



This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

A Comparison of Two Measures of Attitude:  
Scalogram Analysis and the Semantic Differential Test

by

Diane Stapleton Cooney

Thesis submitted for the degree of Doctor of Philosophy.

University of Keele

July 1969

TABLE OF CONTENTS

	page
Acknowledgements	iv
Abstract	v
List of Tables	vi
Classification of Inserts	viii

I. BASIC CONCEPTS

1. Introduction	1
2. Definition of Attitude	3
3. Measurement of Attitude	5
4. Definition of Measurement	5
5. Methods of Collecting Data	8
6. Scaling Attitudes	11
7. Some Attitude Scales	12
8. The Relationship of Attitudes to Behaviour	15

II. SCALOGRAM ANALYSIS

1. Guttman's Definition of a Scale	18
2. Cumulative Scales	20
3. Methods of Scale Analysis	22
4. The Universe of Attributes	33
5. Statistical Test of 'good fit' for Scalogram Data	37
6. Further Components of Scales	40
7. Criticisms of Scale Analysis	41
8. Sampling the Universe of Content	49
9. Unidimensionality	50
10. Evaluation of Scale Analysis	56

### III. THE SEMANTIC DIFFERENTIAL

1. Description of the Semantic Differential	58
2. Osgood's Learning Theory	63
3. Definition of Meaning	66
4. The Dimensionality of Semantic Space	68
5. Osgood's Evaluation of the Semantic Differential	76

### IV. THE DIMENSIONALITY OF DATA

1. Levels of Measurement used by the Semantic Differential and Scale Analysis	81
2. Metric Assumptions of the Semantic Differential	85

### V. THE SEMANTIC DIFFERENTIAL AS AN ATTITUDE SCALE

1. Attitude as a Component of Meaning	90
2. The Measurement of Attitudes with the Semantic Differential	93
3. Relationship to other Attitude Scales	96

### VI. THE SCALE ANALYSIS OF SEMANTIC DIFFERENTIAL DATA

1. Investigation of the Metric Properties of the Semantic Differential	102
2. Scale Analysis of Semantic Differential Data	107
3. Scale Analysis of Semantic Differential Pilot Study Data	108
4. The Relationship of Scale Analysis to Factor Analysis	111

VII. EMPIRICAL COMPARISON OF GUTTMAN SCALE AND SEMANTIC DIFFERENTIAL

1. Construction of Guttman Scale	114
2. Construction of Semantic Differential	122
3. Results of Pilot Studies	127
4. Empirical Comparison of the Two Techniques	128
5. The Extent to which the Semantic Differential measures Attitude	135
6. Correlation with Other Techniques	140

VIII. THE STRUCTURE OF SEMANTIC DIFFERENTIAL DATA

1. The Evaluative Factor : Some Qualifications	144
--	-----

IX. DISCUSSION AND CONCLUSIONS 150

BIBLIOGRAPHY 155

APPENDICES

1. Least Squares Method of Scale Analysis	163
2. D-method of Factor Analysis	172
3. Questionnaires:	
i. Pilot Study A	176
ii. Pilot Study B	182
iii. Main Study	188

ACKNOWLEDGEMENTS

Thanks are due to the Sir Halley Stewart Trust for financing the Study of Social Attitudes, from which the data are drawn, and to Mr. R. Mapes, of the University of Keele, for supervising this study. Also, I would like to thank Dr. P. Hetherington, of the University of London, who spent many hours assisting with the preparation of tables and graphs.

- v -

ABSTRACT

The supposed identity of the evaluative factor of the semantic differential with another measure of attitude, a Guttman scale, is discussed and empirically investigated. Theoretical considerations suggest that the degree of association observed may vary, according to the operation of a number of factors relating to the content and form of constructed scales. An empirical investigation, based on the responses of 1008 subjects, to the concept OBEDIENCE, and to a comparable Guttman scale, is reported. The degree of association found,  $C = 0.41$ , did not permit the postulation of a relationship of near identity between the two measures.

