub-legal items (e.g. Heise, 1968). The large sample featured a range of young people with many different convictions, on remand or from the general population, rather than just students (e.g. Scott, 1959).

The details of the data set are as follows:

- **Questionnaire**

The questionnaire contained questions on personality, background and biographical details, as well as items asking about criminal and delinquent actions. For the purposes of this study, mainly the latter are used. In total, 55 items tested respondents about the commission of the criminal acts and contained a five level response range of 'never', 'once or twice', 'a few times (not more than ten)', 'quite often (between 10 and 50)' and very often (more than fifty times)'. The responses were also dichotomised to reflect participation in the acts ever i.e. 'have never done' and 'have done', as explored later.

These 55 questions asked in natural language about particular criminal acts that the sample had committed. Examples of the statements included for example 'Used a club, knife or other weapon to get something from someone' and 'Taken a car belonging to someone you didn't know for a ride without the owner's permission'. 10 items were removed on the basis that they were extremely low frequency (e.g. rape and obscene phone calls) or because they were not strictly illegal *per se* (e.g. begging and acting as a look out). The remaining 45 items were all illegal, and ranged from extremely trivial to very serious, from 'Not gone to school when you should have been there?' to 'Used or carried a gun to help you commit a crime?'

Therefore this is a good self-report questionnaire which also asked for details of previous convictions - those crimes detected and prosecuted by the authorities.

- **Sample**

The questionnaire was given to 207 males to complete anonymously. The sample comprised of 13.5% from the general population, 25.1% from a remand centre, 43.0% from a Young Offenders Institution and 18.4% serving probation orders. Clearly the sample included a large proportion of 'proven' juvenile delinquents, as well as those awaiting sentencing and a small proportion of people not expected to be serious offenders.

The breakdown of court convictions in the sample is given in Table 12.1. Thus in the sample 90.3% reported having criminal convictions, with the median number of convictions was 10. Only 9.7% stated that they did not have a criminal record. Of those with convictions, the mean age of first conviction was 14.1 years, with the median age of first conviction being 15 years.

Although the age range of the sample was from 14 to 28 years, the mean was 18.8 years, with median 19. The sample was skewed towards the younger ages, with

Number of Convictions	Number of Subjects	Percent of Total Sample
0	20	9.7
1-10	97	46.9
11-20	40	19.3
21-30	22	10.6
31-40	6	2.9
41-50	6	2.9
> 50	16	7.7
	207	100

Table 12.1. Court convictions in Youngs data set of male juvenile criminal actions

93.6% being 20 years or less. Of those who had a previous conviction or convictions, the mean lag between age at first conviction and age at time of completing the questionnaire was 3.9 years, median 4, with a range of 0 to 12. Overall there was no correlation between age and number of convictions, Spearman's ρ = 0.069, p = 0.376, even when those few who did not report any previous convictions were excluded from the analysis.

Therefore the sample contained typically a 19 or 20 year old well known to the police for at least 4 years with several previous convictions.

- **Results of questionnaire**

The range of criminal behaviours reported on the questionnaire was from 0 to 43, with only one individual reporting no offences. The breakdown of the responses is shown in Table 12.2.

Number of Criminal Behaviours	Number of Subjects	Percent of Total Sample
0	1	0.5
1-10	16	7.7
11-20	34	16.4
21-30	73	35.3
31-40	72	34.8
41-45	11	5.3
	207	100

Table 12.2. Total criminal behaviours reported

This range was approximately normally distributed with mean 26.9 and standard deviation 10.0, as Figure 12.1 shows with a normalised curve superimposed.

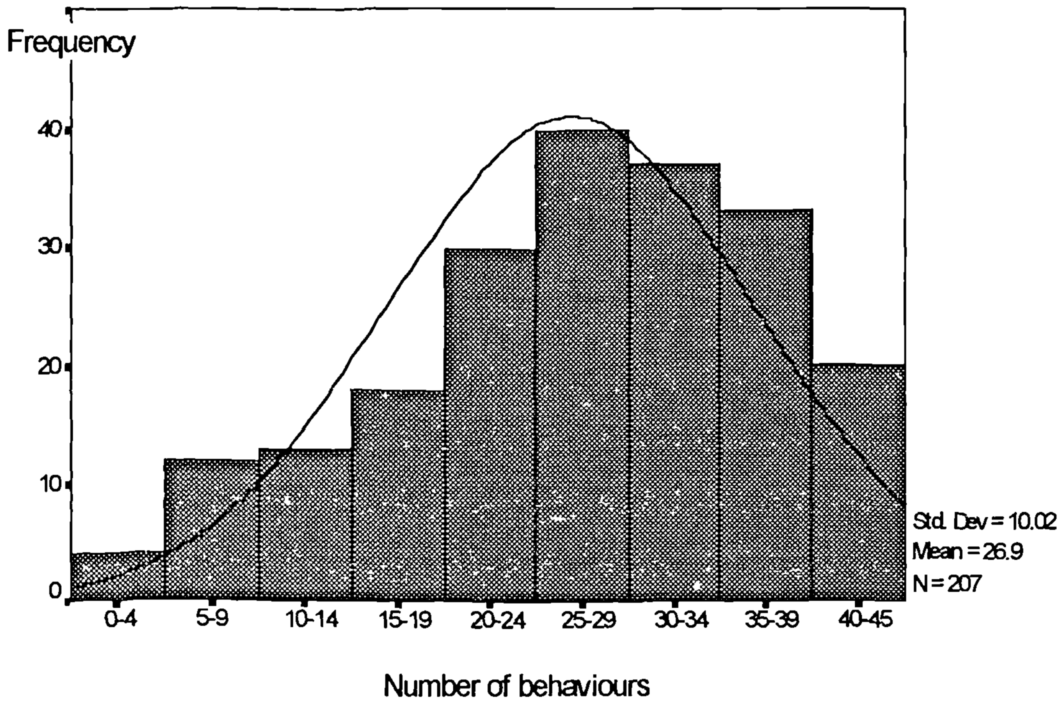


Figure 12.1. Graph of total criminal behaviours reported with normal distribution superimposed

The Spearman's ρ correlation between the sum of criminal actions in the questionnaire reported and the respondent's number of previous convictions was +0.363, p = 0.000. This suggests that the questionnaire was a strong predictor of offence history. Since there was no correlation between age and number of criminal behaviours (ρ = 0.036, p = 0.614) this would indicate that the correlation between self-report and convictions was not accounted for by age.

- **Responses to items and missing data**

On average, each of the 45 items contained on average 5.3 missing responses, with a range of 2 to 10. Of the 207 subjects, 56 failed to respond to at least one variable. Clearly then the issue of missing data was important in this data set.

Those in the sample who had left any answer blank tended were compared to those who did not leave any answer blank. The sub-samples were of similar age (18.9

years and 18.8 years, missing to non-missing); were similar in proportion having convictions (89% and 91%) with similar age at first conviction (14.2 and 14.0); had a quite close median number of convictions (12 and 8.5); but the non-missing sub-sample described themselves as white slightly more than the missing sample (82% and 91%). Non-responders also tended had no GCSE's more often (68% and 59%). However, the hypothesis that disadvantage and illiteracy played a large part in the missing data considering must be tempered with the fact that such a high proportion still had no GCSE's yet completed the questionnaire fully.

Table 12.3 gives the proportions in the sample reporting involvement with the various acts. It also shows the number of missing values for that question.

The Youngs delinquency data set are eminently suitable for the comparison of representation of juvenile delinquent actions by factor analysis and SSA-I, and the consequent structural conclusions drawn. These are the ordinal or dichotomous response ranges with Pearson's PMCC (Braithwaite and Law, 1978; Loeber and Waller, 1988); the inherent dimension of seriousness underlying the data (e.g. Nutch and Bloombaum, 1968); and the structural hypothesis integrating specialism and versatility model of juvenile delinquent and criminal actions (Farrington, Snyder and Finnegan, 1988). A variety of factor analyses were run on the data under different rationales and conditions to test the structure of juvenile delinquency and whether the structure was of versatility and specialism or co-existance.

Empirical Study 12.1: A Factor Analysis of Juvenile Delinquency and Criminality

- **Creating the correlation matrix**

The usual coefficient with factor analysis is Pearson's PMCC (Gorusch, 1988). This was the case in the literature of factor analyses of juvenile delinquent and criminal actions (e.g. Quay and Blumen, 1963; Loeber and Waller, 1988). It was already noted in chapter 1 the uncertainty over the nature of information from rating scales, even

Item description	Involvement (%)		Missing Cases
	Yes	No	
Played truant	92	8	2
Used cannabis	88	12	4
Refuse coop. police	86	14	2
Bought stolen goods	86	15	7
Acted rowdy	85	15	5
Been drunk under 16	81	19	10
Travelled without ticket	79	21	3
Dropped litter	79	22	7
Fireworks in public	78	22	8
Gang fights	78	22	6
Break into house to steal	77	23	5
Broken into car to steal	76	24	7
Insulted stranger	76	24	5
Shoplift goods <£5	74	26	6
Not returned XS change	74	26	3
Shoplift £10-£100	71	29	5
Used barbs./amphet.	71	29	6
Resisted arrest	70	30	5
Broken windows	68	32	6
Stolen bike	67	33	5
Gone joyriding	63	37	4
Sex in public place	63	37	4
Stolen purse/wallet	62	38	6
Fought stranger	62	38	2
Beaten up	62	38	8
Cheat at school	60	40	5
Stolen car part	60	40	5
Driven while drunk/drug	59	41	6
Carried weapon	58	42	5
Used fake money	56	44	5
Shoplift goods >£100	55	45	5
Sniffed glue	51	49	2
Threatened violence	51	50	7
Used ecstasy	50	50	4
Pulled weapon	49	51	4
Dialled 999 as joke	44	56	6
Forged cheque	44	56	5
Stolen cash fr. home	43	57	5
Enter+damage building	42	58	5
Used heroin/cocaine	34	66	6
Used weapon	33	67	9
Mugging	27	73	7
Public disturbance	27	73	6
Set fires	25	75	5
Carried gun	24	76	4

Table 12.3. Frequencies of criminal behaviours reported in Youngs data set Involvement is expressed as a proportion of valid cases i.e. (207 - number missing)

where the response range is symmetric and regular. The difficulties arise from the PMCC requiring among other things a bivariate normal distribution and an interval scale of measurement, like other parametric measures (von Eye, 1988).

However, the use of the PMCC on the present data set was additionally problematic due to difficulties with the full five response range, with a range of 'never', 'once or twice', 'a few times (not more than ten)', 'quite often (between 10 and 50)' and very often (more than fifty times)'. The range clearly was ordered but not numerical (i.e. ordinal not interval) or symmetric and regular.

To overcome this difficulty, the response range can be dichotomised to create the meaning 'Have you ever been involved in this activity', with response yes or no. The use of the PMCC will still then be justified by using its dichotomous equivalent. This would also partly answer the criticism of Micceri (1989; cited by Macdonald, 1997) about the widespread inappropriate use of normality assumptions in published large sample psychometric research.

Similarly, the issue of missing data may be resolved in a number of ways, such as interpolation of mean, pairwise exclusion and listwise exclusion. For interpolation of the mean, the missing value is simply replaced with the average value for that item. This was not considered in these analyses since this may introduce extrinsic error to distributions that may already be skewed, especially where many missing items were found. Pairwise exclusion of missing values excludes that particular value from contributing to the analysis. This may result for example in the values of correlation matrices being derived from different degrees of freedom (i.e. different numbers of respondents), due to different numbers of valid responses and meaning that conclusions about statistical significance must take the differing critical values into consideration. Listwise exclusion of missing values - the SPSS default for factor analysis - removes all those cases from the analysis where *any* value is missing. This procedure would reduce the juvenile sample size by 27% from 207 down to 151, since 56 respondents failed to answer to one or more questions. This would bring the number of respondents well below 200 minimum given by Kline (1994) and very close to the approximate minimum of 150 calculated from the formula of Baggaley (cited by Child, 1990, p. 115).

The PMCC correlation matrix was created for full five response range and dichotomised version using listwise and pairwise exclusion of missing data. Table 12.4 shows the characteristics of these four matrices.

	Dichotomous		Full response	
	Pairwise	Listwise	Pairwise	Listwise
Mean	0.21	0.22	0.29	0.30
St dev	0.11	0.12	0.13	0.14
Minimum	-0.09	-0.14	-0.09	-0.10
Maximum	0.63	0.63	0.80	0.81
% sig. (p<0.05)	62	53	78	75
% sig. (p<0.01)	14	15	10	9

Table 12.4. Descriptive statistics for correlation matrices calculated from Youngs data set
Proportions of statistically significant results are corrected for different degrees of freedom

As can be seen from the table, the matrices were broadly very similar, with the full response range being overall higher and having a larger spread. This results in a greater proportion of statistically significant results. In all matrices the overwhelming manifold was of positive correlation, with mostly non-negative values. This is important for understanding that the first principal component to be extracted using a principal components extraction will be of 'general' delinquency (Kline, 1994).

The values contained in the four correlation matrices were correlated together using Pearson's PMCC to assess their closeness and are shown in Table 12.5.

	Dichot Pairwise	Dichot Listwise	Full range Pairwise	Full range Listwise
Dichot Pairwise	-			
Dichot Listwise	94	-		
Full range Pairwise	80	77	-	
Full range Listwise	77	78	97	-

Table 12.5. Correlations between correlation matrices calculated from Youngs data set
Decimal places are omitted

The largest influence on the difference between emerges between dichotomous and full range response, with the treatment of missing data less so but still having an effect. However, since missing data excluded with the listwise procedure would have reduced the ratio of respondents to variables from 4.6:1 down to 3.4:1, which would have been unfavourable so pairwise exclusion was used.

Given the criticisms made in previous chapters about inappropriate coefficient use, the dichotomised response range was used for these data. Furthermore, the pairwise exclusion was followed, since the listwise procedure would reduce the respondents down to 151, which would be well below 200 minimum given by Kline (1994) and very close to the approximate minimum of 150 calculated from the formula of Baggaley (cited by Child, 1990, p. 115).

- **'Factorability' of the correlation matrix**

The dichotomous pairwise exclusion matrix was tested for factorability. The data were found to be 'factorable' using the conventional measures. Thus firstly, the anti-image matrix was found to be close to zero. Secondly, the Kaiser-Meyer-Olkin (KMO) measure of 0.88460 was close to the ideal value of unity and well over the cut-off of 0.5 (Kinnear and Gray, 1997). Finally, the Bartlett Test of Sphericity was 3545.9 ($p < 0.000$), indicating the factor analytical model was appropriate.

- **Number of factors**

To discover the 'correct' number factors in the correlation matrix, the two main methods as commonly found in the factor analytical literature - especially on juvenile delinquency - were used, namely the Kaiser criterion and Cattell's scree test. The Kaiser criterion, as mentioned previously in chapter 10, states that factors with an eigenvalue greater than unity should be rotated to simple structure. Table 12.6 indicates the details of the first 13 factors extracted from the dichotomous PMCC correlation matrix with missing data excluded pairwise.

In this principal-axis factor extraction, 12 factors had eigenvalues greater than unity. It must be noted that while factor 12 with an eigenvalue of 1.00664 explained 2.2% of the variance, factor 13 with an eigenvalue of 0.95898 still explained 2.1% of the variance. Even though the criterion would accept 12 but reject 13, in terms of

Factor number	Eigen-value	Variance accounted	Cumulative Variance
1	10.84	24.1	24.1
2	2.92	6.5	30.6
3	2.60	5.8	36.4
4	2.03	4.5	40.9
5	1.54	3.4	44.3
6	1.35	3.0	47.3
7	1.28	2.8	50.1
8	1.26	2.8	52.9
9	1.16	2.6	55.5
10	1.11	2.5	58.0
11	1.06	2.3	60.3
12	1.01	2.2	62.6
(13	0.96	2.1	64.7)

Table 12.6. Details of first 13 factors from Youngs data
Note: Dichotomised correlation matrix used. Brackets indicate eigenvalue less than unity.

variance explained in the model the difference is slight between these two factors. This 12 factor suggestion was not far from what Velicer and Jackson (1990) suggested would be found with the Kaiser criterion regardless of the true structure of the correlation matrix, as suggested in chapter 10. Velicer and Jackson stated that it typically gave n 3 factors, namely 45 / 3 = 15 for the Youngs data set.

Child (1993, p. 38) cited Cattell as suggesting that the Kaiser criterion was most reliable when the numbers of variables was between 20 and 50. Since there were 45 items, it would be expected therefore the criterion 'reliably' suggested that no less than 12 factors explain the structure of juvenile delinquency.

The Cattell (1966) scree method suggests the number of factors from the factor scree plot of eigenvalues against principal components. The scree plot for the dichotomised matrix is reproduced in Figure 12.2, showing the first 20 factors. In this plot, the solid line of eigenvalues against factors starts can be seen to level out after factor 6. Extrapolating the solid line with the dotted line shows that the 'scree' begins between factors 5 and 6, meaning that the number of factors to be extracted is either 5 or 6. Extraction of 5 factors may lose some valuable variance but the extraction of 6 may include some scree. However, there is a distinct 'kink' in the plot, making the elbow somewhat double-jointed elbow after the second factor.

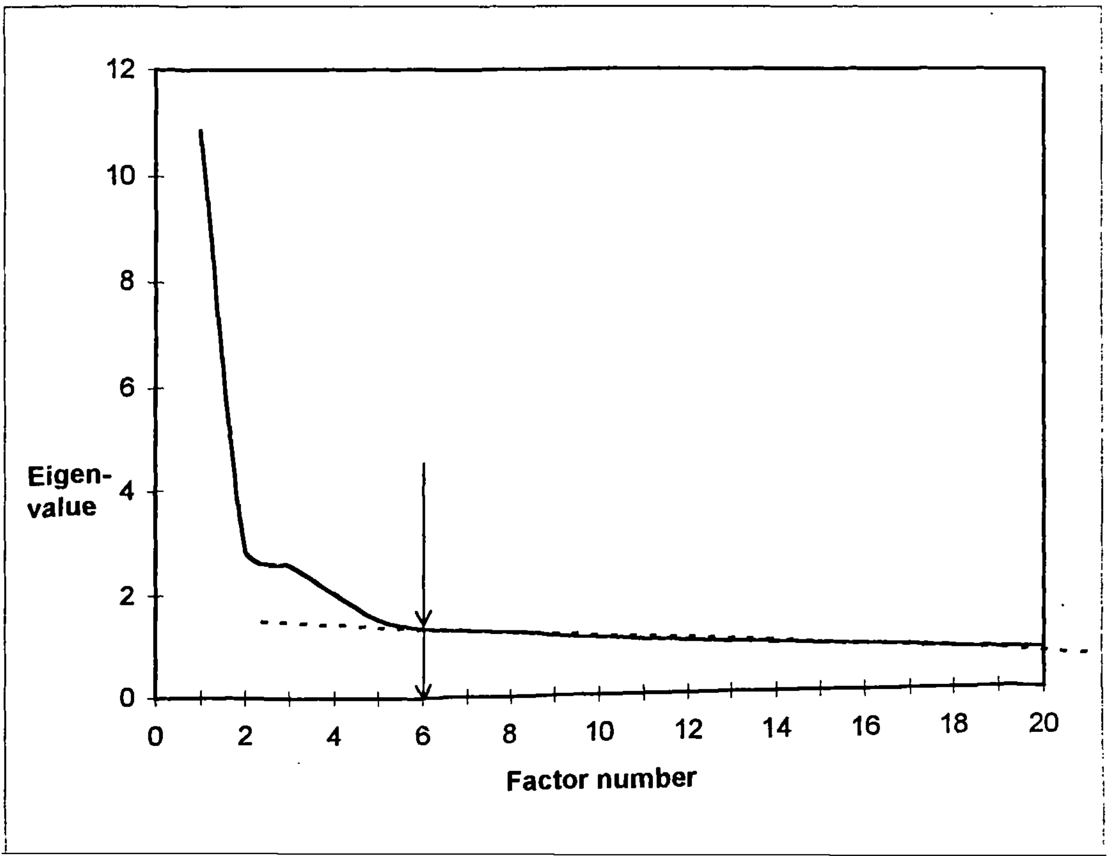


Figure 12.2. Scree plot for factor analysis on Youngs data

The final clue as to the number of factors is from Maximum Likelihood Estimations (MLE) of the factor analytical model. The advantage of the MLE model is that it provides a statistical test to examine the hypothesis that '*m* factors are required to analyse the data, given that the sample is representative of the population' and assuming multivariate normality. For the full response range, this assumption was stated earlier to be suspect, and indeed the MLE extraction for the correlation matrix using the full response range repeatedly gave improper models where communalities were greater than 1. However, for the dichotomised response range, the χ^2 in the MLE models improved up to 6 factors but then failed at 7 and above, again producing communalities were greater than 1. In other words, MLE suggested that 6 was the correct number of factors, assuming that the sample was representative of the population.

In summary, then the different numbers of factors calculated by the various methods tended to converge around 6 factors. As a final check, the original correlation matrix was recovered from the principal-axis solutions for 4 through to 8 factors. This revealed that 6 factors was the best compromise in terms of minimising the size of the recovery residuals but also keeping the solution parsimonious.

- **Item loadings of orthogonal rotation of principal factors**

Having decided that that six was the correct number of factors for the dichotomised correlation matrix with pairwise deletion, these were then rotated to Thurstonian simple structure using the orthogonal varimax method.

The loadings of the 45 items on the six factors are given in Table 12.7, also showing the item communalities and variance explained by the factors. The items are listed by factors 1 to 6 and in descending order of loading. The significance of each loading was assessed by two methods: the ± 0.3 rule of thumb and the Burt-Banks calculation (Child, 1990). The former simply states that loadings (i.e. correlations of items with factors) greater than ± 0.3 are significant. The Burt-Banks value is calculated from the number of items, respondents and factors in the analysis. For the six factors, these were 0.182, 0.184, 0.186, 0.188, 0.191 and 0.193 respectively using the 1% level of statistical significance, as recommended by Child (1990, p. 39). However, it should be noted that the 5% level of significance was only approximately 0.05 less than this, being approximately 0.170 or greater.

As can be seen from the table of factor loadings, the arbitrary ± 0.3 criterion gave a clearer and more interpretable set of values than the rational Burt-Banks criterion. However, the arbitrary criterion also indicated nearly a third of items loaded greater than ± 0.3 on more than one factor. This collinearity of items on factors was even worse with the rational Burt-Banks criterion was used, since only 6 items did *not* load significantly on more than one factor. In other words, 39 items were collinear.

This might mean that too many factors were extracted, causing this undesirable collinearity of items. However, some items were poorly accounted for, as measured by the communalities or total variance explained of the item. In fact, the poor commonalties could be taken as suggesting that more factors should be extracted.

Item description	Factor 1	2	3	4	5	6	Communality
Used weapon	**82***	16	02	10	02	-01	71
Pulled weapon	**72***	02	**31***	08	-03	09	63
Threatened violence	**65***	21*	-08	24*	18	14	59
Mugging	**61***	17	-06	07	17	-02	44
Carried weapon	**57***	04	28*	11	24*	06	48
Beaten up	**47***	16	**42***	08	10	14	46
Carried gun	**46***	16	06	-04	-02	9	25
Used fake money	**41***	**37***	09	17	06	00	35
Fought stranger	**35***	07	05	**34***	27*	17	35
Sex in public place	**29***	24*	18	07	29*	09	27
Public disturbance	18	13	08	06	12	10	09
Used ecstasy	16	**63***	03	-11	27*	18	54
Shoplift goods >£100	22*	**62***	15	14	-02	21*	52
Used heroin/cocaine	18*	**61***	00	-13	18	12	47
Forged cheque	21*	**55***	23*	09	-16	10	45
Break into house to steal	07	**49***	23*	25*	-03	**31***	46
Used barbs./amphet.	15	**49***	23*	25*	**34***	13	50
Stolen bike	02	**48***	23*	28*	-03	**35***	48
Shoplift £10-£100	13	**43***	**31***	20*	19*	15	40
Sniffed glue	13	**43***	02	**38***	26*	-01	42
Resisted arrest	29*	**36***	08	08	18*	**30***	35
Bought stolen goods	14	22*	**57***	04	13	19	44
Fireworks in public	03	04	**51***	22*	18*	22*	39
Gang fights	**39***	07	**50***	-02	23*	-08	47
Refuse coop. police	19*	14	**47***	-05	20*	26*	39
Played truant	-02	08	**44***	14	15	11	25
Used cannabis	16	**33***	**41***	10	22*	07	36
Broken windows	-13	06	**30***	**55***	**36***	21*	59
Stolen purse/wallet	11	29*	01	**54***	12	-08	40
Dialled 999 as joke	01	02	22*	**53***	02	12	35
Stolen cash fr. home	11	05	02	**45***	18	-05	26
Enter+damage building	08	19*	00	**40***	05	25*	27
Shoplift goods <£5	-07	17	26*	**37***	**33***	15	37
Set fires	22*	01	01	**36***	00	11	19
Cheat at school	04	-06	04	28*	18	-06	13
Insulted stranger	22*	-01	13	27*	**56***	13	46
Acted rowdy	20*	18	10	20*	**42***	09	31
Dropped litter	13	01	20*	26*	**41***	07	30
Been drunk under 16	08	07	25*	08	**39***	15	26
Travelled without ticket	14	16	**32***	23*	**32***	00	30
Not returned XS change	00	11	18	05	**31***	-03	14
Gone joyriding	03	26*	17	00	12	**77***	72
Broken into car to steal	16	**34***	**31***	04	-05	**62***	62
Driven while drunk/drug	21*	24*	19*	11	28*	**47***	46
Stolen car part	10	13	21*	24*	20*	**38***	31
Percentage variance	22.8	5.4	4.6	3.2	2.1	1.8	

Table 12.7. Details of 6 factor solution from Youngs data

Decimal places are omitted; **Bold figures** indicate significance with value greater than ± 0.3

* indicates significance with value greater than Burt-Banks minimum at 1% (Child, 1990, p. 110)

Clearly each factor is accounting for each item, though some to a greater degree. The Thurstonian simple structure could not obtain clear and distinct loadings. As Figure 12.3 shows, the item loadings on the first two axes are anything but clear and distinct.

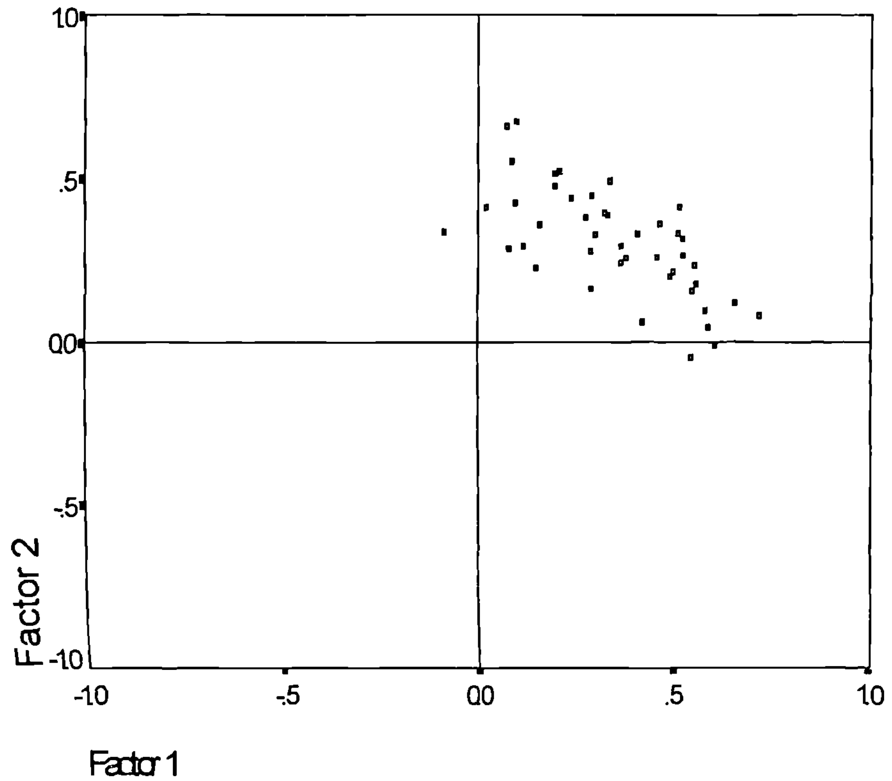


Figure 12.3. Item loading plot for factors 1 and 2 of Youngs data

However, this problem of collinearity across all the factors literally disappears when the conventional factor analytical representation is shown, with loadings less than 0.3 ignored. Such a representation is given in Table 12.8. This representation is taken to be the 'factor structure' matrix (e.g. Kline, 1994; Child, 1990). It is from such a matrix that the meaning of the analysis is taken. Naturally, the elimination of values below the 0.3 cut-off does imply that the items mostly coalesce around one of six factors. Each of these factors is independent of each other due to the Varimax rotation. Table 12.8 would lead to the conclusion that the structure of juvenile delinquent and criminal actions was dictated by a sixfold typology of distinct offence styles.

Item description	Factor 1	2	3	4	5	6
Used weapon	82					
Pulled weapon	72		31			
Threatened violence	65					
Mugging	61					
Carried weapon	57					
Beaten up	47		42			
Carried gun	46					
Used fake money	41	37				
Fought stranger	35			34		
Sex in public place	(29)					
Public disturbance	(18)					
Used ecstasy		63				
Shoplift goods >£100		62				
Used heroin/cocaine		61				
Forged cheque		55				
Break into house to steal		49				31
Used barbs./amphet.		49			34	
Stolen bike		48				35
Shoplift £10-£100		43	31			
Sniffed glue		43		38		
Resisted arrest		36				30
Bought stolen goods			57			
Fireworks in public			51			
Gang fights	39		50			
Refuse coop. police			47			
Played truant			44			
Used cannabis		33	41			
Broken windows			30	55	36	
Stolen purse/wallet				54		
Dialled 999 as joke				53		
Stolen cash fr. home				45		
Enter+damage building				40		
Shoplift goods <£5				37	33	
Set fires				36		
Cheat at school				(28)		
Insulted stranger					56	
Acted rowdy					42	
Dropped litter					41	
Been drunk under 16					39	
Travelled without ticket			32		32	
Not returned XS change					31	
Gone joyriding						77
Broken into car to steal		34	31			62
Driven while drunk/drug						47
Stolen car part						38

Table 12.8. Factor structure matrix from six factor solution on Youngs data
Decimal places are omitted
Values are greater than ± 0.3, or else shown in brackets

- **Interpretation and reliability of factors**

Farrington (1973) observed that that few researchers had attempted to measure internal consistency. Therefore, the reliability of the factors was measured using Cronbach's α, a measure of the average correlation between items and scale. The drawback with this procedure as implemented on SPSS version 6 is that it excludes cases with missing data listwise i.e. it deletes the whole case. This meant that on average 17 cases were excluded from each reliability analysis, which may or may not have influenced the α score due to inconsistent responding.

Using the factor analytical representation in Table 12.8, the names and Cronbach's α reliabilities of the six orthogonal rotated principal factors in descending order of loading are as follows:

1. Street violence and robbery. $\alpha = 0.848$. 11 items.

Used weapon, pulled weapon, threatened violence, mugging, carried weapon, beaten up, carried gun, used fake money, fought stranger, sex in public place, public disturbance.

These items show that the juvenile has threatened and used violence in order to get his way or to obtain cash or goods. He is accustomed to carrying knives and/or guns. However, the items 'used fake money [in a machine]' and 'sex in public place' do not seem to fit.

2. Drugs and thieving. $\alpha = 0.857$. 10 items.

Used ecstasy, shoplift goods >£100, used heroin/cocaine, forged cheque, break into house to steal, used barbs./amphet., stolen bike, shoplift £10-£100, sniffed glue, resisted arrest.

The use of drugs - with theft and forgery to support the habit - is indicated by this factor. Also loading on this factor were 'used fake money', 'broken into car to steal' and 'used cannabis', the first of which fits better than in factor 1.

3. Group delinquency. $\alpha = 0.741$. 6 items.

Bought stolen goods, fireworks in public, gang fights, refused to cooperate with police, played truant, used cannabis.

This factor was unclear as to its meaning, despite having a reasonably high Cronbach's α reliability value. If the highest loading item, 'bought stolen goods', was removed, then it becomes clearer as a group minor delinquency factor, probably characterised by younger juveniles. Also loading highly were 'beaten up' and 'pulled weapon', which in this context would seem to be related to the gang fighting.

4. Anti-social acts and petty thieving. $\alpha = 0.705$. 8 items.

Broken windows, stolen purse/wallet, dialled 999 as joke, stolen cash fr. home, enter and damage building, shoplift goods less than £5, set fires, cheat at school.

These items were essentially minor anti-social acts and petty thievery. There was a degree of overlap both in meaning and loadings between this and the next factor. The thieving here is of a lower order than that in factor 2, presumably since this factors contains no element of drug use.

5. Loutish behaviour. $\alpha = 0.673$. 6 items.

Insulted stranger, acted rowdy, dropped litter, been drunk under 16, travelled without ticket, not returned excess change.

The least reliable of all the factors, this indicates public order offending and general 'loutishness', with 'used barbs./amphet' and 'sniffed glue' also loading on the factor.

6. Vehicle thieves ('TWOCers'). $\alpha = 0.764$. 4 items.

Gone joyriding, broken into car to steal, driven while drunk or on drugs, stolen car part.

Despite only having four items - which Kline (1994) suggested was too low to be reliable - the factor is quite a strong 'joyriding' factor, indicating the taking of and from motor vehicles. Other high loadings included thefts of bikes and from houses (i.e. burglary), though interestingly not drugs offences which indicates that alcohol rather than substances were meant by 'driven while drunk or on drugs'.

The average Cronbach's α reliability for the six factors was 0.765.

- **Oblique rotation of principal factors**

In order to see if the factors suggested by the representation truly were independent, a further factor analysis was undertaken whereby the factors were allowed to correlate with each other. Thus a oblique rotation was compared to the orthogonal rotation. The oblique factor analysis was a Direct Oblimin with delta equal to zero, using a six principal factors extraction.

Overall, the factor loadings were not markedly changed by the oblique rotation, which allowed the factors to correlate with each other for the sake of achieving simple structure. There were only two differences in factor interpretability of the oblique solution: firstly, 'Stolen bike' and 'Resisted arrest' were included in the Vehicle thieves factor rather than the Drugs and thieving factor; and secondly, 'Travelled without ticket' item was included in Group delinquency rather than the Loutish behaviour factor. The overall average Cronbach's α reliability of the oblique solution was slightly less at 0.76255 compared to the average reliability of 0.76475 for the varimax solution. The loosening of the restriction of orthogonality of factors did not improve the factor analytical solution, which would be taken to suggest that independence of dimensions was indicated by the analysis.

However, the instability of some variables across the analyses also alludes to the fact that many of the items within the same factor analytical solution cannot be thought of as exclusively 'belonging' to one factor or another, but are in fact best described by an amalgamation of factors. Such items are readily seen to load highly on more than one factor, such as the 'Travelled without ticket' variable which loaded 0.319 on factor 3 (Group delinquency) and 0.323 on factor 5 (Loutish behaviour) in the orthogonal solution, a difference of 0.004 which when squared is a tiny amount of variance.

The rotation to simple structure is intended to maximise loading on one factor and minimise loadings on all other for each item (Kline, 1994), thereby increasing the 'belongingness' of items to single factors. For some items this patently works, such as 'Used weapon' with loadings of **0.82**, 0.16, 0.02, 0.10, 0.02 and -0.01 for factors one to six, using **bold** to indicate greater than ± 0.3. Therefore of the communality of 0.71 (i.e. item variance explained by all six factors), 90% was due to the one factor into which it was grouped. But with the 'Travelled without ticket' item, the figures are

0.14, 0.16, **0.32***, 0.23*, **0.32*** and 0.00, with only 34% due to one factor into which it was grouped, again using **bold** to indicate greater than ± 0.3 and * to indicate greater than the Burt-Banks value (Child, 1990).

- **Second order factor analysis of oblique factors**

Although the orthogonal solution was interpreted as the main factor analysis, the correlation matrix between factors for the oblique rotation was also factor analysed, given a higher second-order (SO) solution. A principal component extraction was done on the correlation matrix, since the previous principal factor extraction of the first order (FO) factor analysis was intended to remove unique variance leaving only common variance in the factor structure, and also therefore the factor correlation matrix. The principal component extraction gave only one component with eigenvalue greater than unity, with a value of 2.22 corresponding to 37.1% of the variance. The second principal component had an eigenvalue close to unity at 0.99 with 16.6% of the variance, with the scree test of eigenvalues against factors indicating this component was not scree. A Maximum Likelihood Extraction also indicated a better fit for two factors rather than one.

Two principal factors were therefore extracted for the SO factor analysis, which were rotated to simple using the orthogonal Varimax method. As can be seen in Table 12.9, the SO factor analytical solution was extremely unclear.

As can be seen from this table, higher factor 1 is loaded most highly by lower factors (inverse) 1, 2, 3, 4 and 6, while higher factor 2 is loaded most highly by lower factor 5. This however must rely on the statistical significance of these loadings being the ± 0.3 criterion. The only real understanding of the higher structure is that the first higher factor has non-violent thieving, that is to say gaining financial advantage through theft against property rather than theft against the person directly, while the second higher factor has loutish and some anti-social behaviour.

Clearly, these results do not add anything to the understanding of the structure and serve to remind that factors are not entities, but merely speculative linear combinations of items acting as summaries of the content implied.

First order factor		Second order factor		Commu-nality
Number	Name	1	2	
One	Street violence and robbery	**-41**	-15	19
Two	Group delinquency	**57**	09	33
Three	Drugs and thieving	**46**	26	28
Four	Anti-social acts and petty thieving	**31**	28	17
Five	Loutish behaviour	12	**94**	90
Six	Vehicle thieves	**65**	.06	43
Percentage variance		37.1	16.6	

Table 12.9. Second order (SO) factor analytical solution from Youngs data
Decimal places are omitted
Bold figures indicate significance with value greater than ± 0.3
Names of first order (FO) factors are taken from orthogonal solution, since oblique solution was similar

- **Conclusions from factor analyses**

The largely positive manifold of the correlation matrix suggested that juvenile delinquency and criminal actions were all related. The first order principal-axis factor analysis suggested that six factors explained the structure provided in the correlation matrix. Many items were collinear, loading on more than one factor, and many item communalities were low, suggesting the factors did not exhaust the correlation matrix.

However, the factor analytical representation in Table 12.8 suggested that a strong six factor structure could be obtained. These factors were named and tested for Cronbach's α reliability, which was good overall.

The independence the factors was tested by comparing with an oblique rotation with principal-axis factoring. There was no overwhelmingly structural change, though with some anomalies. A second order factor analysis of the mildly related oblique factor correlation matrix did not produce any strong higher order structure, reinforcing the idea of the six factor extraction being adequate.

The nature of the response range was found not to be as influential as had been proposed by Loeber and Waller (1988) in determining the number of factors *viz.* types of delinquency. The full range and dichotomised correlation matrices were similar and gave a similar number of factors to be extracted under Maximum Likelihood and scree tests, though not using the Kaiser criterion of eigenvalues greater than unity. It is possible that this was the actual source of difference for the Loeber and Waller (1988) study, though this cannot be confirmed since the eigenvalues were not published by Loeber and Waller.

Also, even though Braithwaite and Law (1978) had criticised the use of the PMCC with the ordinal ratings data, it was shown that the treatment of missing data was more important than the dichotomisation to reduce 'scales of measurement' violations. This is important for the next chapter, where a different coefficient is used to deal with the problem of missing.

Within the six factors it was not possible to detect any dimension of seriousness within the factors themselves, as had been suggested by various studies summarised in chapter 11.

A plot of the item loading on the first two principal axes showed that the items were extremely poorly distributed in the plot, not least because the two components only explained 30.6% of the variance and hence gave an inaccurate representation of the items. The weaknesses of factor loading plots was also noted in the previous chapter in the context of Wiggins' interpersonal circumplex.

Finally, no indication was given as to a general 'pool' of delinquency from which various specialisms may arise or be superimposed, as Farrington, Snyder and Finnegan (1988) proposed in their hypothesis of integrated specialism and versatility.

Summary of Chapter 12

The Youngs data set on juvenile delinquent and criminal actions was introduced as a large scale survey overcoming many of the problems associated with the surveys in the previous chapter. It was shown that the data were suitable to be factor analysed, which was done in Empirical Study 12.1. However, difficulty was in noted with the response range and its unsuitability for Pearson's PMCC, as well as the large number of missing responses. Dichotomisation and pairwise exclusion was suggested to resolve these problems. Various numbers of factors were suggested by different methods, but six factors were extracted using principal-axis factoring and rotated to simple structure with orthogonal Varimax rotation. Contrary to the goal of simple structure, the factors were shown *not* to have a few high loadings and the rest near zero. The conventional interpretation of the factor representation ignored this, implied six different distinct types of juvenile delinquent and criminal action. Oblique rotation and second-order factor analysis confirmed this view of independent factors. The conclusions from factor analysis were reiterated.

Chapter 13
structure and representation in a faceted analysis of juvenile delinquent and criminal action

The Youngs Survey of Juveniles' Delinquent and Criminal Actions

The previous chapter described a factor analysis of the Youngs data set on juvenile delinquent and criminal actions, a large scale survey of 207 young males. A factor analysis was performed and interpreted as would be expected by convention. Furthermore, it was suggested that a conventional interpretation of the factor representation in the factor loading chart of Table 12.8 would indicate independent types of criminality - the dimensions of the factor space. In other words, representation would be mistaken for true structure. Structural hypotheses of versatility and generalism

Using the same data set, this chapter considers a different representation and interpretation: a faceted SSA-I. The results of this analysis are then compared to the factor analytical solution, and the advantages of the faceted interpretation are explored.

The Nature of Data on Criminal Actions

The value and need for a non-metric approach to the juvenile delinquency data was shown by Braithwaite and Law (1978) and Nutch and Bloombaum (1968). However, Smith, Smith and Noma (1986) rejected both the non-metric scaling and the factor analytical approaches to understanding juvenile delinquency due to poor conceptualisation of the data. Since it is argued in this thesis that the nature of criminal actions data has a strong impact on the way structural hypotheses may be made and modelled in geometric MDS representations, this critique by Smith, *et al.* is especially relevant.

Smith, *et al.* (1986) argued analyses must take into consideration the fact that criminal arrest records used by delinquency researchers were an instance of 'pick any/*n*' data (i.e. 'pick any of *n* objects'; e.g. Coombs, 1964, p. 295). Such data reflect the fact that individuals choose objects or stimuli (i.e. commit crimes) from a fluid, individual-defined universe varying between respondents; each individual has a subcontent universe. For arrest histories, this picking process occurred twice: firstly when a particular crime was chosen to be committed by a juvenile, and secondly, when the authorities chose to arrest and charge the offender with that crime. Even for self-report studies such as the Youngs data, however, this is relevant since the respondents have picked the acts from their own universes of potential acts.

The importance of the 'pick any/*n*' data concerns what is meant by *not* picking an object. Does this mean that the object was considered and then rejected? Or does it mean that it was simply not considered? Smith *et al.* argued that 'This calls into question the strong rejection assumption behind most scaling [and factor analytical] techniques, where a nonchosen alternative is taken as evidence that the alternative has been rejected.' (Smith *et al.*, 1986, p. 332)

Therefore Smith *et al.* (1986) proposed that Variance Centroid Scaling (VCS) should be used to take into account the nature of the data. VCS could derive the dimensions of arrests that reflected the whole 'career' of the juvenile, rather than what was just picked. VCS was described as being a variant of Correspondence Analysis (e.g. Weller and Romney, 1990), which in turn is the initial approximation of MSA-I (Guttman, 1985). VCS would therefore overcome the bias stated by Smith *et al.* to be inherent in arrest records by treating offences *not* in the 'criminal career' as missing data rather than not present. This could suggest therefore that a faceted analysis using SSA-I would be unsuitable for the Youngs data. However, the strong argument put forward by Smith *et al.* (1986) does contain some flaws and can be overcome in many respects by the faceted analysis presented in this chapter.

Rejoinder to Smith, Smith and Noma (1986)

Arrest records are influenced in adverse ways by the two 'pick any/*n*' factors of juveniles and authorities. Several studies (e.g. Arnold, 1965; Nutch and Bloombaum, 1968) have indicated that juveniles tend to commit minor crimes more than serious crimes, i.e. 'pick' less serious ones. But consider also the 'picking' of crimes by the authorities: for example, a higher proportion of the sample had been arrested for an atrocious assault than had been arrested for loitering (10.2% and 6.5% respectively, Smith *et al.*, 1986, p. 350). Yet clearly the more serious assault would be more likely to result in an arrest than standing on a street corner, even though the juveniles would expected to loiter more than do 'atrocious assaults'; the authorities pick more serious ones, even if they are more frequent. Though these biases cannot be assumed to cancel each other out, they will definitely interact to diminish structure recoverable. Weaker non-metric techniques are therefore suggested.

The rejection by Smith *et al.* (1986) of MDS also reflects an generalisations and assumptions made from an over-simplification which was acknowledged by Smith *et al.* This was their restriction to regard only classical scaling models and the single proximity measure of Euclidean distance (Smith *et al.*, 1986, footnote p. 336; table 1, p. 338). In factor analysis only one proximity measure is permissible: the Pearson's PMCC, or its equivalent (Gorusch, 1988). Non-metric MDS in particular does not suffer from this restriction, and the Jaccard's index can overcome this problem of uncertainty over lack of reporting, as was shown in chapter 8. Therefore it is possible to still use 'pick any/*n*' data as Similarities, and even more so when using self-report data rather than arrest records, as in the Youngs data.

These self-report data are the first and far more reliable stage of the Smith *et al.* 'pick any/*n*' data, namely the reliable reporting of the choices made by the juveniles. The use of the Jaccard's coefficient as a similarity measure would diminish the rejection assumption in dichotomised data. Jaccard's only takes joint occurrence as worthy of note, ignoring joint non-occurrence (see Forumla 8.1; Table 8.1). Furthermore, a careful sampling of the content universe of the full range of illegal acts would ensure that each person's individual content universe would be represented.

Furthermore, it will be shown later that structural hypotheses about individuals using Single Stimulus data can be guided by the SSA-I representation of Similarities data. This would overcome the last advantage of the paradigm proposed by Smith *et al.* (1986) that structure on individual differences was shown by VCS and its 'pick any/*n*' Single Stimulus data.

It is proposed that the faceted regional interpretation of the non-metric SSA-I space has the ability to identify structural aspects ignored by the factor analyses, Guttman scaling and VCS methods mentioned previously. These would also be concealed by a dimensional interpretations of the non-metric space. The faceted reanalysis of the Youngs juvenile criminal actions data was done in Empirical Study 13.1.

Empirical Study 13.1: A Faceted Analysis of Juvenile Delinquency and Criminality

The same set of variables as in Empirical Study 12.1 were taken from the Youngs data set on juvenile criminal actions, namely 45 items from 207 respondents. The items in the delinquency data were associated using Jaccard's coefficient, using the dichotomised raw data. The rationale for this was given in the previous section, namely a recognition of the points made by Smith *et al.* (1986). Missing data were coded the same as 'not ever committed', meaning that presence and absence in the items was to be understood as 'involvement admitted' and 'involvement not admitted'.

Table 13.1 shows some descriptive statistics on the Jaccard's association matrix.

	Jaccard's Value
Mean	0.49
Maximum	0.85
Minimum	0.17
St. dev.	0.15

Table 13.1. Descriptive statistics on Jaccard's association matrix from Youngs data set of male juvenile criminal actions

As can be seen, the values are not spread on the full possible range of the coefficient from 0 to 1. Instead these values are roughly normally distributed around a value of approximately 0.55 but with a slightly disproportionate number of values from 0.25 to 0.30, hence the mean of 0.49. The non-metric transformation of the SSA-I minimises this skew.

The unusual use of the Jaccard's coefficient with questionnaire data was assessed by comparison with the correlation matrices explore in Empirical Study 12.1 in the previous chapter. These are shown in Table 13.2.

		Pearson				
		Dichot Pairwise	Dichot Listwise	Full range Pairwise	Full range Listwise	Jaccard
Pearson	Dichot Pairwise	-				
	Dichot Listwise	94	-			
	Full range Pairwise	80	77	-		
	Full range Listwise	77	78	97	-	
Jaccard		55	48	49	47	-

Table 13.2. Correlation between Jaccard's association matrix and Pearson's correlation matrices from Youngs data set of male juvenile criminal actions

As can be seen from this matrix, the overall impact of ignoring the conjoint absence as with the Jaccard's coefficient is not so radically different from the Pearson's coefficients, which use conjoint absence. The Jaccard's matrix had most in common with a pairwise exclusion of the dichotomised data, which would be expected since the pairwise exclusion of missing data is most similar to the Jaccard's treatment of missing data. Also, Jaccard's is a dichotomous coefficient, unlike the full range Pearson's matrices. However, the similarities with the other types of correlation matrix were not far from this value at all. The values in the above matrix were all highly statistically significant.

Figure 13.1 shows the two dimensional SSA-I solution of the matrix in Table 13.2. The alienation in this solution was 0.00135 with global monotonicity and shows Jaccard's being closer to dichotomous than non-dichotomous data, and the independence of the treatment of missing data. However, the treatment of missing data is less important relatively than the choice of coefficient, which backs up chapter 8. This solution differs from the one in Figure 8.2 in Empirical Study 8.1 since only

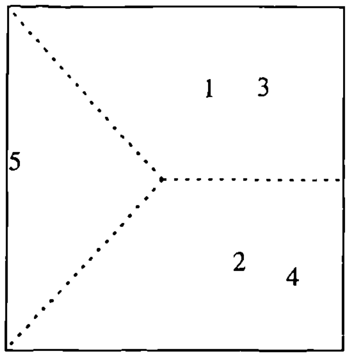


Figure 13.1. SSA-I of correlations between Jaccard's association matrix and Pearson's correlation matrices
Key: 1. Dichotomous Pairwise; 2. Dichotomous Listwise; 3. Full range Pairwise; 4. Full range Listwise; 5. Jaccard's

the associations and correlations are being used here, as opposed to the distance matrices as well.

- **Fit and dimensionality**

Although the issue of 'extracting' principal components is relevant to SSA-I, it is limited to the initial approximation. More importantly, the issue of dimensionality - the closest to the factor analytical 'number of factors' problem - is related to the representation with MDS, rather than the structure with factor analysis. Provided that the representation is adequate to model the hypothesised structure (see chapter 6), then dimensionality is not overwhelmingly important.

A preliminary SSA-I was run on the Jaccard's association matrix from 1 to 5 dimensions using local monotonicity. It was immediately apparent that the one-dimensional and two-dimensional solutions were degenerate and had placed most items into a 'clump' with a couple of items outside the 'clump'. The local monotonicity weighting had pulled the points together, and dimensional slurring (Lee and Canter, submitted) had carried forward the error of the one-dimensional configuration, creating clearly sub-optimal solutions. As the plot of dimensionality against alienation in Figure 13.2 shows, there would seem to be an elbow at dimensionality of three indicating that the data were too noisy to be scaled in anything less.

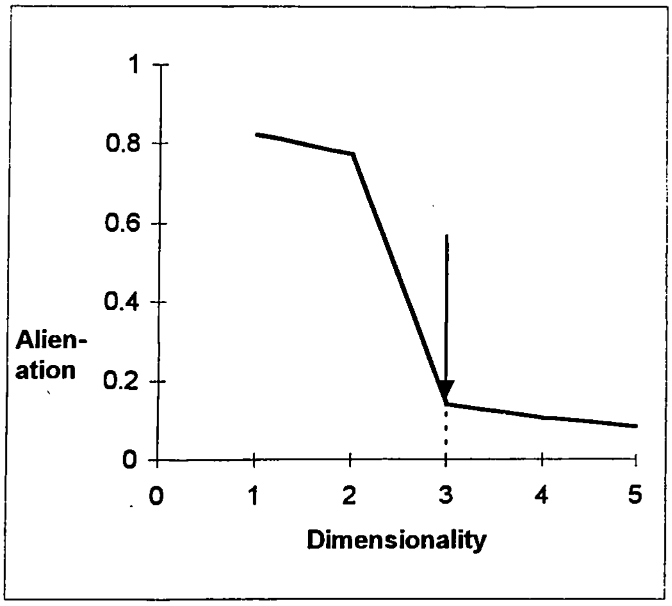


Figure 13.2. 'Elbow plot' of alienation against dimensionality for SSA-I using local monotonicity and starting dimensionality of 1

However, when this was repeated except starting at two dimensions rather than at one dimension, this 'elbow' disappears. Figure 13.3 shows the new plot for the solution starting at one dimension, with the values from Figure 13.2 superimposed as a dashed line for comparison.

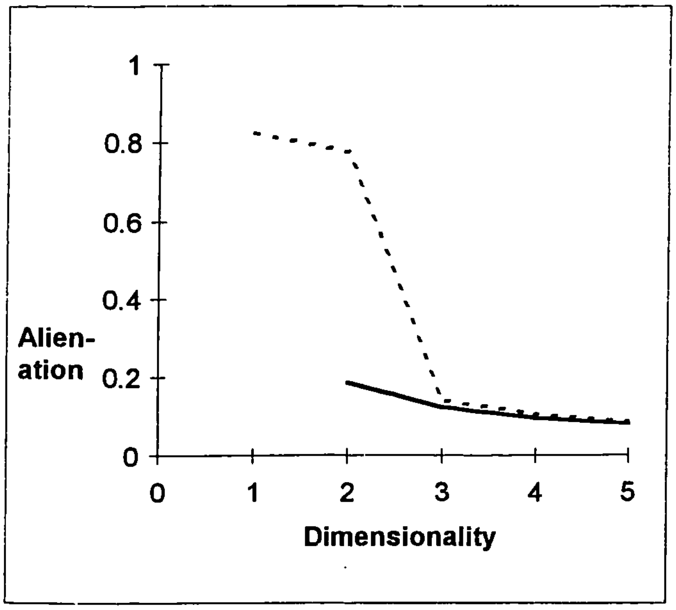


Figure 13.3. 'Elbow plot' of alienation against dimensionality for SSA-I using local monotonicity and starting dimensionality of 2, with plot of starting dimensionality of 1 shown as dashed line

A comparison with a globally monontonic solution was made and it was noted that the alienation decreased to 0.17 in two dimensions. However, the slightly higher alienation was justified to make use of local monotonicity because of the large number of high similarities that were close together in the centre of the SSA-I solution.

A two-dimensional local monotonicity solution was therefore used, and the plot obtained is reproduced in Figure 13.4.

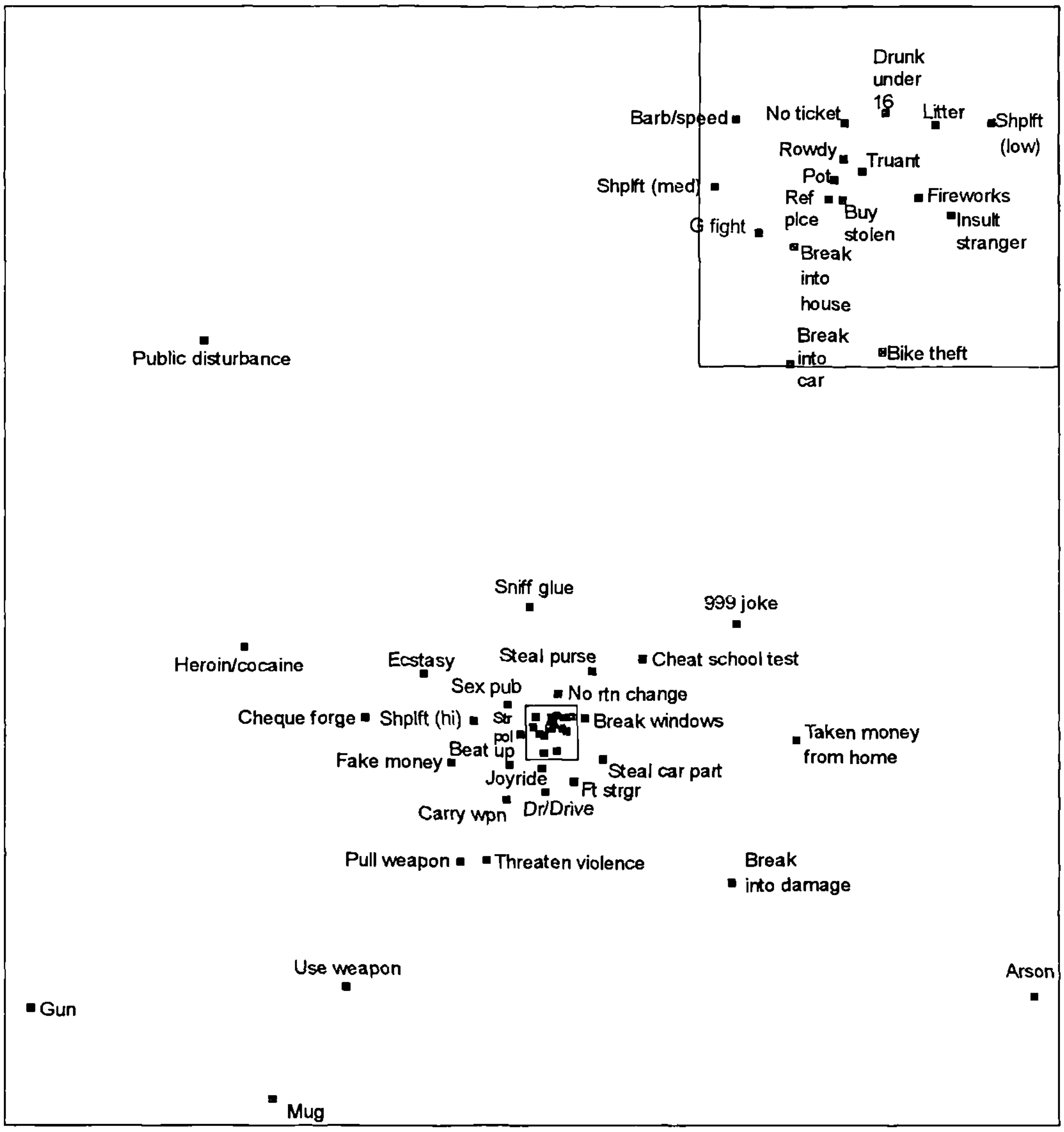


Figure 13.4. Two dimensional SSA-I of Youngs data set using local monotonicity

This solution had an alienation of 0.19, which was surprisingly good given that there were 45 items and far less than Braithwaite and Law (1978) who obtained a two-dimensional solution with alienation 0.23 using only 17 items.