LIST OF TABLES

TABLE NUMBER	INSERT	page
1	Table I	17(a)
2	Methodological Note I	19(a)
3	Example I	22(a)
4	Methodological Note II	23(a)
5	Example II	30(a)
6	Example III	32(a)
7	Methodological Note III	35(a)
8	Example IV	37(a)
9	Methodological Note IV	49(a)
10	Methodological Note V	51(a)
11	Example V	53(a)
12	Methodological Note VI	55(a)
13	Example VI	62(a)
14	Methodological Note VII	65(a)
15	Methodological Note VIII	81(a)
16	Example VII	89(a)
17	Example VIII	107(a)
18	Factor Analysis I	108(a)
19	Example IX	108(b)
20	Scalogram I	109(a)
21	Scalogram II	109(b)

TABLE NUMBER	INSERT	page
22	Factor Analysis II	110(a)
23	Table of distribution of sample variables I	117(a)
24	Scalogram III	121(a)
25	Scalogram IV	121(b)
26	Table II	126(a)
27	Table of distribution of sample variables II	129(a)
28	Graphs I and II	130(a)
29	Scalogram V	131(a)
30	Table of frequency distribution of response I	131(b)
31	Scalogram VI	132(a)
32	Table of frequency distribution of response II	132(b)
33	Table of frequency distribution of response III	133(a)
34	Table of frequency distribution of response IV	133(b)
35	Graph III	134(a)
36	Graph IV	134(b)
37	Factor Analysis III	147(a)

CLASSIFICATION OF INSERTS

TYPE OF INSERT	page
Examples	ix
Factor Analyses	x
Graphs	xi
Methodological Notes	xii
Scalograms	xiii
Tables	xiv
Tables of distribution of sample variables	xv
Tables of frequency distribution of response	xvi

All inserts are labelled TABLE and each is accorded a table number. This is so the reader will not confuse them with the text. For the same reason the pages of inserts are not numbered as part of the text, but are distinguished by the use of lower case letters as suffixes. For example, page 17(a) follows page 17 in the text. Inserts can be referred to as mentioned in the text, or read independently, but the reader is advised, in the first instance to follow the text.

Though this arrangement is in some respects cumbersome, its aim is to facilitate the easy reading of the text. The discussion and illustration of points which could interrupt the flow of the text are confined to inserts. Footnotes, which are usually too brief to do justice to methodological issues, are avoided.

EXAMPLES

	TABLE NO.	page
I The Cumulative Scale	3	22(a)
II The Scalogram Board Technique	5	30(a)
III The Pattern of Errors in Scalogram Analysis	6	32(a)
IV Statistical Test of Significance of Scale Findings	8	37(a)
V Example of a Scale (Hypothetical)	11	53(a)
VI Plotting Concepts in Semantic Space	13	62(a)
VII The Identity of Continua	16	89(a)
VIII Scale Analysis and Factor Analysis (Equal Intervals)	17	107(a)
IX Scale Analysis and Factor Analysis (Common Origin)	19	108(b)

FACTOR ANALYSES

	TABLE NO.	page
I Coordinates of evaluative, potency and activity scales selected from Osgood, based on the responses of 60 Pilot Study B subjects to the concept OBEDIENCE	18	108(a)
II Coordinates of evaluative scales selected from Osgood, based on the responses of 60 Pilot Study B subjects differentiating the concept OBEDIENCE	22	110(a)
III Coordinates of scales, based on the responses of 1008 main sample subjects to the concept OBEDIENCE	37	147(a)

GRAPHS

	TABLE NO.	page
I Age distribution of 1008 subjects completing semantic differential	28	130(a)
II Age Distribution of 1008 subjects failing to complete semantic differential	28	130(a)
III Relationship of Guttman scale and semantic differential responses of 1008 main study subjects based on cumulative frequency of response	35	134(a)
IV Relationship of Guttman scale and semantic differential responses of 1008 main study subjects based on deciles	36	134(b)

METHODOLOGICAL NOTES

	TABLE NO.	page
I The Process of Scaling	2	19(a)
II The Matrix of Data in Scalogram Analysis	4	23(a)
III The Universe of Attributes	7	35(a)
IV Categories of the Scalogram Matrix of Data	9	49(a)
V Trace Lines of Scale Models	10	51(a)
VI The Meaning of Unidimensionality	12	55(a)
VII Paridigm of the Sign Process	14	65(a)
VIII The Matrix of Data in Scalogram Analysis and the Semantic Differential	15	81(a)

SCALOGRAMS

	TABLE NO.	page
I Semantic Differential Evaluative Scales.		
Study B. (N = 60). (i)	20	109(a)
II Semantic Differential Evaluative Scales.		
Study B. (N = 60). (ii)	21	109(b)
III Guttman Scale Items.		
Study A. (N = 92)	24	121(a)
IV Guttman Scale Items.		
Study B. (N = 99)	25	121(b)
V Scale Pattern of Guttman Scale Data.		
(N = 1008)	29	131(a)
VI Scale Pattern of Semantic Differential		
Data. (N = 1008)	31	132(a)

TABLES

	TABLE NO.	page
I    Summary of Studies of Correspondence between Measured Attitude and Behavioural Patterns	1	17(a)
II   Table of Squared Distances between Semantic Differential Profiles of 99 Pilot Study B Subjects on the Concept OBEDIENCE	26	126(a)

TABLES OF DISTRIBUTION OF SAMPLE VARIABLES

	TABLE NO.	page
I Description of Pilot Study Samples		
Sample A; N = 92 : Sample B; N = 99.	23	117(a)
Distribution of		
i. Age		117(a)
ii. Sex		117(a)
iii. Marital Status		117(b)
iv. Education		117(b)
v. Religion		117(c)
vi. Place of Birth		117(c)
II Description of Main Sample.		
N = 1008.	27	129(a)
Distribution of		
i. Age		129(a)
ii. Sex		129(a)
iii. Marital Status		129(b)
iv. Education		129(b)
v. Religion		129(c)
vi. Place of Birth		129(c)
vii. Occupation		129(d)

TABLES OF FREQUENCY DISTRIBUTION OF RESPONSE

	TABLE NO.	page
I    Distribution of responses of 1008 main study subjects on Guttman scale items	30	131(b)
II   Distribution of responses of 1008 main study subjects on semantic differential scale items	32	132(b)
III  Bivariate frequency table of Guttman scale scores by semantic differential evaluative factor scores of 1008 main study subjects (before category combinations)	33	133(a)
IV   Bivariate frequency table of Guttman scale scores by semantic differential evaluative factor scores of 100 main study subjects (after category combination)	34	133(b)

## CHAPTER I

### BASIC CONCEPTS

#### 1. Introduction

The aim of this work is to discuss and compare two techniques which are commonly used to measure attitudes. Scale analysis was developed by Guttman to deal with some of the problems of attitude measurement. The semantic differential was designed by Osgood to measure meaning, but is commonly used as an attitude scale. As the two techniques have developed from different theoretical standpoints, giving emphasis to different problems and procedures, the comparison is not an obvious one, and can be approached in a number of ways. It does seem important to try to make the comparison, as attitude research underlies a number of important disciplines, and a research worker may, at the moment, have to choose one of these techniques rather than the other to measure the same attitude, while having little information about their relationship at his disposal.

Both the semantic differential and the Guttman scale are thought to measure attitude. If they are both valid, and if they both purport to measure the same thing, similar results should be obtained using either method, and only practical considerations need differentiate between them. Osgood et al. (1957) note a high correspondence between the two, in one instance, and use this as evidence that they do measure the same thing. The danger in equating

the two instruments, because they both seem to measure something which can be described as an attitude, will become apparent when some of the problems involved in measuring attitudes are discussed. A single definition of the term "attitude" is by no means universally accepted. Attitudes cannot be directly observed, and in any one application of the two techniques it is not possible to know that they are measuring the same variable. If a high correspondence is found to exist, and this cannot be inferred from only one application, they may be measuring different, but highly correlated, variables. Where a purely empirical approach is acceptable, the techniques could be used interchangeably, if they were found repeatedly to give highly similar results under differing conditions. This empirical approach is often accepted in psychology. For instance, in the field of psychological testing variables are defined operationally and tests are evaluated in terms of their usefulness. But purely empirical criteria are not as easily applied in this case, because the form and content of the measuring techniques under consideration, is not fixed. The psychological test is rigidly structured and standardized; these techniques are not. An unknown, but probably very large, number of other variables may be introduced in different applications of the two techniques. A relationship between the techniques, observed to hold for a large number of instances, might break down when a change of content, or a change in the type of subject occurred. Even if it is accepted that both instruments can be used to measure attitude, it is necessary to define those conditions under which they measure the same attitude, before

they can be used interchangeably. It is necessary to consider the rationale of each technique, to see if the relationship between the two techniques follows theoretically, before accepting that they do measure the same thing.