The literature on juvenile delinquency has reported several distinct types or styles of actions, as well as a general, versatile delinquency from which various these specialisms may arise (Farrington, Snyder and Finnegan, 1988). Evidence for versatility comes from for example Nye and Short (1957) and Loeber and Waller (1988), and different degrees of versatility from Braithwaite and Law (1978). Evidence for distinct types of theft, vandalism, and aggression is found in Loeber and Waller (1988), of theft against persons known and theft against persons unknown in Scott (1959) and for Guttman scales of theft, nuisance and violence offences by Arnold (1965). The qualitative meta-analysis of Klein (1984) pointed to conceptual similarity among acts constituting assault, theft, auto offences, drug offences and status offences. Lastly, an element of seriousness across the types of offences was indicated by Braithwaite and Law (1978) and Nutch and Bloombaum (1968). The plot in Figure 13.4 was examined for evidence of these constructs given in chapter 11.

It was noted that central to the plot and therefore the meaning of the content universe of juvenile delinquency was a tightly bonded 'cluster' of items. Clearly these items must be understood as a whole, so a regional interpretation must take these items as a region in themselves. However, outside of this central region, there is clear differentiation between different criminal behaviours with some similarities in intentions. In this outer region, distinct types emerge. Therefore it is proposed that the regional interpretation of the plot should include a classification into themes reflecting types of juvenile delinquency in a polar facet - a geometric circumplex - with the central region forming a theme in itself, as shown in Figure 13.5.

The details of the regional hypotheses in Figure 13.5 are as given below, along with the Cronbach's α calculations. As was cautioned in chapter 7, these values must be taken with a modicum of caution since the associations were made using a different coefficient to the one use in the Cronbach's α calculation, i.e. Jaccard's rather than Pearson's.

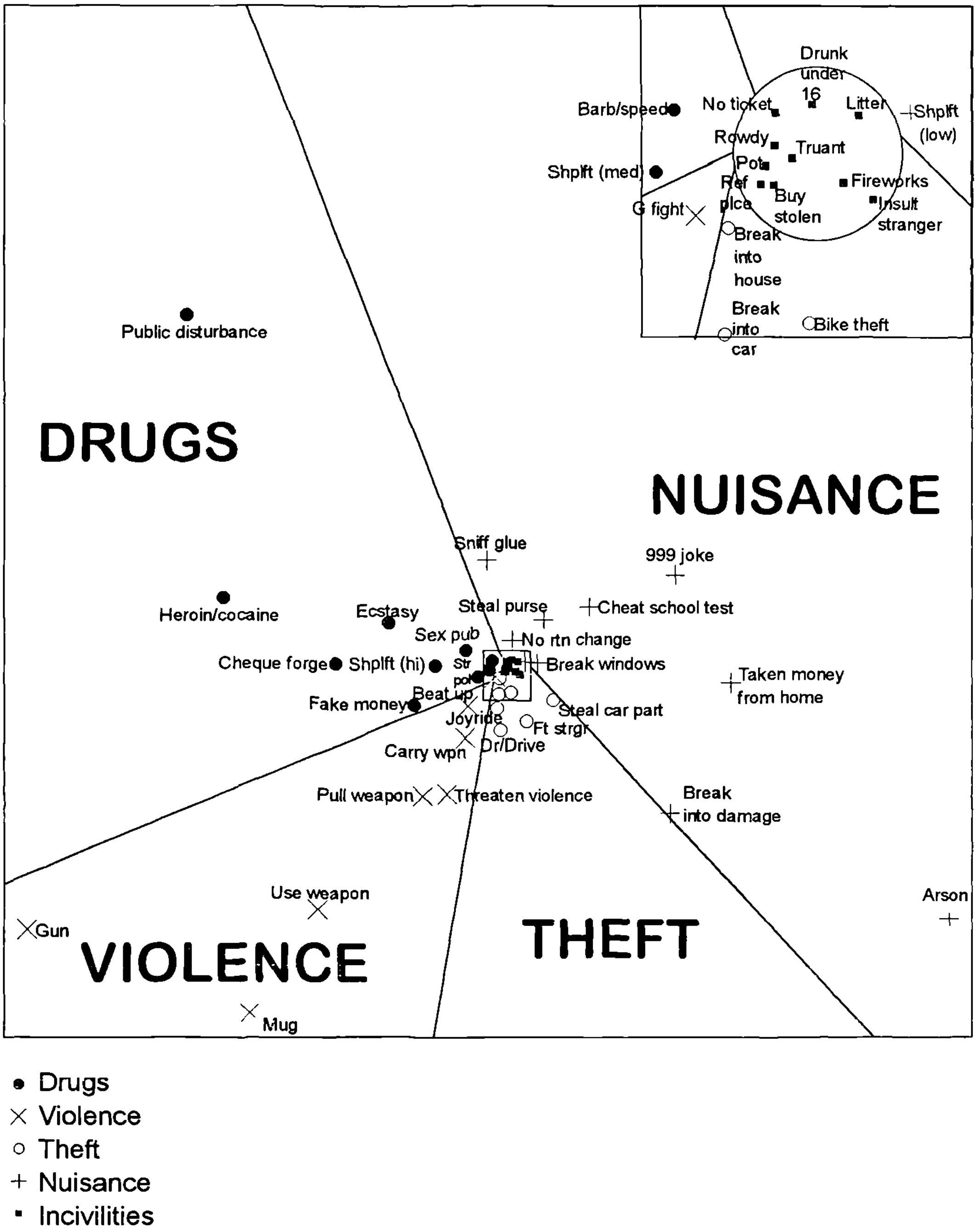


Figure 13.5. Regional interpretation of a polar facet in two dimensional SSA-I of Youngs data set

Incivilities. $\alpha = 0.773$. **10 items.**

Insulted stranger, fireworks in public, dropped litter, travelled without ticket, been drunk under 16, acted rowdy, bought stolen goods, used cannabis, refuse coop. police, played truant.

The items here are mainly minor delinquent activities or crimes, indicating the beginnings of marginalisation and rebellion against society, and in particular against authority figures. Respondents admitted on average to just over 8 of the 10 items. The position of the region in the plot is a central 'cluster' from which others radiate out, which is developed later in its description as part of a modular facet; it is a region that serves as a link to the others.

Drug Lifestyle. $\alpha = 0.820$. 10 items.

Shoplift goods >£100, used heroin/cocaine, used ecstasy, used barbs./amphet., forged cheque, shoplift £10-£100, used fake money, resisted arrest, sex in public place, public disturbance.

The items in this region contain three of the four drugs items, the other being 'used cannabis' which quite common and less serious hence in the minor Incivilities region. However, the other three drugs are class A and would probably attract more police attention. Also in the region are some dishonesty offences and higher value thefts which could be used to get the money required to support a habit of heavier drug use or addiction. One unusual item though was 'public disturbance'. The full question for this item was which was 'Have you attended a demonstration or sporting event to cause a disturbance or be violent?' This question may have picked up on a different type of violence to the Violent Transaction region, and it is suggested that this item may be detecting involvement in the riots over the Criminal Justice Act (1994) in 1993 when the questionnaire was administered. The riot was reported to involve many following a club culture who are more likely to be committed to a drugs lifestyle. Nevertheless, the item had poor LSB (chapter 7), given in Figure 13.6 which shows the 5 highest associations with this item.

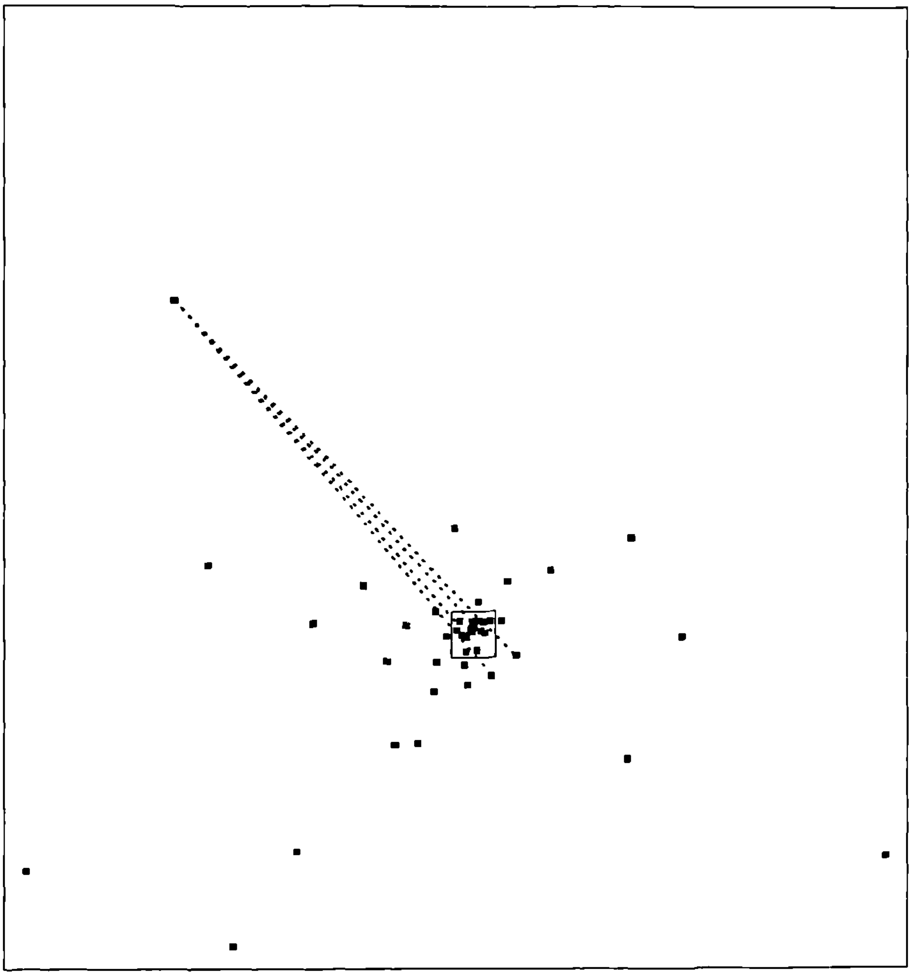


Figure 13.6. Poor LSB in Public Disturbance ite n in SSA-I of Youngs data

Theft. $\alpha = 0.789$ **7 items.**

Gone joyriding, driven while drunk or on drugs, broken into car to steal, stolen bike, stolen car part, break into house to steal, fought stranger.

Most of the targets of this Theft region are cars, with both stealing of and stealing from cars, though there are also thefts of bikes and from houses (i.e. burglaries). The item 'fought stranger' is somewhat out of place, though its proximity to the Violent Transactions region explains this. The closeness of the 'enter and damage building' could have been understood as attempted burglary by the respondent, though it is more likely to refer to aggravated vandalism which would not make it a Theft item. Similarly, the 'shoplift goods less than £5' was thought to imply a different type of theft to the vehicle and house items classified as Theft, though its empirical closeness is noted. Interestingly, the theft items in the Theft region are differentiated from the theft items in the Drugs Lifestyle which are more indicative of dishonesty.

Nuisance. $\alpha = 0.726$. 10 items.

Broken windows, stolen purse/wallet, sniffed glue, stolen cash fr. home, dialled 999 as joke, shoplift goods less than £5, enter and damage building, set fires, cheat at school, not returned excess change.

These items are all property offences except for 'dialled 999 as a joke'. They are also in some sense 'immature', and could even represent what might be termed the symptoms of conduct disorder or an unstable home rather than criminality. Hence the region is named Nuisance since these expressions of an immature self acting on the environment, tending towards vandalism rather than aggression. The item 'sniffed glue' is close to the Drugs Lifestyle region, but classified as Nuisance since the other drugs items seem more sophisticated and mature, rather than solvent which characterise younger juveniles. The bonding of this item was indeed spread across Nuisance, Drugs Lifestyle and Incivilities themes.

Overall, the items in this region are quite dispersed, and for this reason the region is slightly less bonded than the others, as shown by the Cronbach's α.

Violent Transactions. $\alpha = 0.853$. 8 items.

Pulled weapon, used weapon, carried weapon, threatened violence, beaten up, mugging, gang fights, carried gun.

The intention of the behaviours in this region is violence directed towards people - or going prepared for violence towards people - rather than property as in vandalism. The region is named Violent Transactions to reflect the way the respondent seems to be prepared to deal with the world. The item 'gang fights' suggests a gang element to these behaviours. This theme is shown by the recent press reports about which were reported recently in so-called 'Triad' gangs in London implicated in the stabbing of Philip Lawrence, the London headmaster in December 1995.

The conceptual tightness of the region is shown by a high Cronbach's α, and a robust Localised Spatial Bonding.

The average reliability was 0.792 for the faceted regional interpretation. This was better than the average for factor analysis, which had an average reliability of 0.765.

There are two important reasons for this improvement of recovering structure in the SSA-I representation as opposed to the factor analytical representation. Firstly, there is a truncation of the structure by rotation to Thurstonian 'simple structure' and the interpretation from the factor analytical representation. Empirical Study 13.2 proposes a non-metric equivalent to the simple structure, reiterating that the meaning of each item is spread across all the themes or factors.

Secondly, there is the ability in the faceted interpretation to make conceptually meaningful regional hypotheses mindful of the demands of LSB on items such as 'sniff glue'. Empirical Study 13.3 examines the way the selection of items for factors in the domain is far less satisfactory than how they are selected in a regional interpretation of SSA-I.

Empirical Study 13.2: Regional Interpretation as Non-Metric Simple Structure

One important point to note is that the use of regions in the faceted interpretation is not intended to imply a strict mutual exclusivity for the items. That is to say, the variance in the items is shared by the regions, though each behaviour adds predominantly to one intention in one region. This is revealed by the SSA-I representation having some items near the boundaries of regions and other items in the centre. If the regions of item were conceptually distinct from other regions of items then items would form tighter 'clusters' within the partitioning boundaries. This was shown in the regional interpretation of the factor loadings in Figure 10.2.

In factor analytical terms, the correlation or loading of items onto more than one factor is unwelcome, and the rotation to Thurstonian simple structure is intended to diminish this. Added to this is the representation of the factor structure as ignoring any variance other than the highest loading, as in Table 12.8. In effect, within factor-variance in items is emphasised and between-factor variance in items is ignored. This Empirical Study demonstrates that if desired it is possible to achieve the same (mis)representation in a faceted regional interpretation, and hence conceal generality of criminality and indicate specialism. A 'non-metric simple structure' is proposed as follows.

1. Regional hypotheses are proposed using the faceted regional interpretation of the SSA-I space. This is analogous to the extraction of factors from the correlation matrix.
2. Each item in the region is then correlated with the scale total for each region - its 'loading' on the region is calculated. This is analogous to finding the unrotated factor structure matrix, though naturally regions are real and substantive, though factors are purely hypothetical and mathematical.
3. The items in the non-metric 'regional structure matrix' are then optimised for within-region variance while ignoring between-region variance. The analogy here is with rotation to simple structure and factor loading less than ± 0.3 being excluded from the factor structure matrix.

The regional hypotheses in stage 1 has already been done in Figure 13.5 and shown in Empirical Study 13.1.

In stage 2, since there were five regions proposed in the SSA-I space of the Youngs juvenile delinquency data, each item therefore has five 'loadings' or correlations against each of the five themes, *viz.* the correlation between the dichotomous item and the sum of the items in the scale. Since this is a non-metric loading with no distributional assumptions, Spearman's ρ is most suitable. Table 13.3 presents the item-region correlations for each of the 45 items 'loading' on each of the 5 regions. The table shows that for each and every item, the highest correlation is indeed found between the item and its own region.

Stage 3 of the creation of 'non-metric simple structure' requires emphasising within-region variance while ignoring between-region variance. The most obvious way to do this would be to keep the highest 'loading' (i.e. item-region correlation) and to drop the rest, as might by done in the factor analytical representation of the factor structure matrix. This is done in Table 13.4, which follows Table 13.3 in its representation but drops any 'loading' other than the highest.

Name	Incivilities	Drugs	Theft	Nuisance	Violence
Insulted stranger	**61**	27	36	41	34
Fireworks in public	**55**	27	38	34	25
Dropped litter	**55**	26	31	38	29
Travelled without ticket	**53**	34	28	37	30
Been drunk under 16	**52**	25	29	29	25
Acted rowdy	**51**	37	29	38	31
Bought stolen goods	**49**	39	40	25	36
Used cannabis	**46**	42	36	29	36
Refuse coop. Police	**45**	34	36	19	35
Played truant	**41**	22	24	27	18
Shoplift goods >£100	30	**74**	53	29	32
Used heroin/cocaine	19	**67**	39	12	29
Used ecstasy	25	**66**	43	18	31
Used barbs./amphet.	44	**65**	50	44	37
Forged cheque	23	**62**	41	21	33
Shoplift £10-£100	45	**62**	49	38	35
Used fake money	31	**59**	35	29	43
Resisted arrest	30	**59**	47	24	37
Sex in public place	37	**52**	35	30	39
Public disturbance	19	**39**	23	16	25
Gone joyriding	38	45	**73**	20	20
Driven while drunk/drug	44	49	**71**	32	35
Broken into car to steal	35	49	**68**	23	34
Stolen bike	33	50	**65**	37	21
Stolen car part	38	35	**64**	39	28
Break into house	31	52	**60**	35	26
Fought stranger	36	37	**51**	38	41
Broken windows	49	25	44	**67**	13
Stolen purse/wallet	26	32	25	**64**	23
Sniffed glue	31	44	31	**58**	25
Stolen cash fr. Home	25	13	17	**56**	18
Dialled 999 as joke	37	17	31	**56**	13
Shoplift goods <£5	42	32	40	**55**	16
Enter+damage building	20	26	36	**55**	18
Set fires	23	17	24	**42**	25
Cheat at school	21	08	09	**40**	09
Not returned xs change	22	18	14	**32**	11
Pulled weapon	35	35	33	21	**78**
Used weapon	29	43	28	22	**77**
Carried weapon	44	38	36	27	**73**
Threatened violence	36	47	41	36	**71**
Beaten up	42	46	44	27	**70**
Mugging	25	38	23	20	**64**
Gang fights	41	37	24	21	**61**
Carried gun	22	33	22	10	**56**

Table 13.3. 'Non-metric simple structure': Item-region correlation using Spearman's ρ on Youngs data

Decimal places are omitted

Bold figures indicates highest Spearman's correlation value for that row

Name	Incivilities	Drugs	Theft	Nuisance	Violence
Insulted stranger	61				
Fireworks in public	55				
Dropped litter	55				
Travelled without ticket	53				
Been drunk under 16	52				
Acted rowdy	51				
Bought stolen goods	49				
Used cannabis	46				
Refuse coop. Police	45				
Played truant	41				
Shoplift goods >£100		74			
Used heroin/cocaine		67			
Used ecstasy		66			
Used barbs./amphet.		65			
Forged cheque		62			
Shoplift £10-£100		62			
Used fake money		59			
Resisted arrest		59			
Sex in public place		52			
Public disturbance		39			
Gone joyriding			73		
Driven while drunk/drug			71		
Broken into car to steal			68		
Stolen bike			65		
Stolen car part			64		
Break into house			60		
Fought stranger			51		
Broken windows				67	
Stolen purse/wallet				64	
Sniffed glue				58	
Stolen cash fr. Home				56	
Dialled 999 as joke				56	
Shoplift goods <£5				55	
Enter+damage building				55	
Set fires				42	
Cheat at school				40	
Not returned xs change				32	
Pulled weapon					78
Used weapon					77
Carried weapon					73
Threatened violence					71
Beaten up					70
Mugging					64
Gang fights					61
Carried gun					56

Table 13.4. 'Non-metric simple structure': Highest item-region correlation using Spearman's ρ on Youngs data
Decimal places are omitted

Clearly, if only the highest value in each row were to be kept then the items would appear to be loading on that region only. This would suggest a clear typology within the regions, and the items would not *seem* to load on more than one factor.

In reality, the non-metric simple structure in Table 13.3 showed that the items were correlating with all regions positively, though they were correlating *mostly* with the region they were classified into by the faceted regional interpretation of Figure 13.5. However, Table 13.4 gives the impression of unique correlations on one region by ignoring the variance of the item 'explained' by the other regions. This procedure would lead to a misrepresentation of the domain and an incorrect hypothesis of structure.

The selection of items for the non-metric simple structure was based on regional interpretation. However, the selection of items for the factor analysis based on the SSA-I space is explored in Empirical Study 13.3.

Empirical Study 13.3: The Factor Analytical Interpretation of the SSA-I Space

Both the factor analytical factor loadings matrix and a partitioned SSA-I space are representations of the similarities between the items sampled from the content universe. The structure of the content universe under the faceted approach is hypothesised to be found in terms of contiguous regions in that space. The factor analytical hypothesis of structure items is a linear combination of variables that explain variance, the representation of which has been shown to be biased.

The way factor analysis selects items from the content universe can be shown by the superimposing the factors onto the SSA-I space. This is not unreasonable, given the close relationship between the SSA-I initial approximation and the loading plots of principal component or factor analysis.

To achieve this, the SSA-I plot of items was examined and each item was labelled according to the factor it belonged to. The factor analytical solution was that one which was found in Empirical Study 12.1 and presented in Table 12.8. (i.e. the 6 factor solution found using a principal factor extraction with orthogonal Varimax rotation of the dichotomous correlation matrix.) This is briefly summarised in Table 13.5.

Factor number	Factor name	Cronbach α*	No. of items
1	Street violence and robbery	0.848	11
2	Drugs and thieving	0.857	10
3	Group delinquency	0.741	6
4	Anti-social acts and petty thieving	0.705	8
5	Loutish behaviour	0.673.	6
6	Vehicle thieves ('TWOCers')	0.764	4

Table 13.5. Details of 6 factor orthogonal solution
*Cronbach's α calculated with listwise exclusion of missing data, so some degree of error is present

The items in the SSA-I plot were labelled according to the factor analytical classification from Empirical Study 12.1, and shown in Figure 13.7, with the regional partitions indicated as the dashed lines.

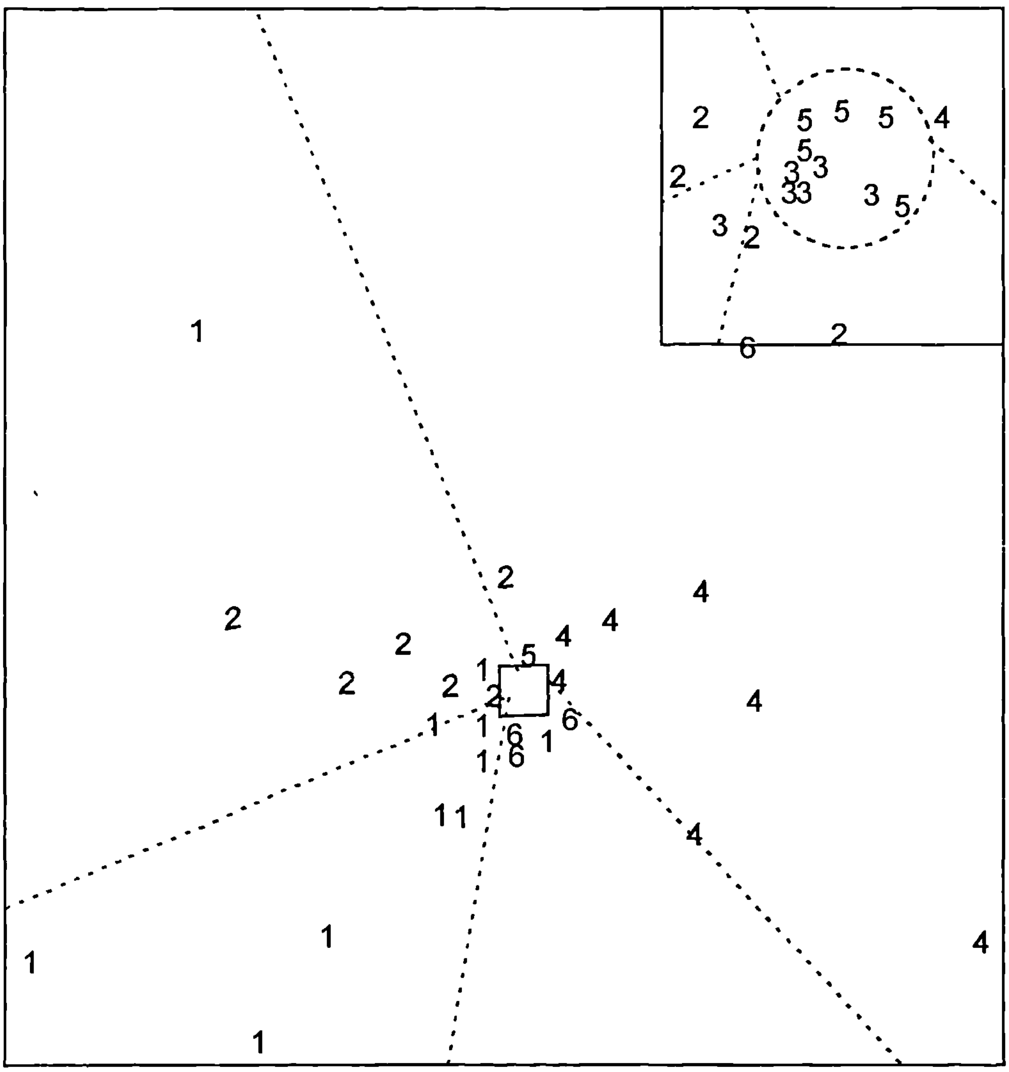


Figure 13.7. Factor analytical interpretation superimposed on SSA-I of Youngs data set

If the factors were broadly equivalent to the regional interpretation proposed in Empirical study 13.1 then the numbers in Figure 13.7 would be in distinct parts of the space and could be readily partitioned into regions. As can be seen from the plot, several of the factors indeed are in contiguous regions of the space. However, there are some marked discontinuities in the sampling of the space by the factorial linear combinations. Although the space was not perfect, with alienation of 0.19, the error in the plot cannot alone account for the fact that some factors overlap each other's 'territory'. This is especially marked in the centre of the plot, the items classified as trivial 'Incivilities' in the regional interpretation. There is no single factor which samples and accounts for all the items in the centre. If there had been then this factor would have indicated the 'general' delinquency of the Incivilities.

While the orthogonal factors are commonly interpreted as independent and not overlapping - as indeed oblique factors would be, if they had been used - this superimposition on this representation shows that the factor boundaries are 'fuzzy' and run into one another. However, the faceted boundaries of the same space were designated explicitly as contours in a continuous space, welcoming the 'fuzziness' and acknowledging the positive manifold of the association matrix.

However, the advantages of the regional interpretation using SSA-I over the factor analysis are not restricted to a favourable 'like-for-like' comparison. The additional advantage of the regional interpretation is shown in Empirical Study 13.4 as being the recovery of a structural hypothesis hidden by the factor analysis but mentioned by several sources in chapter 11, namely the extra component of intensity - the seriousness of the offending and the involvement in the offence theme.

Empirical Study 13.4: Regional Interpretation and the Extra Component of Juvenile Criminality

In chapter 10 it was proposed that there was an 'extra component' in the geometric distance plot of personality related to distance from the centroid. This was noted on account of the variability of this property even within personality scales or themes, and was suggested to be related to involvement. A similar configuration was found in the two-dimensional SSA-I plot of the juvenile delinquent and criminal actions data.

It is hypothesised that this configuration refers to what Nutch and Bloombaum (1968) referred to as the 'dimension of seriousness'. In other words, there is another facet in the plot.

This extra component to the structure of juvenile criminality is proposed to be modelled by a modulating facet of seriousness of the crime. The facet implies that the more intense the crime then the more serious the offender's criminality who commits it. Thus offenders who commit crimes at the periphery of the proposed modulating facet are more involved in that particular theme of offending. The facet is shown in Figure 13.8.

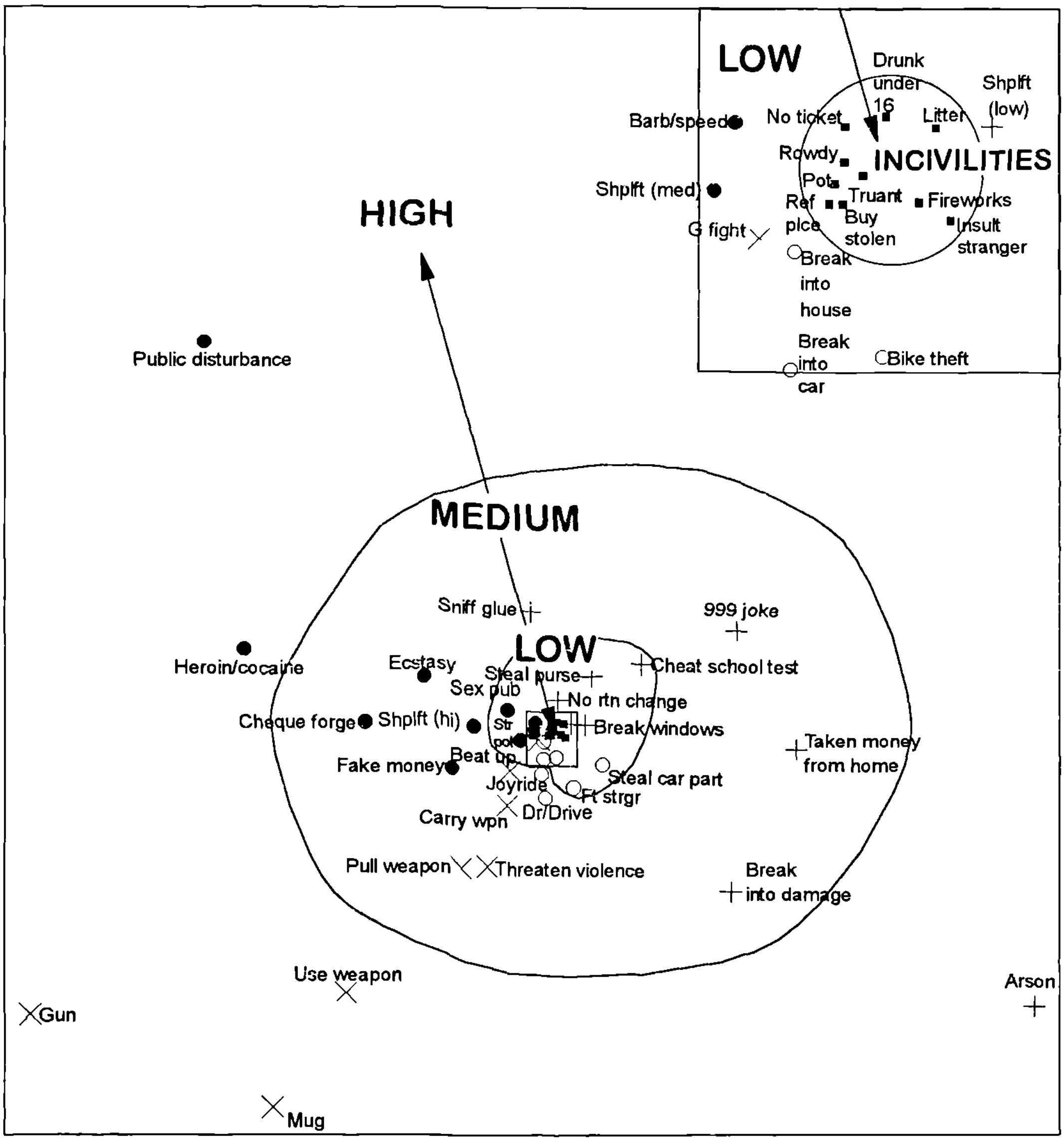


Figure 13.8. Modulating facet of involvement in the SSA-I of Youngs data set

Since the involvement decreases as the distance from the centroid decreases, this implies that the least involved crimes are those at the centre of the plot - the incivilities region. Consequently, it is hypothesised that the crime themes have their common origin in the minor delinquency items at the centre of the plot, the items that could not be differentiated in terms of the offending theme in the polar facet. However, outside these low involvement items it is suggested that there is an expansion and increase in involvement in the shape of concentric rings.

The seriousness facet acts a moderator of the other facet in the plot hypothesised in Empirical Study 13.1 and shown in Figure 13.5, namely the polar facet of offending theme. This modulating 'facet of seriousneess of involvement' acts on each region of the polar facet, meaning that an individual can be highly involved in one offence style but not necessarily in any other. In other words, there is a simplex within each element of the circumplex, giving a radex structure (Guttman, 1954b).

The interaction of the involvement in the modulating facet and the offence theme in the polar facet gives and idea of the seriousness of the crime or delinquency items. The two facets are superimposed on the same SSA-I to give a radex as in Figure 13.9. For each of the circumplicial regions, the facet of seriousness is as follows:

Drug Lifestyle.

Low: used barbs./amphet., resisted arrest, sex in public place, shoplift £10-£100

Medium: used ecstasy, forged cheque, used fake money, shoplift goods >£100

High: used heroin/cocaine, public disturbance

Escalation in this region involves both the nature of the drug and the value of thefts to pay for them. Anecdotal evidence support for the escalation of drug use from trivial to serious comes from a case study of a cocaine user who claimed 'I've tried cocaine a couple of times now, having worked my way up through the hierarchy starting with dope and proceeding through LSD, speed and E [then cocaine]' (Sphinx, 1996, p. 11). Elsewhere, Osgood, Johnston, O'Malley, and Bachman (1988) also suggested that use of marijuana led onto an escalation into more serious drugs. Using cannabis (dope) was in the Incivilities region, indicating that this Drug Lifestyle geometric simplex does indeed start in the central region then radiate out in increasing

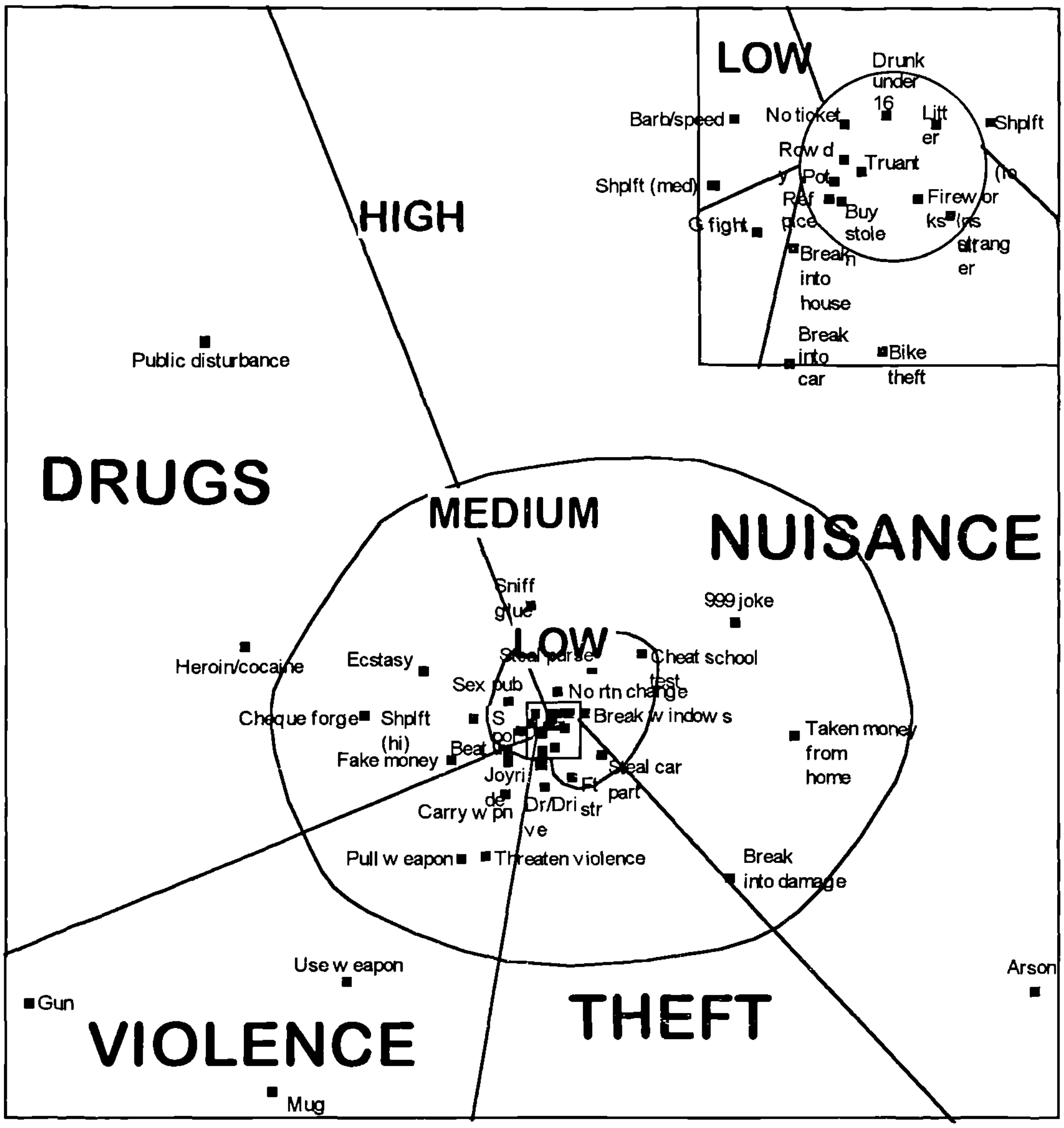


Figure 13.9. Radex of seriousness and offence theme in the SSA-I of Youngs data set

seriousness. Caution is noted with the interpretation of 'public disturbance' as being highly serious, though, due to its poor LSB.

Theft.

Low: break into house to steal, fought stranger, stolen car part, broken into car to steal, stolen bike

Medium: gone joyriding, driven while drunk or on drugs

The Theft region mainly relates to vehicles, but there are no highly serious vehicle theft items in here, explaining the space in the periphery of the region. Items which may be hypothesised to be highly serious, for future questionnaires, could include stealing cars to order and selling to professional gangs, being involved in a car chase with police, or being involved in organised thefts. The burglary item is more serious than its place in the region would warrant; however, it is a core activity for many themes and was probably 'pulled' closer by the items in the Drugs Lifestyle region.

Nuisance.

Low: shoplift goods less than £5, stolen purse/wallet, broken windows, not returned excess change, cheat at school

Medium: enter and damage building, stolen cash fr. home, dialled 999 as joke, sniffed glue

High: set fires

The escalation here can be viewed in terms of the amount of damage caused by the vandalism, or the potential for harm. A 'highly serious nuisance' would be the setting of fires, which may escalate from breaking and entering property and sniffing glue on these premises. Less serious are acts such as breaking windows, which would not require trespass onto the property. Theft of a purse is less serious than theft from home since the target is known in the latter, and is therefore more intense.

Violent Transactions.

Low: gang fights, beaten up

Medium: pulled weapon, carried weapon, threatened violence

High: used weapon, mugging, carried gun

The escalation in the Violent Transactions theme is in terms of possession of weapons, from none, to carrying and pulling knives, to using weapons for robbing, and finally carrying a gun. In contrast to the Theft region, there is a full range of seriousness items in this region.

While the elements of the polar facet of type of delinquency were validated using Cronbach's α to give a guide to test internal consistency, the modulating facet of seriousness cannot be tested in the same way. Similarly, LSB is not relevant for items hypothesised as equal seriousness since the bonding between these items is not necessarily local, and in fact equally serious items may be lawfully at opposite ends of the SSA-I plot.

Therefore, an external form of validation for the consistency must be found. This was done in Empirical Study 13.5.

Empirical Study 13.5: Validating the Extra Component of Seriousness in Juvenile Criminality

To validate the modulating facet as an extra component of seriousness, a sample of university students was asked to rate on a scale of 0-100 the seriousness of each of the 45 acts. The questionnaire given to the students used identical phrasing of the acts as was given to the delinquent sample *viz.* natural language descriptions. (The questionnaire is described fully in chapter 12.)

The sample of students consisted of 54 people aged between 18-24 years, with a mean of 18.9 years and a modal age of 18 years old. Of the 54 students, 19 said they were male and 30 said they were female, with 5 choosing not stating either gender or age. The mean seriousness given by the student sample is in Table 13.6, with ratings in decreasing order.

If these seriousness ratings reflected the hypothesised seriousness of the modulating facet then there would be a pattern of lower seriousness at the centre of the plot which gradually increased with distance from the centre. This would therefore follow the concentric circles of the modulating facet.

These mean values were superimposed as external variables onto the SSA-I space, which gave the plot in Figure 13.10.

Item	Serious -ness
Carried gun	83
Set fires	78
Break into house to steal	76
Carried weapon	76
Pulled weapon	75
Broken into car to steal	74
Used weapon	73
Mugging	71
Driven while drunk/drug	71
Gone joyriding	71
Beaten up	71
Threatened violence	68
Shoplift goods >£100	66
Stolen purse/wallet	66
Forged cheque	66
Dialled 999 as joke	66
Stolen bike	64
Enter+damage building	64
Shoplift £10-£100	64
Stolen car part	63
Gang fights	59
Used heroin/cocaine	57
Fought stranger	55
Resisted arrest	54
Shoplift goods <£5	49
Used barbs./amphet	48
Public disturbance	47
Broken windows	44
Bought stolen goods	44
Used ecstasy	42
Refuse coop. police	41
Sniffed glue	41
Fireworks in public	40
Stolen cash fr. home	38
Dropped litter	36
Insulted stranger	34
Used fake money	30
Travelled without ticket	29
Been drunk under 16	28
Used cannabis	27
Cheat at school	27
Played truant	25
Sex in public place	23
Not returned XS change	22
Acted rowdy	21

Table 13.6. Seriousness of items from Youngs data set

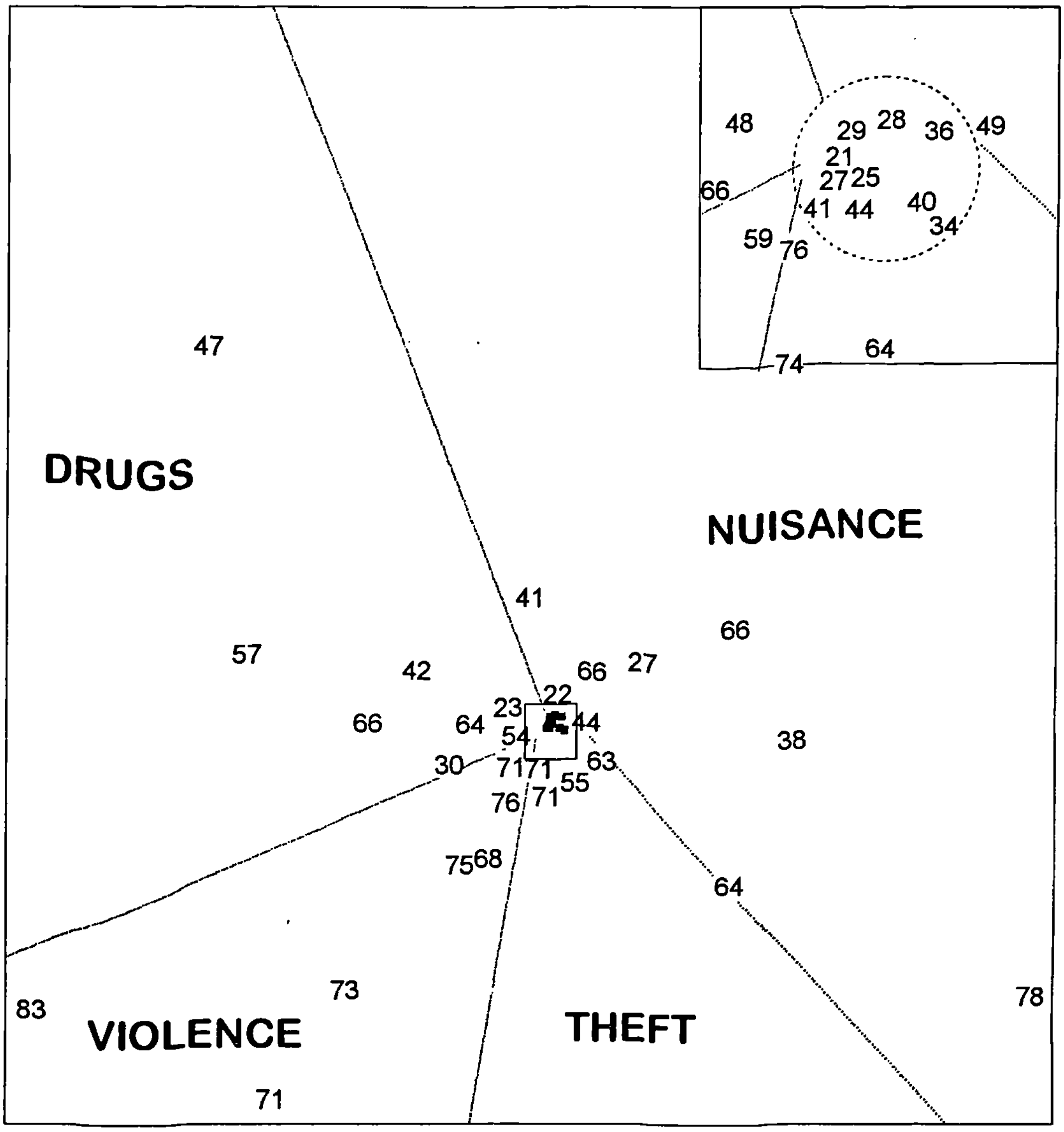


Figure 13.10. Seriousness ratings of items superimposed on SSA-I of Youngs data set

As can be seen from the plot, there is indeed a general trend for the items to increase in seriousness rating moving away from the centre, though with some are exceptions. Most notably, the central core of items in the Incivilities region are all rated as less serious than the other items outside the core in offending themes. The very outer periphery of the plot has items rated high on seriousness. There was a strong positive correlation between the distance from the centre of the Incivilities region in the middle of the plot and the mean rating of seriousness (ρ = +0.503, p < 0.000). In other words, as distance from the centre of the plot increased then mean rating of serious increased significantly.

It is also interesting to note that the Violent Transactions region is rated as more serious overall than the other regions. It would be hypothesised from this plot that the 'empty' space in the outer edge of the Theft region would contain theft items that would be rated as highly serious criminal acts by the sample. For example, an item such as a robbery on a post office or bank would be hypothesised to be in this part of the space near the boundary with the Violent Transactions region.

From this Empirical Study it can be concluded that there is external empirical support for the interpretation of the modulating facet as an extra component of seriousness in the SSA-I. On the basis of this data set, the structure of juvenile delinquent and criminal actions includes variation in both type and degree of offending. The polar facet of offence theme indicates a distinction between types of offence style which can be delineated in the highly positively intercorrelated matrix. The modulating facet of offence seriousness shows that there is are degrees of offence seriousness radiating out from the minor petty incivilities to more serious crimes. Together these two facets create the radex structure. The factor analytical representation of the data set could not reveal the radex structure, unlike the SSA-I representation.

It was hypothesised earlier that offenders who commit the highly serious criminal acts were more involved in the crime theme. These were suggestions concerning the structure of individual respondents rather than the associations themselves. These sorts of relations are usually investigated by techniques using Single Stimulus data rather than Similarities data as in the SSA-I (Coombs, 1964). Guttman scaling is one such method of looking at individuals. This endeavour was noted in Smith *et al.* (1986), who commented that analyses should simultaneously offer insights into both the interrelations among crimes and the interrelations among the careers of offenders.

The next chapter integrates Guttman scaling of the data set into the radex structure on the basis of a particular property of the modulating facet. This property is the hypothesised equivalence of the modulating facet, the simplex configuration and the Guttman scale (Guttman, 1954b). It suggests an alternative way to investigate the structure of the juveniles' responses to the questionnaire. It develops this idea and also shows how the faceted regional interpretation can provide a rationale for testing structure with the strict representation provided by Guttman scaling.

Summary of Chapter 13

The special nature of the Youngs data was considered using the perspective of Smith *et al.* (1986). It was argued that the data was suitable for non-metric analysis using Jaccard's on account of this nature and the coefficient's handling of missing data. A faceted reanalysis of the data set on juvenile delinquent and criminal actions was performed using SSA-I, in Empirical Study 13.1. Regional hypotheses indicated some similarity of styles of offending with the factor analysis, though with better Cronbach's α reliabilities and an indication of a common generalist origin of the themes of minor incivilities. Empirical Study 13.2 used an analogy of a factor loading table - 'non-metric simple structure' - to reinforce that items correlated highly with most regions but one in particular. A biased representation of this non-metric structure would lead to conclusions of offence specialism, just like in Thurstonian simple structure and factor structure representation. Empirical Study 13.3 suggested that factor extraction partially concealed the structure in the way factor analysis would have sampled from the SSA-I space less systematically that the regional interpretation. Furthermore, Empirical Study 13.4 proposed that a further facet was present in the SSA-I representation of juvenile criminality, being a modulating facet of involvement in the offence style. This was hypothesised to be related to crime seriousness, which was confirmed by Empirical Study 13.5 with a sample of student rating the items on seriousness of criminality.

It was concluded that the faceted interpretation offers a better representation of juvenile criminal and delinquent actions for the following reasons. It exploits more fully, the nature of the data, allows full testing of structural hypotheses of versatility and specialism, emphasises the overlap of these two explanations in the present data set and discovers an extra aspect of the structure hidden by the factor analytical representation.

Chapter 14
guttman scales and the modulating facet of juvenile delinquent and criminal action

Integrating Guttman Scales into the Radex

In the previous chapter, the elements of the polar facet of theme of delinquency (intention of criminal behaviour) were validated using Cronbach's α to indicate internal consistency. These were better than the factor analytical interpretation, it was argued. Furthermore, the faceted regional interpretation hypothesised the radex structure, including a modulating facet of seriousness which was not be revealed in the factor analytical representation. The validity of the modulating facet could not be tested for internal consistency so it was necessary to find another external form of validation for the consistency and meaning of the modulating facet of seriousness. This was done using a sample of students rating the offence seriousness.

This chapter explores another possible structural property of the modulating facet. This property is the hypothesised equivalence of the modulating facet, the simplex configuration and the Guttman scale (Guttman, 1954b) and it suggests an alternative way to investigate the structure of the juveniles' responses to the questionnaire. The chapter develops this idea and also shows how the faceted regional interpretation in SSA-I can suggest items for inclusion in the strict structural hypothesis of a Guttman scale.

The Modulating Facet and the Guttman Scale

The Similarities data derived from the Youngs data set on juvenile delinquent and criminal actions were used to propose a radex structure of juvenile delinquency, with polar facet of offence styles and modulating facet of seriousness. The data point configurations in the geometric representation that are associated with these regional interpretation are a circumplex and simplex respectively (Guttman, 1954b). As was shown in chapter 3, the fundamental difference between the regional interpretation and the data point configurations is that the regions partition the whole geometric space whereas the data point configurations join the points together.

Guttman (1965b) intended that the examination of triangular Similarities matrices for configurations such as the simplex and the radex should be with quantitative data. However, it was also suggested that 'with appropriate changes in the algebra required, the entire theory can be restated for qualitative data. The perfect simplex is then analogous to a perfect scale' (Guttman, 1954b, p. 340).

This analogy can be demonstrated readily by the following example. A two-way two-mode Single Stimulus data in a perfect Guttman scale is shown in Figure 14.1.

Items					Scale Sum
1	2	3	4	5	
0	0	0	0	0	0
1	0	0	0	0	1
1	1	0	0	0	2
1	1	1	0	0	3
1	1	1	1	0	4
1	1	1	1	1	5

Figure 14.1. Typical Guttman scale

This Guttman scale shows a uniform increase from item t_1 to t_5. When this two-way two-mode matrix in Figure 14.1 is correlated on its columns with the ϕ coefficient to create a two-way one-mode matrix of Similarities data, the values shown in Figure 14.2 are found.

	t_1	t_2	t_3	t_4	t_5
t_1	100				
t_2	63	100			
t_3	45	71	100		
t_4	32	50	71	100	
t_5	20	32	45	63	100

Figure 14.2. Correlation matrix from a typical Guttman scale
Values are multiplied by 100 for clarity

As can be seen from visual inspection of the matrix, this is a perfect nonequally-spaced simplex, obeying the pattern required by Figure 3.6 of decreasing similarity when moving towards the bottom-left cell. The simplex also shows an increase in intensity going from t_1 to t_5.

Returning to the modulating facet in the SSA-I representation from Figure 13.8, it was suggested that the intensity of items at the centre of the plot was lowest and increased with distance from the centre. This increase was uniform in all directions from the lowest intensity of the centroid. Two different vectors radiating from the centroid are equal in intensity when they are equal in distance, but are different in style or themes unless they are in the same direction. This applies for all possible vectors from the centroid. In short, therefore the modulating facet would be hypothesised to be equivalent to infinitely many is a Guttman scales radiating out from the centroid, each qualitatively distinct but quantitatively equal.

However, a modulating facet acts upon (i.e. modulates) a polar facet, as was shown in Figure 13.9. Since the polar facet comprises regions hypothesised to be conceptually distinct from each other, the modulating facet acts on each different polar region *in the same way*. Therefore a modulating facet in a radex would be hypothesised to create *n* Guttman scales given a polar facet with *n* regions. For the Youngs data in chapter 13, this would be four Guttman scales. It is hypothesised that there is a functional equivalence of the two different representations of SSA-I and Guttman scale, even though 'the concepts of [content] universe and a scale are distinct and separate' (Guttman, 1950a, p. 82).

But performing Guttman scaling on the Youngs data set on juvenile delinquent and criminal actions requires the use of the original recorded observations as Single Stimulus rather than Similarities data. In other words, the two-way two-mode matrix of observations (i.e. the 'raw data') is required rather than the two-way one-mode triangular matrix. This creates no conceptual difficulties at all since the creation of data from recorded observations is a distinct phase in the Coombsian Research Model (CRM) explored in chapter 1. As Jacoby (1991) put it:

> There is never any single correct type of data that must be extracted from a given set of empirical observations [i.e. data matrix]. The interpretation of the data is always based on a combination of substantive considerations (which interpretation of the observations makes the most sense?) and analtyic objectives (which scaling procedure will produce the kind of information desired?) (Jacoby, 1991, p. 72)

In other words, derived data may be ordered and related in different ways even though they can be derived from the same set of recorded observations.

The hypothesis of the functional equivalence of the two different representations of SSA-I and Guttman scale was tested in Empirical Study 14.1.

Empirical Study 14.1: The Guttman Scale and the Simplexes of Juvenile Delinquency

For the data set on juvenile delinquent and criminal actions, the polar facet of the SSA-I representation gave the content of four themes of offending, with the Incivilities region at the centre of the plot not hypothesised to differentiate between offence themes. By the argument in the previous section, it was hypothesised that the items in each of the four themes would each create a Guttman scale. In other words, there are four Guttman scales in the representation. The rationale for this was that the themes were also differentiated by a modulating facet of seriousness.

To examine this hypothesis, the items in the four regions were then tested for scalability to find the best fitting Guttman scale. Missing data were coded as absence and the dichotomised version was used. The scales were then ordered by column and row in Microsoft Excel to allow the calculation of errors and hence the Coefficient of Reproducibility, which is 1 - [errors/(items*subjects)].

The best fitting Guttman scales increasing in order from left to right are given in Figure 14.3. These were calculated for all 207 respondents on the items defined by the regional interpretation of the SSA-I space, with the number of errors or deviations from the predicted perfect scale pattern, and the derived Coefficient of Reproducibility (Guttman, 1950b) for the scale.

Theft

Broken into car to steal	Gone joyriding	Break into house	Stolen bike	Stolen car part	Driven while drunk/drug	Fought stranger

LESS INTENSE MORE INTENSE

Errors: 281, Reproducibility: 0.806

Drug Lifestyle

Shoplift £10-£100	Used barbs./ amphet	Resist arrest	Shoplift goods >£100	Used fake money	Sex in public place	Forged cheque	Used ecstasy	Used heroin/ cocaine	Public disturb-ance

LESS INTENSE MORE INTENSE

Errors: 496, Reproducibility: 0.760

Violent Transactions

Gang fights	Beaten up	Carried weapon	Pulled weapon	Threaten violence	Used weapon	Mugging	Carried gun

LESS INTENSE MORE INTENSE

Errors: 218, Reproducibility: 0.868

Nuisance

Not return excess change	Cheat at school	Shoplift goods <£5	Broken window	Stolen purse/ wallet	Sniffed glue	Enter+ damage building	Dialled 999 as joke	Stolen cash fr. Home	Set fires

LESS INTENSE MORE INTENSE

Errors: 513, Reproducibility: 0.752

Figure 14.3. Best fitting Guttman scales from Youngs data using regional interpretation from Empirical Study 13.1

As can be seen from the Reproducibility scores, no scale exceeded the 'rule of thumb' given by Suchman (1950a) of 0.90, where 1 is a perfect Guttman scale. These do not compare favourably to Arnold (1965) who found three types of juvenile criminality in Guttman scales of theft, nuisance and violence offences, though the violent Transactions scale above was the best here but the worst for Arnold.