After defining some of the concepts involved, the rationale of each technique will be described in Chapters II and III. In Chapters IV and V a comparison between the two techniques will be made. In Chapter VI an attempt will be made to transform one type of technique, the semantic differential, into the other, a Guttman scale. In Chapter VII an empirical application of the two techniques will be discussed. Finally a comparative evaluation of the two techniques will be made, and the question of whether they do in fact measure the same thing will be considered.

## 2. Definition of Attitude

The term attitude is derived from the Latin "aptus", which means fitness or adaptedness. Like its byeform aptitude it connotes a mental state of preparation for action. The mental state may, of course, be conceived as being anchored in, or composed of, changes in neural substrata, and hence basically physical in nature. Osgood (1957) for example defines attitude as a component of meaning in terms of representational mediational responses. But whether the attitude is regarded as essentially mental or physical it is still treated as a psychological variable insofar as it is incipient and unobservable, rather than overt. In the terms of Coombs (1953) it is not behaviour

at the phenotypic or manifest level, and can be studied only through behaviour at this level, at the present time.

In his classic study of attitudes Allport (1935) examines a number of definitions of attitudes from diverse sources, and finds that they have in common this idea of a preparation or readiness for response. He defines attitude as:

"a mental and neural state of readiness, organized through experience, exerting a directive and dynamic influence on the individuals response to all objects and situations with which it is related. (p.810)".

This definition is consistent with the conception of attitude as a construct to explain the consistencies of human behaviour. Eysenck (1954), for example, uses attitudes as part of a hierarchical model of underlying variables to explain consistencies in the behaviour of individual people. Writers primarily interested in the measurement of attitudes tend to use more specific and limited definitions of attitude. Thurstone (1929), for example, defines attitude in terms of feeling or affect for or against an attitude object. This type of definition implies a linear continuum running from positive, through neutral, to negative feelings about the attitude object.

It would be difficult to agree on a correct definition of attitude. In fact, as Allport points out, the situation offers certain similarities with the field of intelligence testing. Intelligence tests have proved their usefulness and validity, while the nature and definition of intelligence are still disputed. Sometimes attitudes are more successfully measured than defined. More precisely, it is sometimes more

fruitful to define attitudes operationally and carry out research based on operational definitions, than to try and define the exact nature of the underlying construct first.

### 3. The Measurement of Attitudes

The most salient feature of attitudes, in relation to their measurement, is that they are not directly observable. This means that it is very difficult to measure them, but it does not follow that the task is impossible. Many constructs used in the physical sciences are not directly observable either; though they tend to be more constant than psychological variables and thus easier to handle. Marshall (1963) expresses the point convincingly:

"What is measured is not a state of mind but its manifestation in certain selected forms of behaviour. Nobody would object to the calculation of a marriage rate on the grounds that love is a passion that defies the measuring rod. Nobody would condemn the study of crime rates because they do not measure criminal propensities. (p.20)."

### 4. Definition of Measurement

Coombs (1953) examines the problems of attitude measurement from the framework of a general theory of data. There are, he suggests, two broad aspects to measurement. On the one hand there is an abstract or formal system of elements with certain properties and operations. On the other hand there is the domain of observable things or objects, with observable properties and relationships. Measurement is the

process of mapping the real object system into one of the abstract formal systems, or mathematical models. In this way operations can be performed in the abstract system to give information about the relations holding in the real object system. The data must fit the model if the operations in the abstract system are to be applicable to the object system.

The simplest form of measurement consists of substituting symbols or names for real objects. When objects are mapped into classes, represented by symbols, the symbols constitute a nominal scale. Certain properties hold between the objects on a nominal scale. A pair of objects either belong to the same class, and are equal, or do not belong to the same class, and are not equal, with respect to the attribute which defines the class. The relation of equality is symmetric and transitive. That is, if  $a$ ,  $b$  and  $c$  are elements then if  $a = b$ ,  $b = a$  and if  $a = b$  and  $b = c$ , then  $a = c$ .

Sometimes a higher relationship than simply "different from" can be found among the objects. It may be possible to say that one class possesses more of an attribute than another. That is, the relation greater than ( $>$ ) may hold. If this relationship holds between some classes a partially ordered scale exists. If this relationship holds between all classes, for all pairs of objects from different classes a simply ordered scale, or ordinal scale exists. The relationship "greater than" is transitive, but not symmetric, and, of course, the axioms of equality pertaining to the lower level of measurement still apply.

If a relationship is observed to hold on the distances between classes, a higher order of measurement is possible. If it is possible

to order classes according to the relation greater than, and also to similarly order the intervals between them, an ordered metric scale exists. If however it is possible to order only some of the intervals, the scale is termed an ordinal-partially ordered scale.

The interval scale is characterized by the fact that the data contain information on the magnitude of the intervals between stimuli on the scale. Objects can be assigned numbers and differences between scale values may be operated on arithmetically. Any linear transformation preserves the relations between intervals. For instance, any constant number can be added to the scale scores as there is no absolute origin. Similarly scale scores may be multiplied by any given value, as the unit of measurement is arbitrary.

The ratio scale has the additional property of possessing an absolute origin. The unit of measurement is arbitrary and relations are preserved under scalar multiplication, but scale values cannot be translated. That is, an arithmetic constant cannot be added to each value to give an equally valid score, because the origin of the scale is fixed. In this system it is possible to express scale values as ratios and say for example, that one object possesses twice as much of the attribute as another.

Each level of measurement is successively more powerful, but a higher level of measurement does not always fit the given data. Measurement may mean, then

"1. mapping an object system into an interval or ratio scale, permitting the assignment of numbers to objects, and permitting at least some of the operations of arithmetic to be performed on these numbers.

2. mapping an object system into at least a simple order, including the ordered metric, interval and ratio scale.

3. a generalization to the extent of mapping an object system into a partial order or even a nominal scale.

4. a generalization including the decomposition of partial orders into sets of simple orders - in effect, measurement in a multi-dimensional space. (p485)".

#### 5. Methods of Collecting Data

The data, from which an attitude scale is constructed, are the subject's check marks. The subject is asked to place check marks against items about the attitude object, to indicate his attitude. It is possible to begin with theories about the structure of a latent variable, or to begin with the data and ask if they satisfy a simple order. The former approach may impose qualities on the data which they do not possess, and hence it may be invalid to generalize from the formal system to the system of objects. It is assumed in attitude measurement that the subject has an attitude, and that his item checking behaviour will be related to that attitude. Two levels are involved in this item checking behaviour. Coombs describes the manifest, observable behaviour, that is the item checking behaviour, as being at the phenotypic level. The inferred, hypothetical or latent level of behaviour which is believed to underly, or generate, this behaviour is described as being at the genotypic level. The degree of attitude that the subject holds towards the attitude

object when he is checking the related items is called, by Coombs, his "ideal". If stimuli or attitude statements are evaluated with respect to an ideal they are classified, by Coombs, as Task A data. For example, if the subject is asked to place a mark against that statement which, of the following, agrees most with his opinion or attitude:

1. No immigrants should be admitted to Britain.
2. A limited number of immigrants should be admitted to Britain.
3. An unlimited number of immigrants should be admitted to Britain.

it is assumed that he will check the statement which comes closest to his attitude or ideal.

A subject may be given a task, however, which specifically excludes his own attitude. He might be asked, for instance, to indicate which of these statements he thinks is most "for" or "against" a policy of unlimited immigration in Britain. Neither the semantic differential, or scale analysis employs this type of behaviour in the collection of data. This type of task, which requires the subject to evaluate the items rather than to indicate his own attitude yields what Coombs calls Task B data.

Another dichotomy is between relative and irrelative behaviour. In relative behaviour the judgement is between two or more alternatives. For instance, the respondent might be asked which of two candidates he prefers, or which of two statements he considers to be more liberal. He may not consider either of the statements to be liberal in an absolute sense, or either of the candidates to be good.

When behaviour is irrelative the judgements involve a single stimulus or attitude statement at a time. The statement may be endorsed or not endorsed (or degrees of agreement may be allowed for). The behaviour in checking either Guttman scale items, or semantic differential items is irrelative. Individuals failing to endorse an item may do so for different reasons. For example, subjects may disagree with statement 2, above, for opposite reasons. Some subjects would disagree on the grounds that there should be no immigration at all, whereas others might disagree on the grounds that immigration should be unlimited. That is different attitudes (genotypic level) may generate the same manifest behaviour (phenotypic level). The difficulties inherent in attempting to gauge an attitude with one item only are clear.

The items in an attitude questionnaire may be non-monotone or monotone stimuli. For example, in a Thurston scale, the subject is asked to select those items which come nearest to his ideal or attitude and to reject all items beyond these, regardless of their direction. The items then are non-monotone stimuli. In a perfect Guttman scale the endorsement of one item implies the endorsement of all less intense items, and a cutting point can be established, below which all items are endorsed, and above which all items are rejected. The stimuli are monotone. In evaluating semantic differential concepts against scales, non-monotone stimuli are involved.