However, this does not mean that the Guttman scales above should be rejected outright, since any guideline on 'how precise is imprecise' must be tempered by regarding the substantive usefulness of the scale, even with a notable degree of error. There were a number of considerations given by Guttman when interpreting the meaning of the figures for Reproducibility. The first is the nature of the response categories, with dichotomies giving higher Reproducibility. Since all items were dichotomised, this was not an issue. Secondly, the frequencies of the items must be examined. Where these are all extremely high or low, Reproducibility will be artificially inflated by these unequal marginalities. Thirdly, the pattern of error may show in the scalogram may show that there are in fact two scales or quasi-scales explaining the pattern of responses.

The impact of Guttman's second and third points about item frequencies and response patterns were explored in greater detail in Empirical Studies 14.2 and 14.3 respectively.

Empirical Study 14.2: The Guttman Scales and Item Frequencies

The initial approximation for the testing of Guttman scales was suggested by Suchman (1950a) to be with the items arranged in order of decreasing frequency. Indeed, as the following four scales with item frequencies show, this was mostly the case. Figure 14.4 shows the percentage of respondents who admitted to the delinquent or criminal item in the region.

Theft

Scale item	Frequency (%)	
Broken into car to steal	76	Less
Gone joyriding	63	Intense
Break into house to steal	77	
Stolen bike	67	↓
Stolen car part	60	
Driven while drunk/drug	59	More
Fought stranger	62	Intense

Figure 14.4. Percentage respondents admitting items in best fitting Guttman scales from Youngs data

Drug Lifestyle

Scale item	Frequency (%)	
Shoplift £10-£100	71	LESS
Used barbs./amphet.	71	INTENSE
Resisted arrest	70	
Shoplift goods >£100	55	
Used fake money	56	
Sex in public place	63	
Forged cheque	44	
Used ecstasy	50	
Used heroin/cocaine	34	MORE
Public disturbance	27	INTENSE

Violent Transactions

Scale item	Frequency (%)	
Gang fights	78	LESS
Beaten up	62	INTENSE
Carried weapon	58	
Pulled weapon	49	
Threatened violence	51	
Used weapon	33	
Mugging	27	MORE
Carried gun	24	INTENSE

Nuisance

Scale item	Frequency (%)	
Not returned XS change	74	LESS
Cheat at school	60	INTENSE
Shoplift goods <£5	74	
Broken windows	68	
Stolen purse/wallet	62	
Sniffed glue	51	
Enter+damage building	42	
Dialled 999 as joke	44	
Stolen cash fr. home	43	MORE
Set fires	25	INTENSE

Figure 14.4. (cont.)

Although the decrease in frequencies was not completely uniform, the overall monotone trend can be noted. In fact, when the four scales are superimposed, this trend is even more marked, as shown in the graph in Figure 14.5.

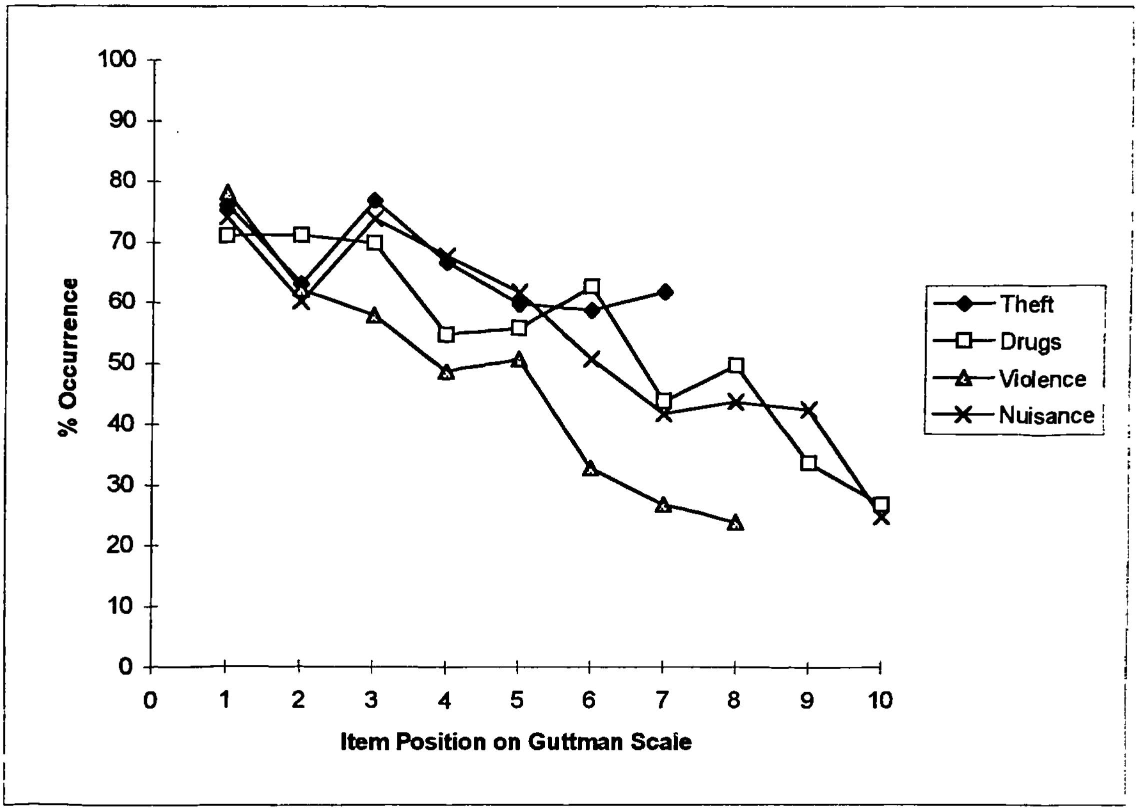


Figure 14.5. Graph of percentage respondents admitting items against position in best fitting Guttman scales

What can also be seen from this graph is that the theft scale - containing only 7 items - would be hypothesised to continue to decrease in frequency as the items became more serious and involved in the crime type, as was suggested in chapter 13.

There are two points to note from the frequencies and the trend. Firstly, the frequencies themselves are not inflating or reducing the Reproducibility in any way on account of highly skewed frequencies. Suchman (1950a) stated that some items should ideally be split around the 50% mark. The distribution of values ranges from 24% to 78%, with average 42.9% and standard deviation 15.9.

In fact, 9 out of the 10 most frequently endorsed items were not scaled since they were in the general Incivilities region. These would have been the items that would have skewed the values for Reproducibility. Therefore the Coefficients of Reproducibility given in the previous section were valid realistic and are not inflated by frequency. Also the more highly reproducible scales of Violent Transactions and Theft were more uniform in the frequency change.

The second and substantive conclusion to be drawn from the pattern of frequencies is that the more highly 'involved' or intense items are committed less frequently in all scales. It will be remembered that in Empirical Study 13.2 the SSA-I plot of items was hypothesised to contain a change in intensity represented by to a modulating facet of offence seriousness. If the Guttman scales were genuinely radiating from the centre of the SSA-I plot and were decreasing in frequency, then if the frequencies were superimposed on the SSA-I plot, it would be expected that the frequency contours would be uniform.

The frequencies associated with each item were plotted, which gave the plot reproduced in Figure 14.6.

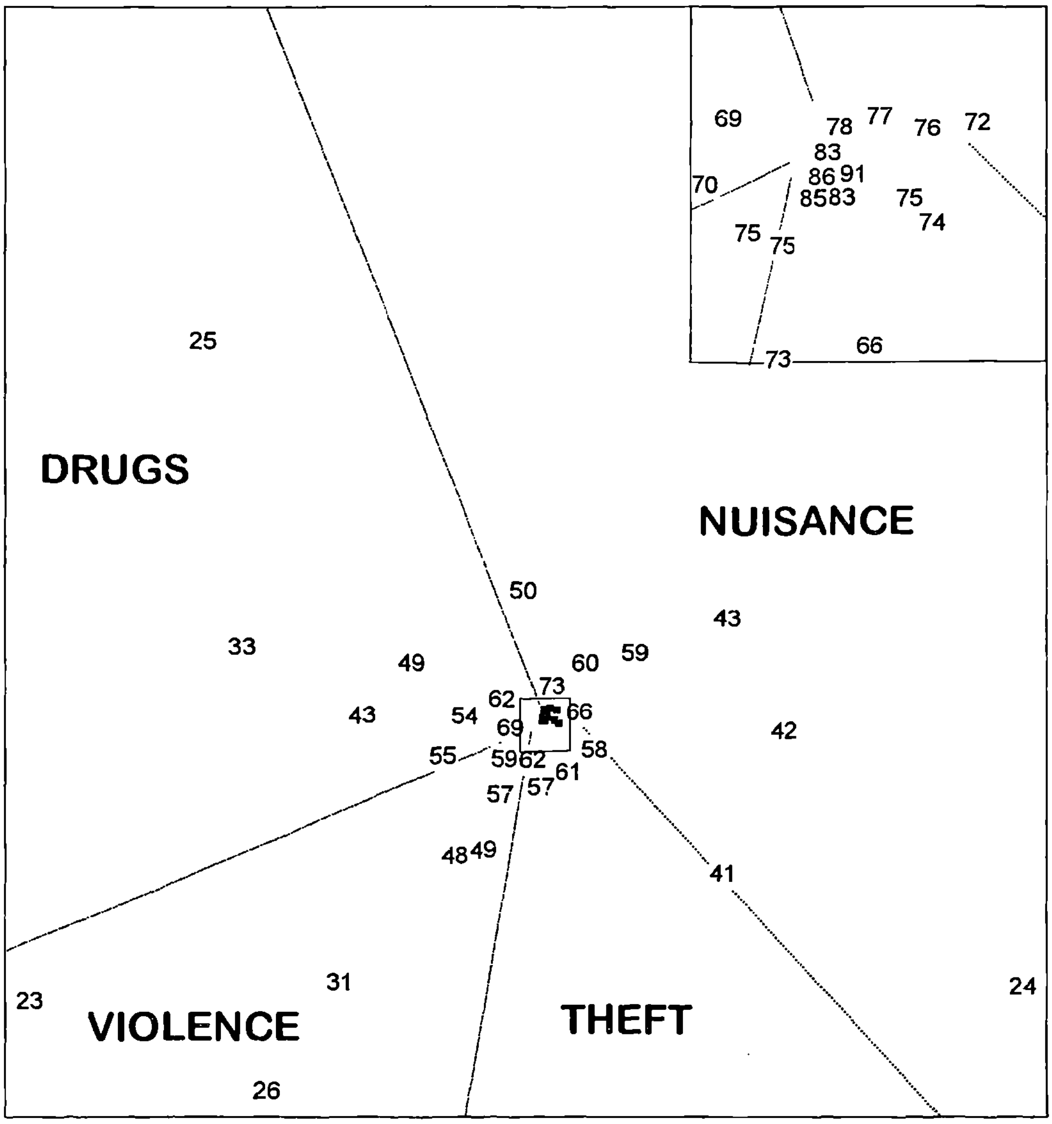


Figure 14.6. Frequencies of items in SSA-I of Youngs data

As can be seen from the plot, there is a strong 'contouring' effect to the frequencies, with the highest frequencies in the centre of the plot and a uniform decrease as the items are further from the centre of the plot. Since the space is continuous, this implies that a regular pattern should be found, limited only by the occurrence of sufficient items to indicate the contours. Bands of 15% were drawn onto this plot, and it was found that with few errors, it is possible to put items into concentric circles containing items with frequencies within 15%. Naturally, if the space were to be filled with more items it would be hypothesised that additional bands could be added with low or no error in placement. The plot in Figure 14.7 shows the bands superimposed on the same space as the original SSA-I solution.

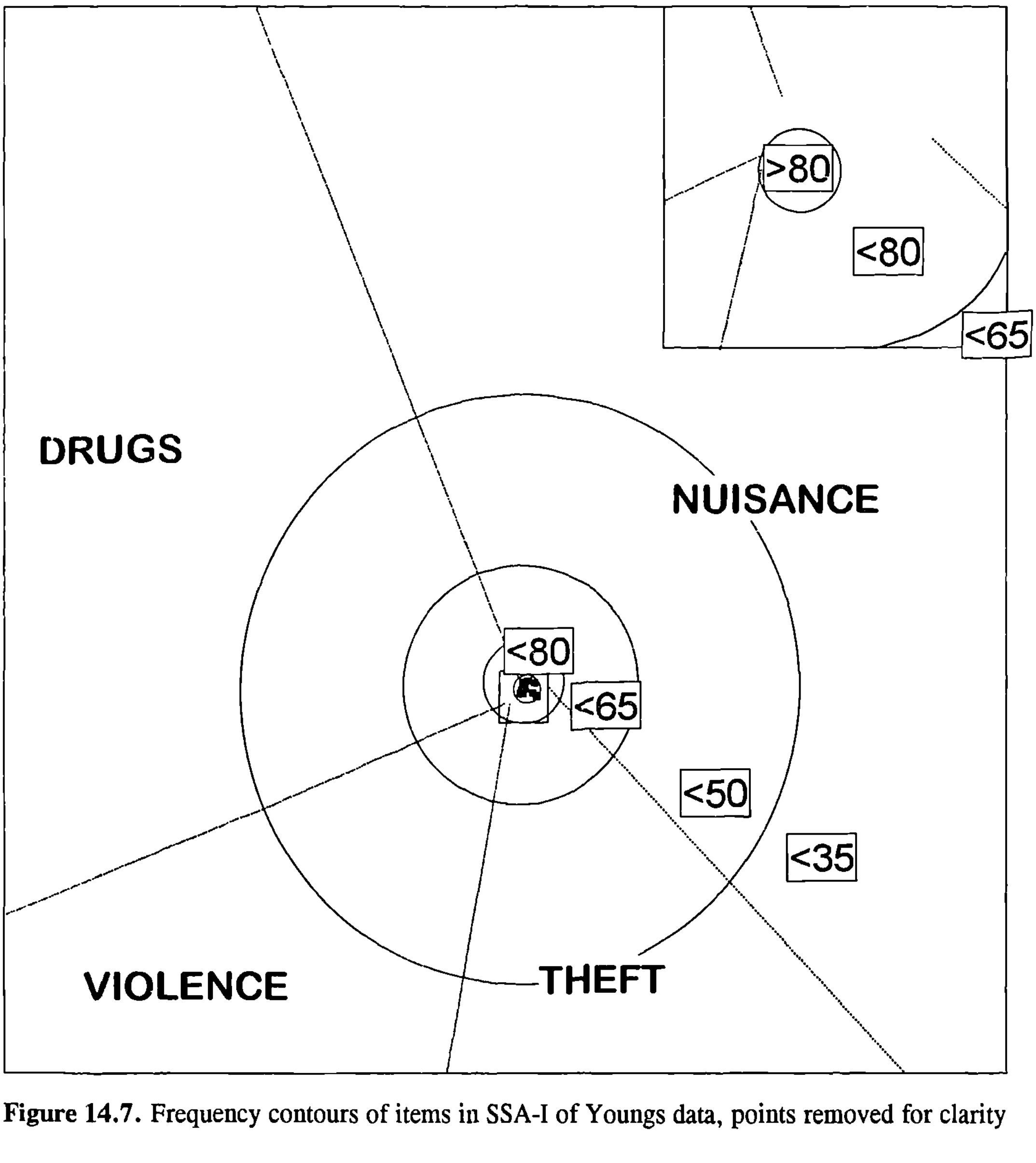


Figure 14.7. Frequency contours of items in SSA-I of Youngs data, points removed for clarity

The plot shows that radiation from the centre of the plot follows a quite uniform decrease. This also shows that the most common acts were the least serious ones as the centre of the plot, the Incivilities region. The modulating facet of seriousness hypothesised from the SSA-I variable space is backed up by the pattern of frequencies of occurrence from the respondents scores - the 'joint space' implied here has a common meaning of seriousness.

It is interesting that the two SSA-I plots in Figures 14.6 and 13.10 show trends of external variables acting in opposite directions, namely as distance increases from the centroid there is decreasing frequency (Fig. 14.6) but increasing seriousness (Fig. 13.10). In fact, there was a strong correlation between the items frequencies and the mean rating of seriousness for the offences, as measured from the sample of students introduced in Empirical Study 13.5 of the previous chapter. The correlation using PMCC between the seriousness of the act as judged by the students and the proportion of the sample who admitted the act was -0.526 (df = 53, $p < 0.01$).

To summarise Empirical Study 14.2, the slightly disappointing values for the Coefficient of Reproducibility which were below 0.9 were not skewed in any marked way by item frequencies. There was a weak relationship between the Guttman scales and the modulating facet in the polar themes. Nevertheless, the interaction between the Guttman scale and the modulating facet highlighted some interesting relations between external variables previously measured.

A second possibility for the low values of Reproducibility for the Guttman scales as suggested in Suchman (1950a) was the possibility that the scales were in fact amalgams of two scales together. This was explored in Empirical Study 14.3 by investigating the response patterns of the Guttman scales.

Empirical Study 14.3: The Guttman Scales and Response Patterns

So far, it has been shown that the four scales were approaching reasonable Reproducibility, and that the values for Reproducibility were not inflated by the frequencies involved in the analysis. Furthermore, the scales have a common meaning which was hypothesised to be seriousness and involvement with the crime theme, - a sense of 'career progression'. It was shown to be inversely related to frequency of reported occurrence in the delinquency sample but directly related to seriousness rating in the student sample. Now the distribution of responses is considered.

Figure 14.8 shows the items in increasing order in the best fitting Guttman scales for the four juvenile delinquency regions. The column headed 'Proportion respondents conforming' shows the number and proportion of the 207 respondents who had the score profile required by that exact response pattern in the Guttman scale. In other words, they conformed to the proposed Guttman scale. The column headed 'As proportion of same scale score' shows the proportion of only those respondents with an identical scale score (equal intensity) who also had the response pattern (equal type) required by that level in the Guttman scale.

Theft:

Score Profile	Scale Score	Proportion respondents conforming	As proportion of same scale score
0000000	0	6	100
1000000	1	1	27
1100000	2	0	5
1110000	3	2	17
1111000	4	1	18
1111100	5	3	15
1111110	6	5	29
1111111	7	25	100
Total conforming		43%	

Drug Lifestyle:

Score Profile	Scale Score	Proportion respondents conforming	As proportion of same scale score
0000000000	0	6	100
1000000000	1	2	29
1100000000	2	1	11
1110000000	3	1	14
1111000000	4	0	5
1111100000	5	0	5
1111110000	6	1	17
1111111000	7	1	11
1111111100	8	4	31
1111111110	9	7	54
1111111111	10	4	100
Total conforming		29%	

Figure 14.8. Respondents conforming to best fitting Guttman scale and as proportion of all respondents with same scale score in Youngs data

Violent Transactions:

Score Profile	Scale Score	Proportion respondents conforming	As proportion of same scale score
00000000	0	13	100
10000000	1	8	63
11000000	2	6	23
11100000	3	3	23
11110000	4	5	37
11111000	5	2	50
11111100	6	3	87
11111110	7	8	55
11111111	8	8	100
Total conforming		57%	

Nuisance:

Score Profile	Scale Score	Proportion respondents conforming	As proportion of same scale score
0000000000	0	4	100
1000000000	1	2	50
1100000000	2	1	13
1110000000	3	2	17
1111000000	4	1	10
1111100000	5	0	5
1111110000	6	2	11
1111111000	7	2	19
1111111100	8	1	11
1111111110	9	1	21
1111111111	10	3	100
Total conforming		21%	

Figure 14.8. (cont.)

For example, in the second row of the 'Theft' Guttman scale, the score profile 1000000 is a scale score of 1. Of all the 207 respondents, 1% had this scale score. This may seem extremely low, but when only looking at those respondents who had a scale score of 1 then 27% conformed to the score profile 1000000.

Nevertheless, what can be readily seen from Figure 14.8 is that as a template to predict scores, even the best Guttman scales had difficulty in restricting the respondents to a parsimonious few set of possible score profiles. A summary of the proportion of respondents conforming to the scale is given in Table 14.1.

	Theft	*Drug Lifestyle*	*Violent Transactions*	*Nuisance*
Proportion respondents fitting scale	43%	29%	57%	21%
Coefficient of Reproducibility	0.806	0.760	0.868	0.752

Table 14.1. Proportion of respondents fitting the Guttman scales and the corresponding Coefficients of Reproducibility from Youngs data

The poor values for Reproducibility would seem to undermine the true worth of the scales when viewed in terms of conformity to predicted score profiles. Nevertheless, any criticism of the worth of the Coefficient of Reproducibility or the Guttman scales proposed above must be tempered with one important fact: even the proportion found for the Nuisance scale was way above the values that would be expected by chance.

The Nuisance scale contained 10 items meaning that there were $2^{10} = 1024$ possible structuples yet the Guttman scale hypothesised that only 11 would be found. Empirically, 21% of scores in the Youngs data conformed to these 11. This proportion of conforming values is far higher than would be expected if there were no structure to the responses - the 'nullest hypothesis' of no structure can therefore be rejected. In other words, the Guttman scales are useful as first approximations but are not rigorous enough for full predictive purposes. Given the nature of the data, this is acceptable though could be improved.

Looking at the distribution of the overall proportions conforming to the scale - the second column from the right of Figure 14.8 - shows that most of those score profiles that do conform tend to be either very high or very low on the scale. This shows a 'U' shaped distribution in the plot of frequencies against scale score.

The distribution of those scores conforming to the scale as a proportion of the score profiles of the same scale sum - the last column on the right of the main tables - shows this even more clearly. That is to say, a scale score of 4 on the Violence scale means that 4 of the 8 acts have been committed, and that of those respondents at this level of intensity then 37% were of the score profile 11110000.

All the figures in Table 14.1 must be considered alongside the fact that the combinatorial possibilities are far greater for those score profiles not near either end of the scale. Mathematically, there are $n!/[(n - r)!\ r!]$ ways of arranging r items in a

Guttman scale n items long. Thus the Violence scale with a scale score of 4 (i.e. 4 violent acts) has $8! / (4!4!) = 840$ possible combinations, only 1 of which is specified by the Guttman scale. By contrast, a scale score of 2 has only $8! / (6!2!) = 28$ combinations. Clearly, then, the finding of even a small percentage in the middle of the scale conforming to the Guttman scale cannot be ignored even if it cannot be taken as proof of unidimensionality. As Guttman suggested that the chances of finding 'a scale by chance for a sample of individuals is quite negligible, even if there are as few as six dichotomous items in the sample and as few as one hundred individuals' (Guttman, 1950a, p. 82).

The Poor Guttman Scales and their Improvement

The Coefficient of Reproducibility counts the number of errors in the scalogram. Thus the theft scale with seven items is hypothesised at level three to have a score profile of 1110000. A respondent whose score profile is 1001100 therefore has two endorsements in the wrong place. But since the items were dichotomous, the Guttman scale also implies what would *not* be found and hence the score profile actually also states by implication a list of the negatives of the items. Thus the lack of endorsement of items 4 to 7 implies endorsement of the opposite of 4 to 7, i.e. endorsing not 4 and not 7.

As was quoted by Smith *et al.* in chapter 13, there is a 'strong rejection assumption behind most scaling techniques, where a nonchosen alternative is taken as evidence that the alternative has been rejected.' (Smith *et al.*, 1986, p. 332) The rejection assumption was unsuitable for the creation of Similarities data from the Youngs survey, it was argued, yet the Guttman scaling technique requires Single Stimulus data that make this rejection assumption. Therefore it is inevitable that the Guttman scales will be of limited reproducibility, though nevertheless the empirically obtained scales in Empirical Study 14.1 are highly reasonable.

The possibilities for improvement for the Guttman scales are twofold. The first is to be more selective in the choice of items for inclusion - akin to item analysis. As has been suggested throughout this thesis, the 'creation' of scales as a goal in itself is unscientific and the faceted approach warns strongly against this in its pursuit of fully representational measurement (Coombs, *et al.*, 1970). The second possibility is to recognise that the representation of Guttman scale is excessively strict and must be weakened to allow more general hypotheses of structure to test the Youngs data. In effect this would be saying that the cost of buying information on criminal actions using Guttman scaling is too much, to borrow from Coombs (1964).

The possibility of item analysis is considered in Empirical Study 14.4 to see its effects on the Guttman scales, and the generalisation of Guttman scaling to weaker partial orders is considered in the next chapter.

Empirical Study 14.4: Creating Guttman Scales with Item Analysis

As an illustration of the problems associated with the Guttman scales, and to test directly the hypothesis of the Guttman scale as a simplex, the SSA-I configuration was taken and onto this was superimposed the directions implied by the Guttman scales. The scales were suggested to be increasing in seriousness and involvement, as was the modulating facet. If the scales were simplexes, radiating out from the centre of the radex i.e. the Incivilities, then the direction would be hypothesised to be non-decreasing in distance from the centre. This would mean that the increase in the scale score did not correspond to decrease in seriousness.

Figure 14.9 shows the results of this analysis. As can be seen, the direction is not uniformly away from the centre. With the scales of lower Coefficient of Reproducibility and score profiles prediction - i.e. Drugs Lifestyle and Nuisance - the lines tend to go back on themselves. These two regions are the ones covering the largest area of the plot, meaning that conceptually they are less strong than Violent Transactions and Theft. Consequently they are less clear in terms of simplexes for Guttman scaling.

Clearly the regions of poorest LSB are those one that have Guttman scales that 'snake back' on themselves, most obviously the 'Nuisance' scale and the 'sniff glue' item. Consequently, this scale has a poor proportion of respondents fitting the scale. Interestingly, the 'public disturbance' item at the extreme of the Drugs Lifestyle region did not distort the Single Stimulus data for the Guttman scales as much as it did the Similarities data for the SSA-I.

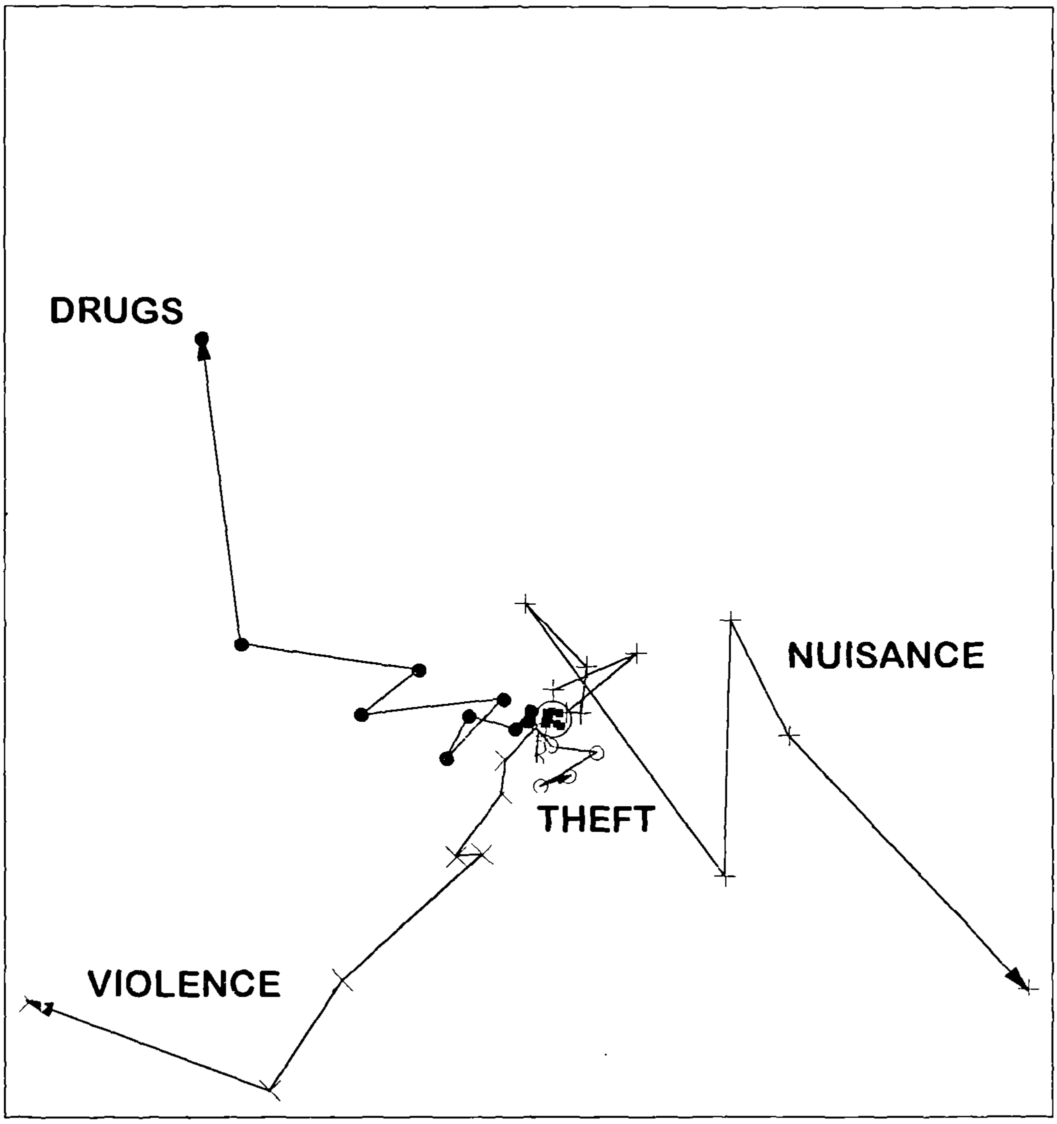


Figure 14.9. Plot of Guttman scales in increasing seriousness on SSA-I of Youngs data

As Table 14.2 demonstrates, the Nuisance region had the poorest Cronbach's α when calculated in Empirical Study 13.1, which was suggested in chapter 7 to act as a guide to LSB though not a perfect index.

It can be readily shown that these more diffuse regions with poor LSB can be improved by 'item analysis' to give better Guttman scales. This requires items to be selected or rejected for the Guttman scales on the basis of whether or not they accounted for greater numbers of score profiles.

	Theft	*Drug Lifestyle*	*Violent Transactions*	*Nuisance*
Proportion respondents fitting scale	43%	29%	57%	21%
Cronbach's α	0.789	0.820	0.853	0.726

Table 14.2. Proportion fitting Guttman scale and Cronbach's α of the corresponding region from Youngs data

Therefore by rejecting items that did not improve the proportion fitting the Guttman scale, numbers could be improved to the values shown in Figure 14.10, where the direction from left to right was of increasing involvement.

Theft

Broken into car to steal	Gone joyriding	Stolen car part	Driven while drunk/drug	Fought stranger

LESS INTENSE MORE INTENSE

Proportion fitting sequence: 58%

Drugs Lifestyle

Shoplift high value	Ecstasy	Forge cheque	Heroin or cocaine	Public disturbance

LESS INTENSE MORE INTENSE

Proportion fitting sequence: 56%

Violent Transactions

Carry weapon	Pull weapon	Threaten violence	Use weapon	Mug	Gun

LESS INTENSE MORE INTENSE

Proportion fitting sequence: 68%

Nuisance

Break windows	Cheat school	999 joke	Take money from home	Arson

LESS INTENSE MORE INTENSE

Proportion fitting sequence: 53%

Figure 14.10. Guttman scales revised using item analysis on Youngs data

Furthermore, when these scales were superimposed onto the SSA-I space it was found that the radiating lines did become more uniform and did not markedly change in direction, with only slight decreases only in Violence and Theft. This plot is reproduced in Figure 14.11.

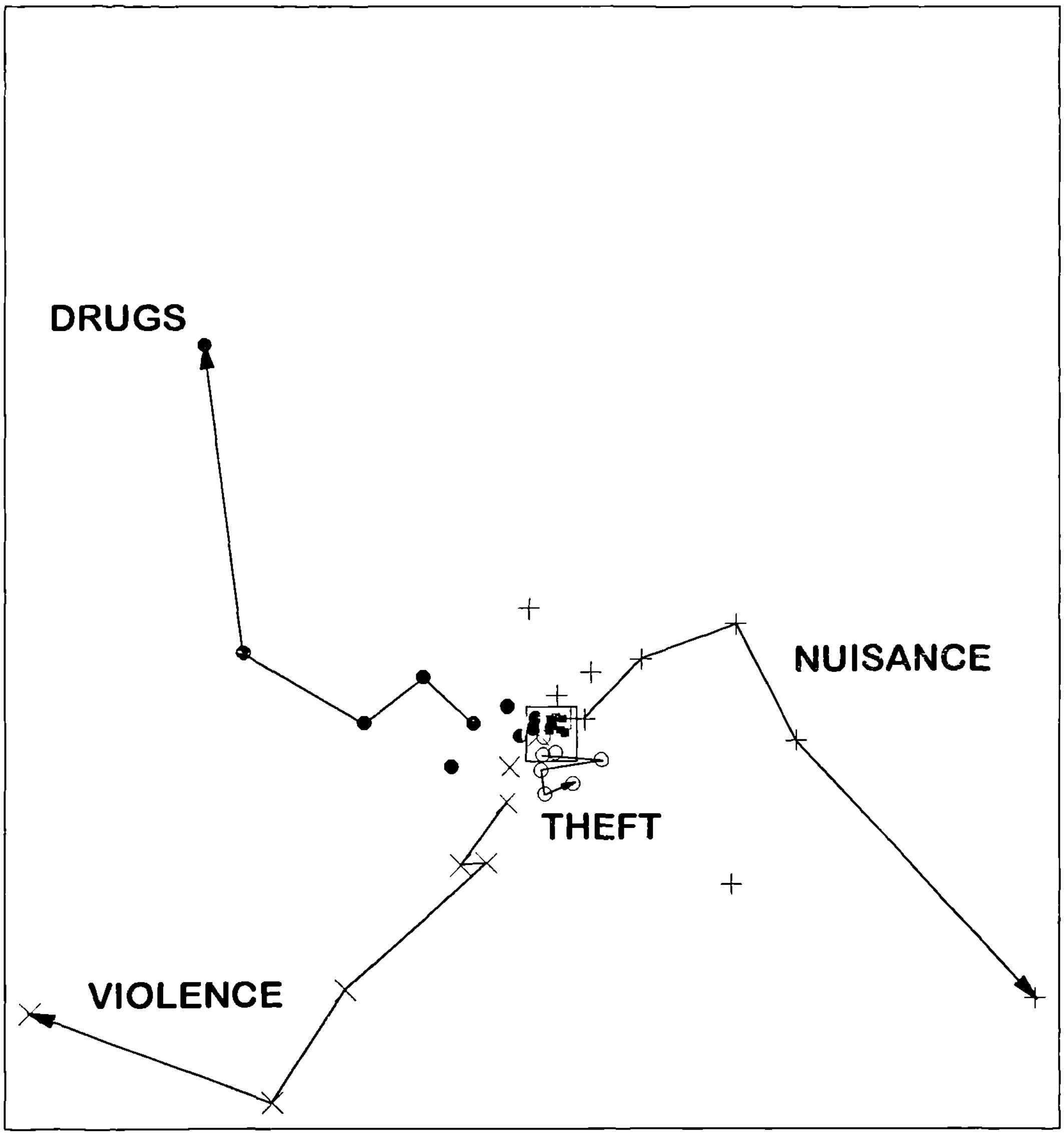


Figure 14.11. Guttman scales revised using item analysis superimposed on SSA-I of Youngs data

These modified scales also indicate that the removal of items nearer the centre of the plot also improve the proportion of score profiles correctly represented in the Guttman scales significantly.

However, removing items when they do not fit the scales rather than the rejecting of the scales when they do not fit items goes against the 'spirit of Facet Theory', to coin a phrase from Shye and Elizur (1994). Furthermore, 'in intrinsic data analysis, one does not "construct a scale", but rather one may hypothesise a [unidimensional] scale for a particular content universe' (Shye and Elizur, 1994, p. 141).

Additionally, where the regions themselves were of poorer LSB then this flaw cannot be removed even by item analysis. Table 14.3 shows that if the new scales created by item analysis were to be the new regions, then a Cronbach's α calculation on these regions would not actually improve as a result of the item analysis. In other words, item analysis cannot overcome poor LSB, as shown by the badly bonded 'Nuisance' region dropping in Cronbach's α further.

	Theft	*Drug Lifestyle*	*Violent Transaction*	*Nuisance*
Proportion respondents in scales created by item analysis	58%	56%	68%	53%
Cronbach's α (revised scale)	0.725	0.699	0.827	0.582
Cronbach's α (original scale)	0.789	0.820	0.853	0.726

Table 14.3. Proportion fitting Guttman scale revised by item analysis and Cronbach's α of the revised and original regions from Youngs data

This also shows that there is no exact equivalence between improving Single Stimulus data as measured by respondents fitting the scale and improving Similarities data by increasing Cronbach's α.

To conclude the functional equivalence between Guttman scales and the modulating facet acting on polar regions was demonstrated only to an extent, but it is preferable to improved representation by weakening the structural hypotheses rather than resorting to item analytical manipulation.

Shye (1997) suggested that Guttman started separately 'two roads to prediction' of firstly scalograms (Guttman, 1950b) and secondly content configurations or SSA variable space (Guttman, 1954b). However, Shye suggested these two could be linked through the common use of the coefficient E^* (Shye, 1985a). The choice of the coefficient is important where it would change the distribution of ranked values of the association matrix. The faceted analysis of the criminal actions data in chapter 13 used Jaccard's index as the association measure since it was the most appropriate for these data. Guttman scaling takes absence in coding items to be more significant than Jaccard's.

Therefore it is possible that the functional equivalence of the modulating facet acting on polar regions and independent Guttman scales could be found with E^* though not with these particular data - an important study to be followed up.

An alternative for the present thesis however would be to find a different model with which to test the Single Stimulus data which retains the possibility of examining the structure of seriousness. One such alternative to use a partial order representation, a weaker but more general form of the strict order (Coombs, 1952), which is tested in Chapter 15.

Summary of Chapter 14

Even though their data types are different, Guttman scales were suggested to be fundamentally the same as simplexes. Since the simplex is the data point configuration of an axial facet, and the modulating facet is a special instance of the axial fact, it was hypothesised that there would be some connection between the modulating facet acting on polar regions and Guttman scales. Empirical Study 14.1 used the data set on juvenile delinquent and criminal actions to test whether the modulating facet in the radex representation acted as four Guttman scales for each of the four polar themes of offence style. Modest support was found for this hypothesis, with poorer scales found in the more conceptually diffuse regions. Explanations for the poor Guttman scale Reproducibility this were explored in Empirical Study 14.2, looking at item frequencies, and Empirical Study 14.3, looking at response patterns. It was found that item frequencies actually supported the view of the modulating facet as indicating seriousness, though the response patters were poor for middle scores in the Guttman scales. In terms of prediction, the Guttman scales were suggested to be quite poor. It was shown in Empirical Study 14.4 that the counter-scientific practice of item analysis could be used to improve prediction in terms of score profiles accounted for. Item analysis was suggested to be ineffective with poor LSB.

Chapter 15
partial orders and the modulating facet of juvenile delinquent and criminal action

Deviation from the Perfect Guttman Scale and the Partial Order

The previous chapter suggested the relation between the modulating facet on polar regions and the Guttman scale. This hypothetical relationship was explored using the data set on juvenile delinquent and criminal actions. However, the hypothesised structure was not shown to represented adequately by Guttman scaling even though this was backed by external criteria of frequency of occurrence and ratings of seriousness by an independent sample. This was particularly so for the more conceptually diffuse Nuisance and Drugs Lifestyle regions, which occupied larger amounts of the SSA-I representation. The poor performance of the scales could be improved by item analysis, though there was suggested to be preferable alternative. This would be to use a partial order representation, of which the Guttman scale is actually a special instance. The ability to model partial order in a rational and non-arbitrary way is essential given Coombs' suggestion that '*behavior is intrinsically partially ordered*' (Coombs, 1964, p. 285, emphasis in original).

Guttman stated from the outset that the regular discovery of highly reproducible Guttman scales was 'not to be expected in practice' (Guttman, 1944, p. 140). As Coombs (1964, p. 280) put it, scalogram analysis for Guttman scales had a 'low tolerance' for inconsistency. Within a Guttman scale, the reasons for endorsing a particular level of intensity by endorsing certain items may be different for different individuals. But even slight differences in emphasis and meaning for individuals are removed through the use of ranks rather than absolute differences among scale variables, with each item being weighted equally in the scalogram. This must be contrasted with for example multiple regression, where weights are applied to each item. However, the weightings in such regression are a mathematical device to maximise fit to the regression line, meaning it is widely accepted that the cherished weights must be recalculated for each new sample, and often even for retesting the same sample. Replication, the essence of a scientific approach, becomes a statistical but not substantive issue in such an approach. Guttman (1944) suggested that this was the advantage of scaling over regressing and prediction, since a scale should remain invariant with respect to the sample.

But while the rigour of the Guttman scale may be excessive, as it was in chapter 14, partial orders are by contrast relatively accommodating. The key difference between strict and partial orders is in the notion of noncomparability of score profiles.

The strict order has score profiles which are all mutually comparable. So for three score profiles x, y and z the following relations are found: $x > y$ and $y > z$, implying further that $x > z$. So in a Guttman scale with two test items, where the symbol '>' means 'more intense', by transitivity it can be reasoned that if 11 > 10 and 10 > 00 then 11 > 00 - which is true. By saying that 11 > 00 by the transitivity rule this implies that both of the items in score profile 11 must be at least equal to or greater than the respective item in the score profile 00. For a strict order, therefore, *all score profiles are comparable.* In other words, as the intensity increases on the Guttman scale then each item in a score profile must be at least as intense as each respective item on any other score profile.

In the partial order, comparability is not always found and two score profiles may have some categories higher and some lower when each category is compared. In the partial order, *some score profiles are noncomparable.* Noncomparability can occur even when the score profiles are quantitatively equivalent - namely if their scale scores were equal - and if each score profile was still transitive with respect to other score profiles. More generally, take any four score profiles x, y, y^* and z where y has the same scale score as y^*. Both the following lines of reasoning are true: $x > y$ and $y > z$, therefore $x > z$; and also $x > y^*$ and $y^* > z$, therefore $x > z$. But here, y and y^* are not necessarily identical, $y \neq y^*$, even though they are quantitatively equivalent though noncomparable. This set of relation can be modelled only if the scale is partially ordered, which denotes the key difference with strictly ordered scales.

For example, with the two item Guttman scale described earlier, $x = 11$, $y = 10$, $y^* = 01$ and $z = 00$. It is true that 11 > 10 and 10 > 00, namely $x > y$ and $y > z$, and it is also true that 11 > 01 and 01 > 00, namely $x > y^*$ and $y^* > z$. Both these lead to the conclusion that 11 > 00 since $x > z$. Yet clearly $y \neq y^*$, and they are noncomparable even if they are quantitatively equivalent with scale scores of 1.

A partial order is required in order to model noncomparable and comparable score profiles at the same time. Using Stevens' typology, this would place partial orders between nominal and ordinal levels of measurement (Coombs, 1952). Strict scales, where all structuples are comparable, would be an ordinal level of measurement, since they contain more information than just statements of equivalence as in nominal scales (Coombs, 1953). The instance of the partial order further demonstrates the inadequacy of the four-fold typology of 'scales of measurement'. In fact, it has even been claimed that if researcher had restricted themselves to measurement on only the four levels, the existence of partial orders would have been permanently obscured (Velleman and Wilkinson, 1993, p. 40).

Hasse Diagrams: Graphical Representations of Strict and Partial Order

Ordered scales of all types from nominal to ratio and beyond can be represented schematically to emphasise different aspects of the mathematical axioms that make the scales (e.g. Coxon, 1982, p. 6). Thus a simple line may graphically represent the unidimensionality of a Guttman scale. Similarly, it is possible to represent partial orders through Hasse diagrams (e.g. Shye, 1978b).

Hasse diagrams are two-dimensional representations of a structure hypothesised to be a partial order. There is no necessary restriction to two dimensions, with the mathematics and representation required to go to higher dimensionalities only recently becoming available (Shye, Magen and Goldzweig, 1997). In a Hasse diagram, the relationship amongst structuples is represented such that lines connecting any two structuples means the structuples are comparable. Hasse diagrams can also represent the direction of intensity, as in the previous section of structuples being 'more intense' on some underlying continuum. In the Hasse diagram in Figure 15.1, the direction of the arrow indicates that the structuple is more intense than the one it points to.

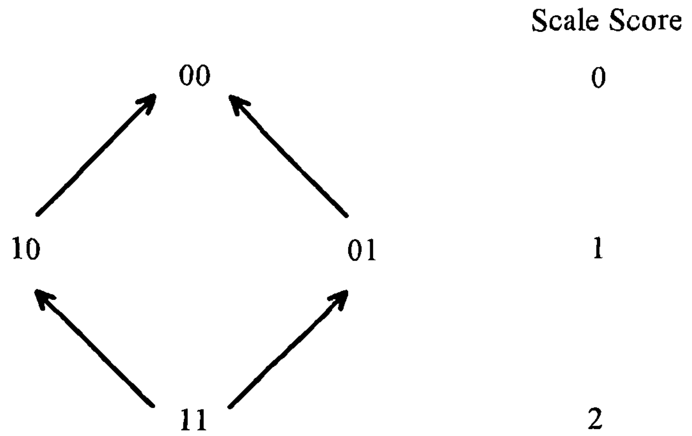


Figure 15.1. Partial order representation of two items

From this Hasse diagram it is possible to see that 11 > 00 by deriving it from both the left and right edges of the diamond; therefore it is not necessary to put an arrow showing comparability between 11 and 00. In a perfect scale, the Hasse diagram would show a succession from the most intense score profile to the least intense score profile, with each score profile connected to the next by a single line.

As the number of items in a score profile increases, the possible number of score profiles increases exponentially. For example, if the items were all dichotomies, then the set of possible score profile is 2^n where n is the number of items. More generally, all possible score profiles are given by the Cartesian product or set of $A = A_1 A_2 .. A_n$ where each component set contains the categories 1 .. i. (Shye, 1985, also uses this as a technical description of a facet, i.e. a component set of a Cartesian set.)

In the special case of a Guttman scale with n items where each item is dichotomous, it would be hypothesised that only n+1 of the possible 2^n score profiles will be found. Therefore, while the set of possible score profiles increases exponentially, the number of hypothesised score profiles increases only linearly. This demonstrates that the more items in an hypothesised scale, or the more categories in an item, the harder it is for score profiles conform to the Guttman scale, if there were a degree of error. That is to say, finding a reproducible Guttman scale where score profiles have many items or many categories in items leads to particularly strong conclusions.

For a partially ordered scale with dichotomous items, the minimum number of score profiles is $n+2$ to a maximum of 2^n. This means that the addition of one extra score profile to a Guttman scale containing $n+1$ score profiles creates a partially ordered scale. As this number of extra score profiles increases, the dimensionality required to model accurately all these score profiles also increases. In a similar way to SSA-I, the smallest dimensionality is sought to do this.

An interesting relationship emerges from the Hasse diagram of the Cartesian set and Guttman scales. The relationship is that any lawful path from lowest scale scoring ('minimal') score profile to highest scale scoring ('maximal') score profile is a strict order, which therefore may be established empirically as a Guttman scale. A lawful path is one which increases scale score each step and follows the lines of comparability. A partial order therefore comprises at least one but usually more strict orders, with some noncomparabilities between the strict orders. The partial order would increase the number of possible score profiles accounted for, as compared to the strict order by itself.

This relation was applied to the Guttman scales found in the delinquency data to see if the partial order and Hasse diagram representations could improve the modelling of the data.

Empirical Study 15.1: Parallel Guttman Scales in the Violent Transactions Scale

To examine whether it would be possible to derive a second parallel Guttman scale from the same data, and whether this improved the number of score profiles represented, the Violent Transactions scale from the study of juvenile delinquents was re-examined.

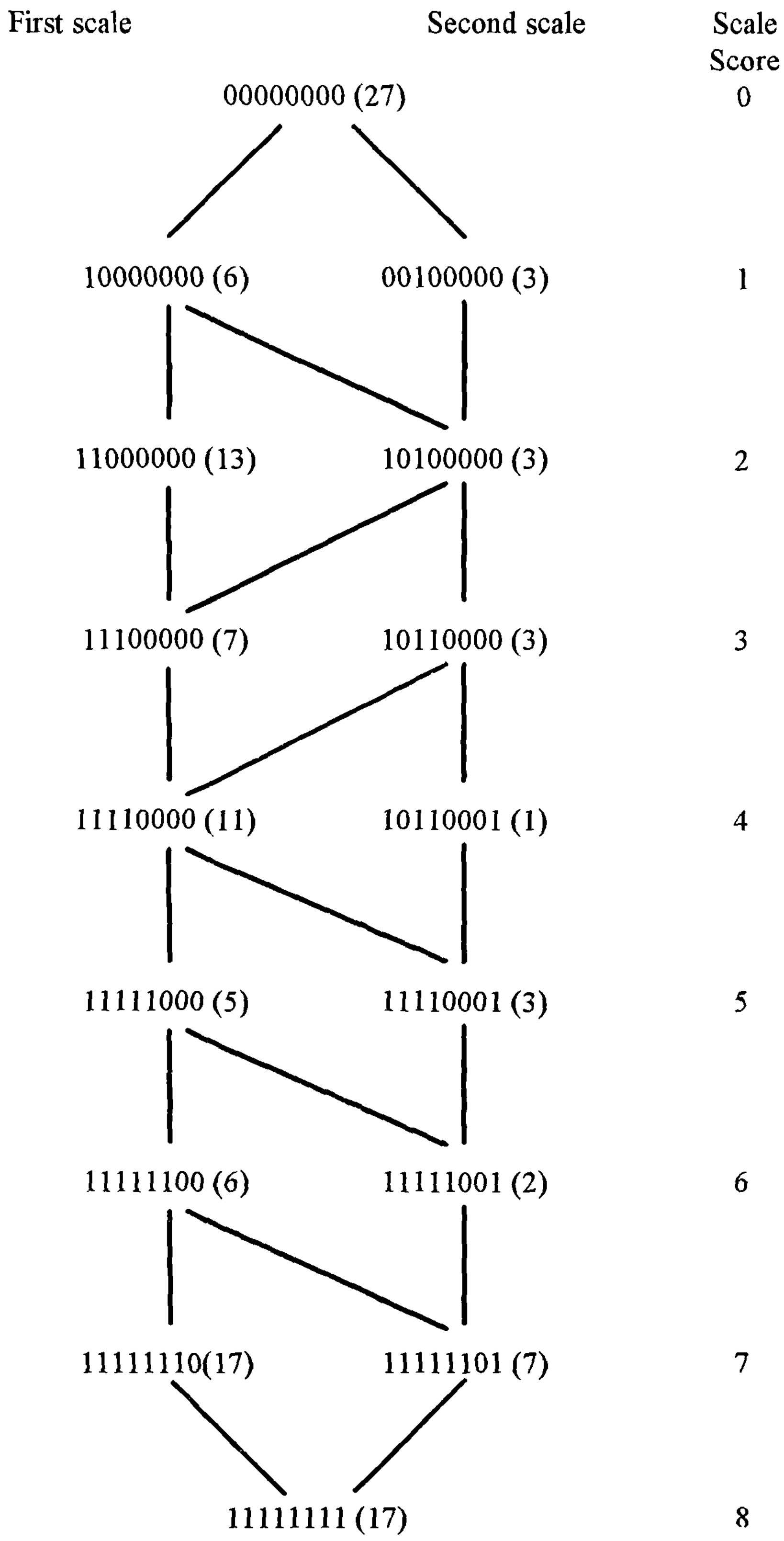
It will be remembered that this particular scale accounted for 51% of the empirically-observed score profiles in the strict order of the best fitting Guttman scale. In a parallel Guttman scale, a partial order exists due to the noncomparability of score profiles. To test for a parallel Guttman scale in the Violent Transactions scale, a second best fitting Guttman scale was taken from the data in addition to the one used in Empirical Study 14.1.

For the Violent Transactions scale, it was found that the second best fitting Guttman scale accounted for 22 extra respondents, not counting the duplicated minimal and maximal score profiles. This would mean that the partial order created by the combination of the two strict scales would contain 62% of the respondents, an increase of 11%. The Hasse diagram in Figure 15.2 shows these two scales combined as a partial order, with the original scale to the left, the additional scale on the right, and the numbers of respondents with that score profile in brackets.

The key differences between these two scales is that the second scale on the right had the items 'carried weapon' (i.e. other than gun) and 'carried gun' earlier than in the original scale. Those appearing later are 'gang fights' and 'beaten up'. Therefore the emphasis and differential meaning of the second scale is more suggestive of going prepared for violence though not actually using it.

Also on the plot was a series of comparabilities between the score profiles. This is inevitable in Guttman scales created from the same data unless the individuals in the sample are responding with precisely two mutually exclusive score profiles in each scale level, creating unrelated Guttman scales. This would not be expected empirically, though approximations to this - i.e. two scales with little comparability - would suggest that there are in fact two types of individual in the sample. The close relation between the first two Guttman scales suggests that they have a common meaning, and should not be taken as independent. In fact, the third best fitting Guttman scale from the remaining score profiles, which explained 15 or 7% more respondents, was midway in meaning between the first two.

The fact that additional scales are explaining a useful amount of extra score profiles and adding extra meaning suggests that the dimensionality of the scales in the data is not unity.



Key:

Gang fights	Beaten up	Carried weapon	Pulled weapon	Threaten violence	Used weapon	Mugging	Carried gun

LESS INTENSE MORE INTENSE

Figure 15.2. Partial order representation of two Guttman scales Violence scale and numbers accounted for in brackets, from Youngs data

Representation in Hasse Diagrams: Partial Order Scalogram Analysis

There is no restriction to finding partial orders from scalograms which are parallel strict orders, though as can be seen above this is a useful model. However, as the number of score profiles in the partial order increases the complexity of the Hasse diagram increases and includes crossed lines which obscure the data relations. Topographically, this does not present a problem, but this may suggest that far more dimensions are required by the partial order. The principle of parsimony would require that the smallest acceptable dimensionality should be chosen. To achieve this - where data are error-prone - the low frequency score profiles could be ignored on the basis that these are unreliable.

A better way around this is to emphasise those score profiles which are repeatedly found in the data, ensuring that these patterns of responses are preserved. (In effect, this acts analogously to the local monotonicity weighting from chapter 7 emphasises error in substantive ways.) This was done in the method of successive best fitting Guttman scales above. A computational method which does this algorithmically is Partial Order Scalogram Analysis (POSA). A program to perform POSA - using an index of goodness of fit and avoiding trial-and-error fitting - was provided by Shye (1985) and was named Partial Order Scalogram Analysis by Base Coordinates (POSAC).

There are two different forms of POSAC: the distributional and structural approaches. The distributional approach considers frequency to be of importance, therefore if the same score profile is repeatedly found empirically then the Hasse diagram must reflect this by ensuring that lines of comparability are most accurately represented on that score profile. Consequently, score profiles of low frequency may represent noise or error in the data so may be deleted if they obscure the partial order representation. In the structural approach, however, all observed score profiles are taken as providing evidence for the existence of a particular hypothesised structure, and is suited to the use of reliable data sources with few items or facets being analysed.

For both these forms of POSAC, a loss minimisation routine is used to find the best-fitting computational solution. For POSAC, the function is simply a maximisation of the proportion of correctly represented comparability-noncomparability relations between all score profile pairs in a two-dimensional space. This is termed the CORREP (CORrect REPresentation) coefficient (e.g. Shye and Elizur, 1994). Perfect POSAC solutions are found when the scalogram is a Guttman scale, a double scale or a diamond configuration (Shye, 1985). In these cases, CORREP is 1. Scores less than 1 but greater than 0 indicate the relative proportion of score profiles that have had their order relations preserved.

The issue of a 'good' or 'significant' value of CORREP is generally guided by the substantive usefulness of the representation - namely, whether or not the solution useful even if *X*% of score profiles are incorrectly summarised. The proportion of correctly represented pairs in percentage terms is calculated by (CORREP+1)/2. An abnormally high value may be indicative of a partial order dimensionality of three or more.

Each point can be identified by the roles played by the items in creating the partial order. This is done after finding the best solution. Since the items are not in a strict order this implies that there must be at least two items that are different qualitatively and provide the poles of the partial order. Items which only accentuate or moderate the effects of these polar variable are quantitatively different but qualitatively equal. Therefore if the items of a perfect Guttman scale were examined then no items would be polar, being all accentuating with respect to each other. The POSAC program tests the fit of each item to the different potential roles they could play in creating the partial order.

Furthermore, for each score profile, a point is placed in space and is identified by a Cartesian (x, y) coordinate. POSAC attempts to place points so that if for a point A either x or y or both x and y is greater than for point B, then A and B are comparable. Additionally, the points in the POSAC solution obey the Principle of Contiguity (Foa, 1958) that items which are conceptually similar with be found close together in a multidimensional concept space. This means that score profiles which have many structs in common are close, while those which have few are separated. Furthermore, if score profile A is as quantitatively different from B as A is qualitatively different from C, then the distance between A and B will be the same as from A to C.

Where score profiles are substantively linked, their response patterns will be similar, or at least different in consistent ways. In other words, the number of item roles will be low since the items will be conceptually related and hence identical in their contributions to the partial order.

Empirical Study 15.2: The Partial Order Structure of the Modulating Facet

The four scales were put into separate distributional POSAC runs using all the items drawn from the regional interpretation of the SSA-I solution earlier. The CORREP coefficients of the four scales are given in Table 15.1, and compared with the results of the Guttman scales.

Structural Hypothesis	Parameter	Type			
		Theft	Drug Lifestyle	Violent Transaction	Nuisance
Guttman Scale	Coefficient of Reproducibility	0.806	0.760	0.868	0.752
	Proportion Correctly Represented	43%	29%	57%	21%
Partial Order	CORREP coefficient	0.8408	0.7167	0.8668	0.6830
	Proportion Correctly Represented	92%	86%	93%	84%

Table 15.1. Comparison of Guttman scale and Partial order representation of four regions of juvenile criminal action in Youngs data

As can be seen from this table, the relaxation of the strict order of Guttman scales to a partial order allowed the representation of far more of the score profiles, giving a more reliable structure while not losing any of the external validation. This is because the Guttman scales are included within the POSAC solutions, and in fact tend to go through the middle of the plot and are accurately represented on account of the higher frequencies found with strict order score profiles.

Although all the offence themes improved in representation from the weakening of the strict to partial order, the best improvement in fit from Guttman scale to partial order was found in the more conceptually diffuse regions of Nuisance and Drugs Lifestyle.

The roles of the items in the four partial orders were also examined. As was stated earlier, there must be at least one polar item for each Euclidean coordinate (i.e. 2 for the two-dimensional POSAC), though the other items are not restricted in their roles beyond this. What would be hypothesised would be that most of the items would be acting to accentuate the combined *X* and *Y* polar roles, as opposed to also being polar. If most items were to be polar, then this would indicate there are at least two very different themes within the partial order.

POSAC supplies a table showing the Guttman's coefficient of weak monotonicity between the observed items and the theoretical roles, somewhat similar to a factor loading table for factor analysis. These were examined for the four delinquency types and summarised in Table 15.2, showing the highest correlation between the items in the types and the theoretical roles.

		Type			
Item Role	Item Label	Theft	Drug Lifestyle	Violent Transaction	Nuisance
Polar	X	1	1	1	1
	Y	1	1	1	1
Increasing	J	2	7*	4	5
	Q	2	0	1	2
				1	
Decreasing	L	0	0	0	0
	P	2	1	0	0
	Total	8	10	7	10

Table 15.2. Roles played by items in partial orders of four regions of juvenile criminal action in Youngs data
*Included one tied correlation with P

The overwhelming conclusion from the table is that the minimum number of polar items were found in the partial order, with most other items playing an increasing role that accentuate the effects of both polar items together. In other words, these items were more closely related to the sum of the poles than to the poles themselves.

This is important because if a perfect Guttman scale were to be put into a POSAC then the item roles would all play an increasing function. Therefore these POSAC solutions allow the investigation of how items themselves (i.e. the questions asking about delinquent or criminal acts) are related to the others, including relations in terms of increasing intensity. Conversely, they also indicate those items that are most dissimilar and are causing conceptual strain in the region.

To integrate the two aspects of the POSAC solutions together, it can be concluded that partial order representation offered a fuller picture for the Single Stimulus scalogram data better than the strict order hypothesis, though the partial order does not reveal that two distinct content universes have been sampled by the items.

Furthermore, this means that although the perfect simplex and Guttman scale have similarly ordered correlation matrices, in reality a radex containing a modulating facet should not be thought of as having simplexes but instead partial orders radiating out from the centre when using criminal actions data. The reason for the lack of finding the stronger Guttman scale structural hypothesis is suggested to be linked to the use of the Jaccard's coefficient, which was necessary on substantive grounds. However, better quality information may produce stronger links using Shye's E^* coefficient (Shye, 1985) between internal consistency and external prediction from Similarities and Single Stimulus data respectively.

Summary of Chapter 15

Guttman scales were hypothesised to explain how the modulating facet interacted with the polar facet, though it was shown that the Reproducibility was poor and the prediction of responses was inadequate. The notion of comparability and noncomparability in score profiles was introduced as an aid to representation in the scales. It was suggested weakening of the strict order to include two parallel Guttman scales could improve representation. However, Empirical Study 15.1 showed that this was also inadequate. But in Empirical Study 15.2 the use of partial orders with POSAC was suggested to improve representation markedly, especially with the Nuisance and Drugs Lifestyle regions. This study also showed that the roles of the items in the POSAC solutions indicated that items (i.e. delinquent or criminal acts) tended to play accentuating roles in the partial orders, as would be hypothesised.

Chapter 16
conclusions

Chapter Summaries and Main Points

This thesis began with an exposition of the structure of criminal actions. It was noted early on that this particular domain presented numerous challenges to scientific analysis due to inherent unreliabilities in the data themselves, in the data collection procedures and the data analysis techniques. To assess the impact and relevance of this state of affairs, it was necessary to explore in detail the very nature of the research process used to understand structure. This thread was developed by revising the notable work of Coombs (1964) on the research process and expanding the Coombsian Research Model (CRM) in Figure 1.2.

The importance of the content universe was noted in chapter 1, and allusions were made about the impact of error and unreliability in the universe throughout the research process. It was necessary to form a framework concerning the ideas of *structure*, *representation*, *data*, *definition* and *secondary information* (chapters 1 and 2). These issue are important at phases 0 and 1 of the expanded CRM, namely the selection of stimuli and responses to those stimuli. Only when these two phases act together are items created for analysis, a concept neglected by Coombs.

However, where the content universe has been defined by people external to the research then inevitably bias creeps into the research process. The use of *secondary information* (chapter 1) means that any of the CRM phases operates under a skewed content universe or *partial content universe*. Such a content universe is partial in the sense of being both incomplete and biased. This may happen in any domain and in any context where the research process relies on secondary information. Researchers in all contexts must appreciate the original information gatherer's perspective and original intended uses for the secondary information. This thesis demonstrated and assessed the effects of three sorts of secondary information in the domain of criminal actions: the FBI data (chapter 7), the Kirby data set (chapter 8) and the Youngs data set (chapter 12).

The structure of the information in the Kirby data was especially significant (chapter 8). The original intention of the information gatherer in this context was as evidence to prepare for prosecution. The information gatherer would clearly only be interested in facts that were known to have happened, could be confirmed and could be used as evidence against a suspect. Therefore in terms of data for analysis, absence of evidence could not necessarily be taken as evidence of absence, making the treatment of conjoint absence of items crucial to any analysis. The *uncertainty ratio* of items varied considerably, with some items in the analysis being highly reliable but others having poor certainty.

Researchers facing such dilemmas have two choices: firstly, the dubious items can simply be deleted to increase the reliability of the analysis, though this will lead to reduced validity by further reducing the completeness of the content universe. Alternatively, the researcher may choose to weaken the analysis so as to capture information at the lowest level of analysis necessary to recover the structure. This alternative has consequences on the choice of coefficient to associate information and create structure (chapter 8), a finding which had implications in other areas discussed later.

It was suggested in this thesis that one solution to the need for weaker analyses which still recovered structure reliably was to use non-metric representations of that information (chapter 2). In particular, Facet Theory (chapter 3) was proposed as allowing a scientific framework for both defining a partial content universe and creating structural hypotheses to test that partial content universe. This was made possible by the use of formal statements in the Mapping Sentence (e.g. chapter 11). However, clear structural hypotheses concerning research domains must be modelled using suitable representations - which then allow valid conclusions to be drawn about that domain.

Facet Theory is specifically geared towards the non-metric representations of SSA-I (chapter 5), Guttman scaling (chapter 14) and POSA (chapter 15). The SSA-I in Liverpool was explored in particular detail, and benchmarked against other implementations of SSA-I and related MDS techniques (chapter 5). The combination of these non-metric procedures and the faceted approach is *intrinsic* in the sense that the influences of external errors of approximation or distributional assumptions is minimised.

The link between hypotheses of structure and non-metric representation was shown to be through a *Contiguity Hypothesis* (chapter 3). This reinterpretation of contiguity returns it to its original meaning as intended by Foa (1958), and it was suggested that Facet researchers had been distracted by the simplicity and intrinsic nature of this requirement for testing structural hypotheses. This idea was developed further to show that contiguity as an hypothesis had powerful implications when used as an analogy to hypothesis testing by statistical significance (chapter 6). That is to say, Type I and Type II errors can equally mislead researchers using multivariate MDS methods. Guidelines were offered to minimise the chances of these errors in any applied research context for both the faceted and non-faceted approaches to structural hypotheses. This requires consideration of the notion of fit in the SSA-I representations as containing a substantive component as well as a statistical component (chapter 6).

Facet Theory as a methodological approach offers a full framework for the research process in the CRM (chapter 3), and can achieve such a goal even where data may be error-prone or gathered from unreliable sources (chapter 4). This would include the research domain of the behaviour and intention in criminal actions. To test the proposed suitability of Facet Theory to overcome the methodological challenges in the domain, the three key data sets on criminal actions were analysed in the light of the different obstacles they presented to structural analysis.

The FBI data set (chapter 7) was taken from a published classification system which was claimed to offer a definitive homicide investigation typology (Ressler *et al.*, 1992). It was shown that structural hypothesis testing with this particular data set required higher values of alienation than would normally be expected, showing that substantive considerations *must* overcome attempts at purely computational minimisation of error. Where researchers in any context have particularly strong hypotheses of structure but know that data quality is poor then it is necessary to do what many in the MDS literature regard as 'unthinkable' and accept high alienations. This was necessary for the FBI data set since the impact of local monotonicity to recover structure in the representation made alienation high. It is recommended that researchers - especially those using faceted regional interpretation - consider again the effects of local monotonicity on their representations. The requirement of global monotonicity, where error is spread evenly in the plot, may be both unrealistic and unnecessary for regional interpretation. With noisy data sources it may be possible to achieve only accuracy of representations at a local level. The impact of local monotonicity had not previously been investigated by the MDS literature.

The consequence of this is *local spatial bonding* in the representation (chapter 7), which may be approximated by Cronbach's α where association coefficients are similar. This requires the researcher to re-examine the original association matrix or correlation matrix to see how well crucial points in a plot have been translated in the non-metric representation given their empirical similarity values. This would also overcome the widely made criticism of non-metric MDS: namely, that it is possible to scale even poorly conceptually linked items, or items that are not even from the same content universe. Combined with careful faceted design and local monotonicity, local spatial bonding should ensure reliable and contiguous regions.

Local spatial bonding is needed where the correct representation of certain items is essential to test hypotheses of structure. These items may be those close to regional boundaries, or - as was shown using the FBI data - core items that determine the success or failure of structural hypotheses. Under standard default conditions of analysis, the structural hypotheses about the FBI classification failed. Practitioners must be prepared to question the utility of 'default values' in all analytical procedures. Rarely does a particular domain have precisely the same demands of an analysis as any other domain. Indeed, even within the same research process there will be differing needs for different analyses that examine slightly different structures. This is especially salient where the representations themselves are used to generate new hypotheses which are then tested iteratively.

Structural and representational influences on analysis were compared and contrasted in the context of unreliable data in the Kirby Child Sexual Abuse data set (Kirby, 1993), which was drawn from a partial content universe of police case files (chapter 8). The issue of conjoint absence and the uncertainty ratio has already been highlighted as a structural issue, and here representational issues were also shown to be significant. It was shown that the choice of association or correlation coefficient was more influential than representational issues such as dimensionality or local monotonicity. The MDS literature has focused mainly on the choice of the 'correct dimensionality' - this analysis showed that the choice of the 'correct coefficient' should be logically prior to any assessment of representational issues such as dimensionality. The use of Monte Carlo data may be more reliable in assessing such parameters as the 'correct dimensionality' or the alienation to be expected given the number of items, but the use of real-world data here increased the validity and showed the relevance of the coefficient, which would be missed by Monte Carlo data. The relevance of Jaccard's and its treatment of conjoint absence in dichotomous items was noted, and this should be extended to other domains where similar biases exist.

The importance of item design and the use of inappropriate coefficients was seen in the Kirby data set (chapter 9). Originally, the representation of the structure contained many incompatible and unusual items associated with an inappropriate coefficient. By rescaling with a better sampling of items and a different coefficient, a stronger and clearer structure was obtained. The classification was made of different types of exclusivities in item design: this added *substantive* and *logical* types of exclusivity to the more familiar *mutual exclusivity*. The deliberate inclusion of unsuitable items in analyses demonstrated their effects in reducing local spatial bonding and inevitably increasing alienation as the SSA-I strained to cope with the irregular association matrix. More broadly, any universe which is sampled by these poorly designed items will be unsystematically biased, irrespective of whether the domain is criminal actions or not, and whether the universe is partial or complete.

The Youngs survey of juvenile offending (Youngs, 1994) was introduced as a large quantitative structured self-report questionnaire design (chapter 12). Consequently, these data were tested for 'factorability' and then factor analysed. However, difficulties with the factor analysis were noted throughout. Again, it was found that using default options would have reduced the sample size drastically by the way SPSS handles missing values. It also would have resulted in 12 independent 'dimensions' of juvenile delinquent and criminal actions by rotating eigenvalues greater than unity. However, issues of the response range and reliability of the items forced the design to be curtailed somewhat. Nevertheless 6 reliable factors with face validity were found that suggested independent types.

But the factor analysis was suggested to be biased in the way it sampled from the domain and how it represented the structure. It was shown to have limited the search for structure in both non-criminal and criminal domains (chapters 10 and 11). The factor analytical interpretation compounds this limitation by removing variance at the stages of factor extraction, rotation and interpretation from the representation. Only if the researcher examines the results of factor analysis carefully will this 'variance stripping' be revealed. Ultimately, the factor analysis *deliberately* turns the complete content universe into a partial content universe in the search for Thurstonian simple structure. Only where the content universe consists of truly independent dimensions can the factor analytical representation be useful, as has been suggested for example in the domain of mental ability. Where the domain consists of many partial measures of the same overall concept - such as juvenile criminal actions where the structural hypothesis was of a blending of offending themes - then factor analysis cannot fail to confirm hypotheses of structural independence of types.

By contrast, a faceted regional interpretation of the Youngs data was shown to be better in creating more reliable regions (chapter 13). The faceted interpretation used Jaccard's coefficient, thereby becoming a more inclusive analysis using the missing information that caused trouble for the factor analysis. The sampling of items from the SSA-I space showed that the linear combinations of items for factors was not as systematic as a regional interpretation of the same space in Figure 13.7. Such a plot of the factors in SSA-I space can be used by researchers to discover which items are included in factors purely for mathematical convenience. Additionally, the offending themes were offered as an alternative to the Thurstonian simple structure to give a *non-metric simple structure* which emphasised the highly bonded nature of the regions rather than ignoring this covariance.

The alternative non-metric representations with the faceted approach to the Youngs data also revealed the extra structural component hidden by the factor analytical representation. This was shown in a modulating facet of offence seriousness, a facet of intensity or involvement with the offence theme. Another area where the component of intensity was hidden by the factor analytical representation was shown in the context of personality (chapter 10), even when visual representation is made using factor analytical results.

The importance of this modulating facet for the domain of criminal actions is to show that frequencies of the endorsement of items by respondents (i.e. commission of the acts in the Youngs data) do not by themselves add meaning as a modulating facet. Facet researchers should use the frequencies as a first step towards understanding the *meaning* of the modulating facet, rather than using them *as* the modulating facet. The facet of offence seriousness in the Youngs data could be externally validated by using a sample of students to rate the offence seriousness. Therefore it was shown how a radex can be validated firstly in terms of internal consistency of the polar facet with local spatial bonding and Cronbach's α, and secondly in terms of external prediction of the modulating facet by cross-reference to an independent source.

Guttman had suggested a link between these regional structures created by the radex to Guttman scales, which would require creating both Similarities data to classify items to Single Stimulus data to classify respondents (chapter 14). This in itself is perfectly viable, and is an instance of secondary data being created from secondary information originally used as a source for a different data type. However, where the data source is unreliable - such as the Youngs data with its 'pick any/*n*' nature - then the strength of the link is undermined. The Guttman scales found in the Youngs data were of poor reproducibility and could only be improved by resorting to item analysis. This finding demonstrates again the inappropriateness in applied domains of a strict 'rejection assumption', namely that absence of evidence is evidence of absence, as shown in the context of conjoint absence in the Kirby data. Such instances would emphasise that Coombs' assertion that 'we buy information by making assumptions' (Coombs, 1964) is valid only if the right assumptions are made.

The poor link between Guttman scales and the radex of the Youngs data should not be taken to mean that researchers should reject the use of the same information for different data types, though. On the contrary, less specific structural hypotheses or alternative representations should be used. For the Youngs data, partial order representations were used as an alternative to show that the data could still offer classifications of respondents (chapter 15).

Conclusion

This thesis demonstrated that methodology can never be divorced from the substantive theories which it is used to support or deny. It showed that methodology is far from merely an adjunct to substantive theory; by integrating methodology into structural hypothesis testing and addressing methodological issues in the real-world context then better theory development is possible. The faceted approach using non-metric representations offers a framework with which to achieve such an integration of methods and theory, though it was shown that merely using the approach is not enough in itself. However, outside the faceted approach the impact of methodological issues will be even more significant.

This thesis is targeted to many different audiences researching both criminal actions and other domains. Topics were investigated that had been neglected even by methodological purists of faceted and non-faceted persuasions. Nevertheless, this thesis still has relevance for even the novice practitioner, to whom the following basic advice is offered: read the print out and do not assume that default parameters are always best.

Young (1997) commented that 'the biggest problem in the social sciences is that researchers either study meaningful questions sloppily, or meaningless questions carefully.' This thesis illustrated a third way between these two extremes by examining meaningful questions carefully using clear structural hypotheses and appropriate representations.

references

Ackroyd, S. and Hughes, J. (1992). *Data collection in context* (2nd edition). London: Longman.

Alison, L.J. and Lee, J. (in press). The feasibility of profiling from sexual fantasies. In D.V. Canter and L.J. Alison (eds) *Profiling crimes: Past, present and future*. Aldershot: Dartmouth press.

Amar, R. and Toledano, S. (no date). *HUDAP manual with mathematics*. Jerusalem: Hebrew University Computing Center.

APA (1987). *Diagnostic and statistical manual of mental disorders* (3rd edition). Washington, DC: Author.

Arabie, P. (1973). Concerning Monte Carlo evaluations of nonmetric multidimensional scaling algorithms. *Psychometrika*, **38(4)**, 473-506.

Arabie, P. (1978a). Random versus rational strategies for initial configurations in nonmetric multidimensional scaling. *Psychometrika*, **43(1)**, 111-113.

Arabie, P. (1978b). The difference between "several" and "single": a reply to Spence and Young. *Psychometrika*, **43(1)**, 121.

Arabie, P., Carroll, J.D. and DeSarbo, W.S. (1987). *Three-way scaling and clustering*. Newbury Park, CA: Sage.

Arnold, W.R. (1965). Continuities in research: Scaling delinquent behaviour. *Social Problems*, **13**, 59-66.

Ault, R.L. and Reese, J.T. (1980). A psychological assessment of crime profiling. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Berkowitz, L. (1993). *Aggression: its causes, consequences and control*. New York: McGraw-Hill.

Bilsky, W., Borg, I. and Wetzels, P. (1995). Assessing conflict tactics in close relationships: A reanalyis of a research instrument. In J.J. Hox, G.J. Mellenbergh and P.G. Swanborn (eds), *Facet Theory analysis and design*. Zeist: SETOS.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, **117(2)**, 187-215.

Booth, D. (1994). 'Tricky Stats' No. 2. *The Psychologist*, 7, 533.

Booth, D. (1995). 'Tricky Stats' No. 3. *The Psychologist*, **8**, 196-7.

Borg, I. (1978). PAL: Point-wise alienation coefficients in multidimensional scaling. *Journal of Marketing Research*, **15**, 478-479.

Borg, I. and Lingoes, J.C. (1987). *Multidimensional similarity structure analysis.* New York: Springer-Verlag.

Borg, I. and Shye, S. (1995). *Facet theory: Form and content.* Newbury Park, CA: Sage.

Borgatti, S., Everett, M. and Freeman, (1992). *UCINET IV Version 1.0.* Columbia: Analytic Technologies.

BPS (1995). Comment on 'Tricky Stats' No. 3. *The Psychologist*, **8**, 197.

Bromberg, W. (1948). *Crime and the mind: An outline of psychiatric criminology.* London: Lippincott.

Brown, J. (1985). An introduction to the use of facet theory. In D.V. Canter (Ed.), *Facet theory: approaches to social research.* New York: Springer-Verlag.

Bryman, A. (1988). *Quantity and quality in social research.* London: Routledge.

Burgess, A.W., Hartman, C.R., Ressler, R.E. and McCormack, A. (1986). Sexual homicide: a motivational model. *Journal of Interpersonal Violence*, **1(3)**, 273-287.

Burn, G. (1993). *Somebody's Husband, Somebody's Son.* London: Mandarin.

Burt, C. (1940). *The factors of the mind: An introduction to factor analysis in psychology.* London: University Press.

Canter,. D.V. (1983). The potential of facet theory for applied social psychology. *Quality and Quantity*, **17**, 35-67.

Canter, D.V. (1985). Editor's introduction: The road to Jerusalem. In D.V. Canter (Ed.), *Facet theory: approaches to social research.* New York: Springer-Verlag.

Canter, D.V. (1993). The wholistic, organic researcher: central issues in clinical research methodology. In G. Powell, R. Young, and S. Frosh (eds), *Curriculum in clinical psychology*. Leicester: BPS.

Canter, D.V. (1994). *Criminal shadows.* London: HarperCollins.

Canter, D.V. (1995). The psychology of offender profiling. In R. Bull and D. Carson (eds), *Handbook of psychology in legal contexts*. Chichester: Wiley.

Canter, D.V. and Alison, L.J. (1998). Precursors to Investigative Psychology. In D.V. Canter and L.J. Alison (eds), *Criminal Detection and the Psychology of Crime*. Aldershot: Dartmouth.

Canter, D.V. and Fritzon, K. (1998). Differentiating arsonists: A model of firesetting actions and characteristics. *Legal and Criminology Psychology*, **3**, 73-96.

Canter, D.V. and Heritage, R. (1990). A multivariate model of sexual offences behaviour: Developments in 'offender profiling'. I. *Journal of Forensic Psychiatry*, **1(2)**, 185-212.

Canter, D.V., Brown, J. and Groat, L. (1985). A multiple sorting procedure for studying conceptual systems. In M. Brenner, J. Brown, and D.V. Canter (eds), *The research interview: uses and approaches*. London: Academic Press.

Canter, D.V., Heritage, R. and Kovacik, M. (1989). *Offender profiling: Interim report to the Home Office*. University of Surrey Psychology Dept.

Canter, D.V., Hughes, D.T. and Kirby, S. (in press). Paedophilia: Criminality, pathology or both? *Journal of Forensic Psychiatry*.

Carroll, J.D. (1961). The nature of the data, or how to choose a correlation coefficient. *Psychometrika*, **26**, 347-372.

Carroll, J.D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalisation of 'Eckart-Young' decomposition. *Psychometrika*, **35**, 238-319.

Carver, R.P. (1978). The case against statistical significance testing. *Harvard Educational Review*, **48**, 378-399.

Cattell, R.B. (1957). *Personality and motivation structure and measurement*. London: Harrap.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, **1**, 144-161.

Cattell, R.B. (1978). *The scientific use of factor analysis*. New York: Plenum.

Cattell, R.B. (1988). Psychological theory and scientific method. In J.R. Nesselroade and R.B. Cattell (eds), *Handbook of Multivariate Experimental Psychology (2nd ed)*. New York: Plenum.

Child, D. (1990). *The essentials of factor analysis* (2nd ed). London: Cassell.

Clark-Carter, D. (1997). The account taken of statistical power in research published in the British Journal of Psychology. *British Journal of Psychology*, **88**, 71-83.

Clarke, R.V. and Cornish, D.B. (1985). Modelling offenders' decisions: a framework for research and policy. In M. Tonry and N. Morris (eds), *Crime and Justice Volume 6*. Chicago: University Press.

Cliff, N. (1973). Scaling. *Annual Review of Psychology*, **24**, 473-506.

Cliff, N. (1993). What is and isn't measurement. In G. Keren and C. Lewis (eds), *A handbook for data analysis in the behavioural sciences: Methodological issues.* Hillsdale, NJ: LEA.

Clifford, H.T. and Stephenson, W. (1975). *An introduction to numerical classification.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* ***(2nd ed).*** Hillsdale, NJ: Lawrence-Erlbaum.

Cohen, J.(1962). The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal and Social Psychology*, **65**, 145-153.

Coombs, C.H. (1952). *A theory of psychological scaling*. Michigan: University Press.

Coombs, C.H. (1953). Theory and methods of social measurement. In L. Festinger, and D. Katz (eds), *Research methods in the behavioral sciences.* New York: Holt, Rinehart and Winston.

Coombs, C.H. (1964). *A theory of data.* New York: Wiley.

Coombs, C.H., Dawes, R.M. and Tversky, A. (1970). *Mathematical psychology: an introduction.* New Jersey: Prentice-Hall.

Costa, P. and McCrae, R. (1992). *Manual for the NEO PI-R.* Florida: Psychological Assessment.

Cox, T.F. and Cox, M.A.A. (1994). *Multidimensional scaling.* London: Chapman-Hall.

Coxon, A.P.M. (1982). *The user's guide to multidimensional scaling.* London: Heinemann.

Coxon, A.P.M. and Jones, C. (1980). Multidimensional scaling: Exploration to confirmation. *Quality and Quantity*, **14**, 31-73.

Dawes, R.M. (1972). *Fundamentals of attitude measurement.* New York: Wiley.

Denzin, N.K. and Lincoln, Y.S. (1994). Introduction: Entering the field of qualitative research. In N.K. Denzin and Y.S. Lincoln (eds), *Handbook of qualitative research*. Thousands Oaks, CA: Sage.

Donald, I. (1995). Facet theory: Designing research domains. In G.M. Breakwell, S. Hammond and C. Fife-Schaw (eds), *Research methods in psychology*. London: Sage.

Donald. I. (1985). The cylindrex of place evaluation. In D.V. Canter (Ed.), *Facet theory: approaches to social research*. New York: Springer-Verlag.

Douglas, J.E. and Burgess, A.E. (1986). Criminal profiling: A viable investigative tool against violent crime. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Douglas, J.E. and Burgess, A.E. (1989). Teaching Crime Classification: A Research Study. Paper presented at offender profiling workshop, University of Surrey Psychology Dept.

Douglas, J.E., Ressler, R.K., Burgess, A.W. and Hartman, C.R. (1986). Criminal profiling from crime scene analysis. *Behavioural Sciences and the Law*, **4(4)**, 401-421.

Dracup, C. (1995). Hypothesis testing - what it really is. *The Psychologist*, **8**, 359-362.

Dudycha, G.J. (1943). A critical examination of the measurement of attitude toward the war. *Journal of Social Psychology*, **39**, 846-860.

Ekman, G. (1954). Dimensions of color vision. *Journal of Psychology*, **38**, 467-474.

Elizur, D. and Shye, S. (1976). The inclination to reimmigrate: A structural analysis of the case of Israelis residing in France and in the USA. *Human Relations*, **29(1)**, 73-84.

Estes, W.K. (1993). Mathematical models in psychology. In G. Keren and C. Lewis (eds), *A handbook for data analysis in the behavioural sciences: methodological issues*. Hillsdale, NJ: Erlbaum Associates.

Everitt, B.S. (1993). *Cluster analysis* (3rd ed). London: Edward Arnold.

Eysenck, H.J. (1957). *Sense and nonsense in psychology*. Harmondsworth: Penguin.

Eysenck, H.J. (1970). *The structure of human personality* (3rd ed). London: Methuen.

Eysenck, H.J. (1973). *The measurement of intelligence*. Lancaster: Medical and Technical.

Eysenck, H.J. and Eysenck, S.B.G. (1970). *Manual of the Eysenck personality questionnaire*. London: Hodder and Stoughton.

Farrington, D.P. (1973). Self-reports of deviant behaviour: predictive and stable? *Journal of Criminal Law and Criminology*, **64**, 99-110.

Farrington, D.P. (1994). Human development and criminal careers. In M. Maguire, R. Morgan and R. Reiner (eds), *The Oxford handbook of criminology*. Oxford: Clarendon Press.

Farrington, D.P., Snyder, H.N. and Finnegan, T.A. (1988). Specialisation in juvenile court careers. *Criminology*, **26(3)**, 461-483.

Fisher, R.P., Geiselman, R.E., Raymond, D.S., Jurkevich, L.M. and Warhaftig, M.L., (1987). Enhancing enhanced eyewitness memory: Refining the cognitive interview. *Journal of Police Science and Administration*, **15**, 177-185.

Foa, U. (1958). The contiguity principle in the structure of interpersonal relations. *Human Relations*, **11**, 229-238.

Foa, U. (1961). Convergences in the analysis of the structure of interpersonal behaviour. *Psychological Review*, **68**, 341-353.

Foa, U. (1965). New developments in facet design and analysis. *Psychological Review*, **72**, 262-274.

Friday, N. (1975). *My secret garden.* London: Virago.

Ghiselli, E.E., Campbell, J.P. and Zedeck, S. (1981). *Measurement theory for the behavioural sciences*. San Francisco: Freeman.

Gibbons, J.D. (1993). *Nonparametric measures of association.* Newbury Park, CA: Sage.

Godwin, G.M. and Canter, D.V. (1997). Encounter and death: the spatial behaviour of US serial killers. *Policing: An International Journal of Police Strategy and Management*, **20(1)**, 24-38.

Gordon, A.D. (1981). *Classification: Methods for the exploratory analysis of multivariate data.* London: New York.

Gorusch, R.L. (1988). Exploratory factor analysis. In J.R. Nesselroade and R.B. Cattell (eds), *Handbook of multivariate experimental psychology* (2nd Edition). New York: Plenum.

Gower, J.C. (1985). Measures of similarity, dissimilarity and distance. In S. Kotz, N.L. Johnson and C.B. Read (eds), *Encyclopedia of statistical sciences Volume 5*. New York: Wiley.

Graef, J. and Spence, I. (1979). Using distance information in the design of large multidimensional scaling experiments. *Psychological Bulletin,* **86(1)**, 60-66.

Gratch, H. (1973). *25 years of social research in Israel.* Jerusalem: Jerusalem Academic Press.

Green, D.E. and Walkey, F.H. (1980). A nonmetric analysis of Eysenck's personality inventory. *Multivariate Behavioral Research*, **2**, 157-163.

Green, P.E. and Carmone, F.J. (1970). *Multidimensional scaling and related techniques in marketing analysis.* Boston: Allyn and Bacon.

Groth, N.A. (1979). *Men who rape: The psychology of the offender.* New York: Plenum.

Guttman, L. (1941). The quantification of a class of attributes: A theory and methods of scale construction. In P. Horst (Ed.), *The prediction of personal adjustment.* New York: Social Science Research Council.

Guttman, L. (1950a). The basis for scalogram analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, and J.A. Clausen (eds), *The American soldier volume IV: Measurement and prediction.* Reprinted (1966) by New York: Wiley.

Guttman, L. (1950b). The relation of scalogram analysis to other techniques. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, and J.A. Clausen (eds), *The American soldier volume IV: Measurement and prediction.* Reprinted (1966) by New York: Wiley.

Guttman, L. (1954a). An outline of some new methodology for social research. *Public Opinion Quarterly*, **18**, 395-404.

Guttman, L. (1954b). A new approach to factor analysis: The radex. In P.F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, IL: Free Press.

Guttman, L. (1956/1967). What lies ahead for factor analysis. Reprinted as: A psychological design for a theory of mental abilities. In D.N. Jackson and S. Messick (eds), *Problems in Human Assessment.* New York: McGraw-Hill.

Guttman, L. (1960). Personal history of the development of scale analysis. Reprinted in S. Levy (Ed.), *Louis Guttman on theory and methodology*. Aldershot: Dartmouth.

Guttman, L. (1964). The structure of interrelations among intelligence tests, Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Services.

Guttman, L. (1966). Order analysis of correlation matrices. In R.B. Cattell (Ed.), *Handbook of Multivariate Experimental Psychology*. Chicago: Rand McNally.

Guttman, L. (1967b). The development of nonmetric space analysis: A letter to Professor John Ross. *Multivariate Behavioral Research*, **2(1)**, 71-82.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, **33(4)**, 469-506.

Guttman, L. (1971). Measurement as Structural Theory. *Psychometrika,* **36(4)**, 329-347.

Guttman, L. (1977). [cited by some sources as 1978] What is not what in statistics. *The Statistician*, **26**, 81-107.

Guttman, L. (1979). Answers to questions submitted in advance for the workshop on Facet Theory. University of Surrey, March 21 1979.

Guttman, L. (1981). What is not what in theory construction. In I. Borg (Ed.), *Multidimensional data representations: When and why*. Ann Arbor, MI: Mathesis.

Guttman, L. (1982a). What is not what in theory construction. Reprinted in S. Levy (Ed.), *Louis Guttman on theory and methodology: Selected writings*. Aldershot: Dartmouth.

Guttman, L. (1982b). Facet Theory, Smallest Space Analysis and Factor Analysis. *Perceptual and Motor Skills*, **54**, 491-493.

Guttman, L. (1985). Multidimensional Structuple Analysis (MSA-I). for the Classification of Cetacea: Whales, Porpoises and Dolphins. In J.-F. Marchotorchino, J.-M. Proth and J. Janssen (eds), *Data Analysis in Real-Life Environment: Ins and Outs of Solving Problems*. Amsterdam: Elsevier.

Guttman, L. (1986). Polytonicity and monotonicity, coefficients of. In Kotz-Johnson (Ed.), *Encyclopaedia of Statistical Sciences Volume 7*. New York: Wiley.

Guttman, L. and Guttman, R. (1976). A theory of behavioural generality and specificity during mild stress. *Behavioural Science*, **21**, 469-477.

Guttman, L. and Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence*, **15**, 79-103.

Hakstian, A.R. and Muller, V.J. (1973). Some notes on the number of factors problem. *Multivariate Behavioral Research*, **8(4)**, 461-475.

Hazelwood, R.R. and Douglas, J.E. (1980). The lust murderer. *FBI Law Enforcement Bulletin*, **49(4)**, 18-22.

Hazelwood, R.R. and Douglas, J.E. (1980). The lust murderer. Reprinted in NCAVC *Criminal investigative analysis: Sexual homicide*. Quantico, VA: FBI.

Hearnshaw, L. (1979). *Cyril Burt: Psychologist*. London: Hodder and Stoughton.

Heise, D.R. (1968). Norms and individual patterns in student deviancy. *Social Problems*, **16**, 78-91.

Heritage, R. (1992). *Facets of sexual assault: First steps in investigative classifications*. Unpublished MPhil thesis, Department of Psychology, University of Surrey.

Hindelang, (1971). Age, sex and the versatility of delinquent involvements. *Social Problems*, **18**, 522-535.

Hollin, C. (1989). *Psychology and crime: An introduction to criminological psychology*. London: Routledge.

Hough, M. and Mayhew, P. (1983). *The British crime survey: First report*. London: HMSO.

Howarth, E. (1976). A psychometric investigation of Eysenck's Personality Inventory. *Journal of Personality Assessment*, **40**, 173-185.

Howitt, D. (1995). *Paedophiles and sexual offences against children*. Chichester: Wiley.

Jacoby, W.G. (1991). *Data theory and dimensional analysis*. Newbury Park, CA: Sage.

Jones, T., MacLean, B. and Young, J. (1986). *The Islington crime survey: Crime, victimisation and policing in inner-city London*. Aldershot: Gower.

Jupp, V. (1989). *Methods of criminological research.* London: Unwin Hyman.

Kendall, D. (1971). Maps from marriages: An application of non-metric multidimensional scaling to parish register data. Reprinted (1982) in A.P.M. Coxon and C. Jones (eds), *Key texts in multidimensional scaling.* London: Heinemann.

Kinnear, P.R. and Gray, C.D. (1997). *SPSS for Windows made simple.* Hove, Sussex: Psychology Press.

Kirby, S. (1993). *The child molester: Separating myth from reality.* Unpublished PhD thesis, Department of Psychology, University of Surrey.

Klahr, D. (1969). A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. *Psychometrika*, **34(3)**, 319-330.

Klein, M.W. (1984). Offence specialisation and versatility among juveniles. *British Journal of Criminology*, **24(2)**, 185-194.

Kline, P. (1993). *Personality: The pyschometric view.* London: Routledge.

Kline, P. (1994). *An easy guide to factor analysis.* London: Routledge.

Kline, P. and Barrett, P. (1994). Studies with the PPQ and the Five Factor Model of Personality, *European Review of Applied Psychology*, **44(1)**, 35-42.

Krippendorf, K. (1980). *Content analysis: An introduction to its methodology.* London: Sage.

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29(1)**, 1-27.

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29(2)**, 115-129.

Kruskal, J.B. and Carroll, J.D. (1969). Geometrical models and badness-of-fit functions. In P.R. Krishnaiah (Ed.) *Multivariate analysis II.* New York: Academic Press.

Kruskal, J.B. and Wish, M. (1978). *Multidimensional scaling.* Beverley Hills: Sage.

Kruskal, J.B., Young, F.W. and Seery, J.B. (1973). *How to use KYST.* Murray Hill: Bell Laboratories.

Lander, B. (1954). *Towards an understanding of juvenile delinquency.* New York: Columbia University Press.

Lea, J. and Young, J. (1984). *What is to be done about law and order?* Harmondsworth: Penguin.

Leary, T. (1957). *Interpersonal diagnosis of personality.* New York: Ronald Press.

Lee, H.B. and Comrey, A.L. (1979). Distortions in a commonly used factor analytical procedure. *Multivariate Behavioral Research,* **14** (3), 301-321.

Lee, J. (in press). FBI homicide profiling and FBI homicide classification - the motivational link. In D.V. Canter and L.J. Alison (eds) *Profiling crimes: Past, present and future.* Aldershot: Dartmouth press.

Lee, J. and Canter, D.V. (submitted). Dimensional slurring: Starting dimensionality and starting configuration in the Smallest Space Analysis of Meaningful Data.

Levy, S. (1985). Lawful roles of facets in social theories. In D.V. Canter (Ed.), *Facet theory: Approaches to social research.* New York: Springer-Verlag.

Levy, S. and Guttman, L. (1975). On the multivariate structure of wellbeing. *Social Indicators Research,* **2**, 361-388.

Levy, S. and Guttman, L. (1985). The partial order of severity of thyroid cancer with the prognosis of survival. In J.-F. Marchotorchino, J.-M. Proth and J. Janssen (eds), *Data analysis in real life environment: Ins and outs of solving problems.* Amsterdam: Elsevier.

Levelt, W., van de Geer, J.P. and Plomp, R. (1966). Triadic comparisons of musical intervals. *British Journal of Mathematical and Statistical Psychology,* **19**, 163-179.

Liebetrau, A.M. (1983). *Measures of association.* Newbury Park, CA: Sage.

Lingoes, J.C. (1965). An IBM 7090 program for Guttman-Lingoes Smallest Space Analysis - I. *Behavioural Science,* **10**, 183-184.

Lingoes, J.C. (1968a). The multivariate analysis of qualitative data. *Multivariate Behavioral Research,* **3**, 61-94.

Lingoes, J.C. (1968b). The rationale of the Guttman-Lingoes series of programs. *Multivariate Behavioral Research,* **3**, 495-507.

Lingoes, J.C. (1971). Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika,* **36**, 333-334.

Lingoes, J.C. (1972). A general survey of the Guttman-Lingoes nonmetric program series. In R.N. Shepard, A.K. Romney and S.B. Nerlove (eds), *Multidimensional scaling: Theory and applications in the behavioural sciences*. London: Seminar Press.

Lingoes, J.C. (1973). *The Guttman-Lingoes nonmetric program series.* Ann Arbor, MI: Mathesis.

Lingoes, J.C. (1977a). Measurement. In J.C. Lingoes (Ed.), *Geometric representations of relational data: Readings in multidimensional scaling*. Ann Arbor, MI: Mathesis.

Lingoes, J.C. (1977b). Identifying regions in the space for interpretation. In J.C. Lingoes (Ed.), *Geometric representations of relational data: Readings in multidimensional scaling*. Ann Arbor, MI: Mathesis.

Lingoes, J.C. (1977c). Additional programs in the Guttman-Lingoes series. In J.C. Lingoes (Ed.), *Geometric representations of relational data: Readings in multidimensional scaling*. Ann Arbor, MI: Mathesis.

Lingoes, J.C. (1981). Testing regional hypotheses in multi-dimensional scaling. In I. Borg. (Ed.), *Multidimensional data representations: When and why*. Ann Arbor: Mathesis.

Lingoes, J.C. and Borg, I. (1977). Identifying spatial manifolds for interpretation. In J.C. Lingoes (Ed.), *Geometric representations of relational data: Readings in multidimensional scaling*. Ann Arbor, MI· Mathesis.

Lingoes, J.C and Roskam, E.E. (1973). A mathematical and empirical analysis of two multi-dimensional scaling algorithms. *Psychometrika*, **38**, monograph supplement 19.

Lipsey, M.W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Loeber, R. and Waller, D. (1988). Artefacts in delinquency specialisation and generalisation studies. *British Journal of Criminology*, **28(4)**, 461- 477.

Lyons, J., Hirschberg, N. and Wilkinson, L. (1980). The radex structure of the Leary interpersonal behavior circle. *Multivariate Behavioral Research*, **15**, 249-257.

MacCallum, R. (1988). Multidimensional scaling. In J.R. Nesselroade and R.B. Cattell (eds), *Handbook of multivariate experimental psychology (2nd ed.)*. New York: Plenum.

McCann, J.T. (1992). Criminal personality profiling in the investigation of violent crime: Recent advances and future directions. *Behavioral Sciences and the Law*, **10**, 475-481.

Macdonald, R.R. (1997). On statistical testing in psychology. *British Journal of Psychology*, **88**, 333-347.

Maguire, M. (1994). Crime statistics, patterns and trends. In M. Maguire, R. Morgan and R. Reiner (eds), *The Oxford handbook of criminology*. Oxford: Clarendon Press.

Maimon, Z. (1978). The choice of ordinal measures of association. *Quality and Quantity*, **12**, 255-264.

Maraun, M. (1997). Appearance and reality: Is the big-five the structure of trait descriptors? *Journal of Personality and Individual Differences*, **22(5)**, 629-647.

Marsden, P.V. and Laumann, E.O. (1978). The social structure of religious groups: A replication and methodological critique. In S. Shye (Ed.), *Theory construction and data analysis in the behavioural sciences*. San Francisco: Jossey-Bass.

Marshall, W.L. (1989). Intimacy, loneliness and sexual offenders. *Behavioural research in therapy*, **27**, 491-503.

Megargee, E.I. (1982). Psychological determinants and correlates of criminal violence, in M.E. Wolfgang and N.A. Wiener (eds), *Criminal violence*. London: Sage.

Merschrod, K. (1980). Partial Order Scalogram Analysis: A technique for scaling qualitative data in two dimensions. *Quality and quantity*, **14**, 635-650.

Miles, M.B. and Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed). Newbury Park, CA: Sage.

Minium, E.W., King, B.M. and Bear, G. (1993). *Statistical reasoning in psychology and education* (3rd ed). Chichester: Wiley.

Mirrlees-Black, C., Mayhew, P. and Percy, A. (1997). The 1996 British Crime Survey. *Home Office Statistical Bulletin*, **19/96**.

Nutch, F.J. and Bloombaum, M. (1968). A Smallest Space Analysis of gang boys' behaviors. *Pacific Sociological Review*, **11**, 116-123.

Nye, F.I. and Short, J.F. (1957). Scaling delinquent behavior. *American Sociological Review*, **22**, 327-331.

O'Conneell-Davidson, J. and Layder, D. (1994). *Methods, sex and madness*. London: Routledge.

Osgood, D.W., Johnston, L.D., O'Malley, P.M. and Bachman, J.G. (1988). The generality of deviance in late adolescence and early adulthood. *American Sociological Review*, **53**, 81-93.

Paddock, J.R. and Nowicki, S. (1986). The circumplexity of Leary's interpersonal circle: A multidimensional scaling perspective. *Journal of Personality Assessment*, **50(2)**, 279-289.

Peterson, R.A., Pittman, D.J. and O'Neal, P. (1962). Stabilities in deviance: A study of assaultive and non-assaultive offender. *Journal of Criminal Law and Criminology*, **53**, 44-48.

Pinizzotto, A.J. (1984). Forensic psychology: Criminal personality profiling. *Journal of Police Science and Administration*, **12(1)**, 32-40.

Pinizzotto, A.J. and Finkel, N.J. (1990). Criminal personality profiling: An outcome and process study. *Law and Human Behavior*, **14(3)**, 215-232.

Poland, J.M. (1983). Statistical characteristics of criminal justice data. *Journal of Police Science and Administration*, **11(3)**, 282-289.

Prentky, R.A., Burgess, A.W., Rokous, F., Lee, A., Hartman, C., Ressler, R. and Douglas, J. (1989). The presumptive role of fantasy in serial sexual homicide. *American Journal of Psychiatry*, **146(7)**, 887-891.

Quay, H.C. and Blumen, L. (1963). Dimensions of delinquent behavior. *Journal of Social Psychology*, **61**, 273-277.

Ramsay, J.O. (1982). Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society*, **145**, 285-312.

Raveh, A. (1986). On measures of monotone association. *The American Statistician*, **40(2)**, 117-123.

Raveh, A. and Landau, S.F. (1993). Partial Order Scalogram Analysis with Base Coordinates (POSAC): Its Application to Crime Patterns in All the States in the United States. *Journal of Quantitative Criminology*, **9(1)**, 83-99.

Reber, A.S. (1985). *The Penguin dictionary of psychology*. London: Penguin.

Ressler, R.K. (1985a). The men who murdered. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Ressler, R.K. (1985b). The split reality of murder. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Ressler, R.K. (1985c). Classifying sexual homicide crime scene: Interrater reliability. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Ressler, R.K. (1985d). Crime scene and profile characteristics of organised and disorganised murderers. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Ressler, R.K. and Burgess, A.W. (1985).

Ressler, R.K., Burgess, A.W., Douglas, J.E., Hartman, C.R. and D'Agostino, R.B. (1986b). Sexual killers and their victims: Identifying patterns through crime scene analysis. *Journal of Interpersonal Violence*, **1(3)**, 288-308.

Ressler, R.K., Burgess, A.W., Hartman, C.R., Douglas, J.E., and McCormack, A. (1986a). Murderers who rape and mutilate. *Journal of Interpersonal Violence*, **1(3)**, 273-287.

Ressler, R.K., Douglas, J.E. and Burgess, A.W. (1983). Rape and rape-murder: One offender and twelve victims. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Ressler, R.K., Douglas, J.E., Burgess, A.W. and Burgess, A.G. (1992). *Crime classification manual.* London: Simon and Schuster.

Ressler, R.K., Douglas, J.E., Groth, A.N. and Burgess, A.W. (1980). Offender profiles: A multidisciplinary approach. Reprinted in NCAVC *Criminal Investigative Analysis: Sexual Homicide*. Quantico, VA: FBI.

Richardson, M.W. (1938). Multidimensional psychophysics. *Psychological Bulletin*, **35**, 659-660.

Roskam, E.E. (1977). A survey of the Michigan-Israel-Netherlands-Integrated series. In J.C. Lingoes (Ed.), *Geometric representations of relational data: Readings in multidimensional scaling.* Ann Arbor, MI: Mathesis.

Roskam, E.E. (1979/1981). Contributions of multidimensional scaling in social science research. In I. Borg (Ed.), *Multidimensional representations: When and why*. Ann Arbor: Mathesis.

Roskam, E.E. (1981). Data theory and scaling: A methodological essay on the role and use of data theory and scaling in psychology. In I. Borg (Ed.), *Multidimensional representations: When and why*. Ann Arbor: Mathesis.

Roskam, E.E. and Lingoes, J.C. (1970). MINISSA-I: A FORTRAN IV (G) Program for the smallest space analysis of square symmetric matrices. *Behavioral Science*, **15**, 204-205.

Roskam, E.E. and Lingoes, J.C. (1981). MINISSA. In S.S. Schiffman, M.L. Reynolds and F.W. Young (eds), *Introduction to multidimensional scaling: Theory, methods and applications*. Orlando: Academic Press.

Schiffman, S.S., Reynolds, M.L. and Young, F.W. (1981). *Introduction to multidimensional scaling: Theory, methods and applications*. Orlando: Academic Press.

Schlesinger, I.M. and Guttman, L. (1969). Smallest Space Analysis of intelligence and achievement tests. *Psychological Bulletin*, **71**, 95-100.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, **1(2)**.

Scott, J.F. (1959). Two dimensions of delinquent behavior. *American Sociological Review*, **24**, 240-243.

Scott, J.P. (1991). *Social network analysis: A handbook*. Sage: London.

Sellitz, C., Jahoda, M., Deutsch, M. and Cook, S.W. (1959). Attitude scaling. Reprinted in M. Jahoda and N. Warren (eds), *Attitudes*. Harmondsworth: Penguin.

Shavelson, R.J. (1988). *Statistical reasoning for the behavioural sciences*. Boston: Allyn and Bacon.

Shepard, R.N. (1962). The Analysis of proximities: Multi-dimensional scaling with an unknown distance function. I. *Psychometrika*, **27(2)**, 125-140.

Shepard, R.N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, **3**, 287-315.

Shepard, R.N. (1972). Introduction, in R.N. Shepard, A.K. Romney and S.B. Nerlove (eds), *Multi-dimensional scaling: Theory and applications in the behavioural sciences*. London: Seminar Press.

Shepard, R.N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, **39(4)**, 373-421.

Shepard, R.N. (1978). The circumplex and related topological manifolds in the study of perception. In S. Shye (Ed.), *Theory Construction and Data Analysis in the Behavioural Sciences*. San Francisco: Jossey-Bass.

Shye, S. (1978a). On the search for laws in the behavioural sciences. In S. Shye (Ed.), *Theory construction and data analysis in the behavioural sciences*. San Francisco: Jossey-Bass.

Shye, S. (1978b). Partial Order Scalogram Analysis. In S. Shye (Ed.), *Theory construction and data analysis in the behavioural sciences*. San Francisco: Jossey-Bass.

Shye, S. (1978c). Glossary. In S. Shye (Ed.), *Theory construction and data analysis in the behavioural sciences*. San Francisco: Jossey-Bass.

Shye, S. (1978d). Facet analysis and regional hypotheses. In S. Shye (Ed.), *Theory construction and data analysis in the behavioural sciences*. San Francisco: Jossey-Bass.

Shye, S. (1985a). *Multiple scaling: The theory and application of Partial Order Scalogram Analysis*. Amsterdam: North-Holland.

Shye, S. (1985b). Nonmetric multivariate models for behavioural actions systems. In D.V. Canter (Ed.), *Facet theory: Approaches to social research*. New York: Springer-Verlag.

Shye, S. (1991). *Faceted Smallest Space Analysis (FSSA) version 3.01*. Jerusalem: Israel Institute of Applied Social Research.

Shye, S. (1997). The developments and prospects of Facet Theory. Paper presented to Sixth International Facet Theory Conference, Liverpool.

Shye, S. and Elizur, D. (1976). Worries about deprivation of job rewards following computerization: A Partial Order Scalogram Analysis. *Human Relations*, **29(1)**, 63-71.

Shye, S. and Elizur, D. (1994). *An introduction to Facet Theory*. Newbury Park, CA: Sage.

Shye, S., Magen, A. and Goldzweig, G. (1997). POSAC for structural theory: New program for intrinsic multidimensional data analysis. Paper presented to Sixth International Facet Theory Conference, Liverpool.

Smith, W.R., Smith, D.R. and Noma, E. (1986). The multidimensionality of crime: A comparison of techniques for scaling delinquent careers. *Journal of Quantitative Criminology*, **4(2)**, 329-353.

Snijders, T.A.B., Dormaar, M., van Schur, W.H., Dijkman-Caes, C. and Driessen, G. (1990). Distribution of some similarity coefficients for dyadic binary data in the case of associated attributes. *Journal of Classification*, **7**, 5-31.

Sparks, R., Genn, H. and Dodd, D. (1977). *Surveying victims*. Chichester: Wiley.

Spence, I. (1972). A Monte Carlo evaluation of three nonmetric multidimensional scaling algorithms. *Psychometrika*, **37(4)**, 461-486.

Spence, I. (1974). On random rankings studies in nonmetric scaling. *Psychometrika*, **39(2)**, 267-268.

Spence, I. and Graef, J. (1974). The determination of the underlying dimensionality of an empirically obtained matrix of proximities. *Multivariate Behavioral Research*, **9(3)**, 331-341.

Spence, I. and Young, F.W. (1978). Monte Carlo studies in nonmetric scaling. *Psychometrika*, **43(1)**, 115-117.

Sphinx (1996). Cocaine. University of Liverpool Student Magazine, **Autumn**, 9-11.

Stenson, H.H. and Knoll, R.L. (1969). Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychological Bulletin*, **71(2)**, 122-126.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, **103(2684)**, 677-680.

Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.), *The handbook of experimental psychology*. London: Chapman and Hall.

Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarsfeld, P.F. , Star, S.A. and Clausen, J.A. (1950). *The American soldier volume iv: Measurement and prediction.* Reprinted (1966). by New York: Wiley.

Suchman, E.A. (1950a). The scalogram board technique for scale analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, and J.A. Clausen (eds), *The American soldier volume iv: Measurement and prediction.* Reprinted (1966). by New York: Wiley.

Suchman, E.A. (1950b). The intensity component. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, and J.A. Clausen (eds), *The American soldier volume iv: Measurement and prediction.* Reprinted (1966). by New York: Wiley.

Takane, Y., Young, F.W. and de Leeuw, J. (1977). Nonmetric individual differences multi-dimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7-67.

Thomson, G.H. (1948). *The factorial analysis of human ability* (3rd ed). London: University Press.

Thurstone, L.L. (1929). Theory of attitude measurement. *Psychological Bulletin*, **36**, 222-241.

Thurstone, L.L. (1935). *Vectors of the mind.* Chicago: University Press.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Trapnell, P.D. and Wiggins, J.S. (1990). Extension of the Interpersonal Adjective Scales to include the Big Five dimensions of personality. *Journal of Personality and Social Psychology*, **59(4)**, 781-790.

Velicer, W.F. and Jackson, D.N. (1990). Components analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, **25(1)**, 1-28.

Velleman, P.F. and Wilkinson, L. (1993). Nominal, ordinal, interval and ratio typologies are misleading. *The American Statistician*, **41(1)**, 65-72.

Von Eye, A. (1988). Some multivariate developments in nonparametric statistics. In J.R. Nesselroade and R.B. Cattell (eds), *Handbook of multivariate experimental psychology* (2nd Edition). New York: Plenum.

Wagenaar, W.A. and Padmos, P. (1971). Quantitative interpretations of stress in Kruskal's multidimensional scaling technique. *British Journal of Mathematical and Statistical Psychology*, **24**, 101-110.

Walkey, F.H. and Green, D.E. (1981). The structure of the Eysenck Personality Inventory: A comparison between simple and more complex analyses of a multiple scale questionnaire. *Multivariate Behavioral Research*, **16**, 361-372.

Weber, R.P. (1990). *Basic content analysis (2nd ed)*. Newbury Park, CA: Sage.

Weller, S.C. and Romney, A.K. (1988). *Systematic data collection.* Newbury Park, CA: Sage.

Weller, S.C. and Romney, A.K. (1990). *Metric scaling: Correspondence analysis.* Newbury Park, CA: Sage.

Wiggins, J.S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, **37(3)**, 395-412.

Wiggins, J.S. (1980). Circumplex models of interpersonal behaviour. In L. Wheeler (Ed.), *Review of personality and social psychology*. London: Sage.

Wiggins, J.S., Steiger, J.H. and Gaelick, L. (1981). Evaluating circumplexity in personality data. *Multivariate Behavioral Research*, **16**, 263-289.

Wiggins, J.S., Trapnell P. and Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research*, **23**, 517-530.

Wiggins, J.S., Phillips, N. and Trapnell P. (1989). Circular reasoning about interpersonal behavior: Evidence concerning some untested assumptions underlying diagnostic classification. *Journal of Personality and Social Psychology*, **56(2)**, 296-305.

Wilson, M.A. (1995). Structuring qualitative data: Multidimensional scalogram analysis. In G.M. Breakwell, S. Hammond and C. Fife-Schaw (eds), *Research Methods in Psychology*. London: Sage.

Wolfgang, M.E., Figlio, R.M. and Sellin, T. (1972). *Delinquency in a birth cohort.* Chicago: University Press.

Young, F. W. (1987). History. In F. W. Young and R.M. Hamer (eds), *Multidimensional scaling: History, theory and applications*. Hillsdale, NJ: LEA.

Young, F. W. and Hamer, R.M. (1987). *Multidimensional scaling: History, theory and applications*. Hillsdale, NJ: LEA.

Young, F.W. (1970). Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, **35**, 455-473.

Young, F.W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, **46**, 357-388.

Young, F.W. (1997). Multidimensional scaling. World Wide Web site accessed 5th May 1997. http://forrest.psych.unc.edu/teaching/p208a/p208tc.html

Young, G. and Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, **3**, 19-22.

Youngs, D.E. (1994). [Self-reported delinquency and criminal behaviour in young British males]. Unpublished raw data.

Zvulun, E. (1978). Multidimensional scalogram analysis: The method and its application. In S. Shye (Ed.), *Theory construction and data analysis in the behavioural sciences*. San Francisco: Jossey-Bass.