## 6. Scaling Attitudes

Coombs draws a simple distinction between scaling and measuring.

"The theory of the ordered metric and less powerful scales may be referred to as scaling theory, and the theory of the interval and ratio scales as measurement theory. (p.484)".

Scaling then consists in the ordering of data, whereas measurement more properly consists in mapping the data on to an interval or ratio scale, though scaling can be thought of as a cruder form of measurement.

"The former may also be thought of as qualitative measurement (if this is not a contradiction in terms) and the latter as quantitative measurement. (p.484)".

Scale analysis orders or scales data to see if the data can be regarded as having a linear structure, and cutting points, or class boundaries are derived from the data. The semantic differential on the other hand assumes that the data form a ratio scale, or a series of ratio scales, which can be treated quantitatively. Linearity, the equality of intervals, and the location of the zero are taken a priori to be properties of the data and are regarded as being a function of the words used to define the scales.

## 7. Some Attitude Scales

The crudest measure of attitude is the single item method which is sometimes used in opinion polls. Respondents are asked whether they agree or disagree with an attitude statement and the population is described in terms of percentages for or against an issue. One of the difficulties involved has already been mentioned. The same phenotypic behaviour may be generated by different underlying attitudes, and there is no information given by the data about the underlying variable or variables. The wording of the single item may introduce other factors. A slight alteration of the wording could lead to a very different result, so most methods of measuring attitude employ a number of items.

An a priori scale is one where the items are ordered and stored according to their content. An example is the Bogardus (1925, 1933) social distance scales. For example subjects are asked if they would admit members of a particular race or group

1. to close kinship by marriage
2. to my club as personal chums
3. to my street as neighbours
4. to employment in my occupation in my country
5. to citizenship in my country
6. as visitors to my country
7. would exclude from my country.

Although these scales give useful results, certain shortcomings are evident. For example the intervals between the alternatives do not

appear to be equal. Further, the scale assumes that agreement with a higher degree of intimacy means agreement with all lower degrees of intimacy. That is, that the stimuli are monotone and can be arranged on a single continuum.

Likert (1932, 1934) also employs an a priori scoring system, somewhat similar to that developed for the semantic differential. In this type of scale subjects are asked to check whether they strongly agree, agree, are undecided, disagree or strongly disagree with a number of statements. As in the case of the semantic differential the properties of linearity and equal intervals, and the location of the zero on the continua of judgement are considered to reside in the language defining them. The justification for his approach lies in the degree of agreement between his scales and the more complex Thurstone scales. However as recent research by Tittle and Hill (1967) questions the reliability and validity of the Thurstone scale, despite its greater attention to the metric properties of data, another criterion of validity should be used. The same research finds the Likert scale to be superior to the Thurstone scale, in respect to reliability and predicting a behavioural criterion, in one instance.

Thurstone (1929) determines the score values of items empirically. The most commonly used method is the method of equal-appearing intervals. A large number of attitude statements about a given topic are assembled from newspapers, books or individuals, and presented to a panel of judges. The judges sort the statements into a number of piles according to the degree of positive or negative

affect towards the attitude object they are thought to represent. Each pile then represents an interval on the scale. Items on which judges fail to agree, that is those with large dispersions, are discarded, so are items which give undue replication of an interval. The final selection of items is made taking into account the brevity and clarity of questions. The construction of the questionnaire involves Task B judgements. Respondents to the final questionnaire endorse those opinions with which they agree. This involves Task A judgements and as in all the scales discussed above the respondents' behaviour is irrelative. The limitation of the Thurstone scale as a measuring instrument is that the attitude continuum is assumed from the start, and the unidimensionality of the scale cannot be inferred from the data. This approach assumes that attitudes are arranged naturally on a single continuum and the problem of the investigator is to find useful intervals to mark off along this continuum. Also inherent in this method is the assumption that the intervals determined by the judges (in the Task B behaviour) will be the same as those used by the population for which the scale is intended (in their Task A behaviour).

Methods of studying attitudes which rely on relative behaviour such as the Method of Paired Comparisons or the Method of Triads can also be used to construct scales but these will not be discussed here as they are less closely related to scale analysis and the semantic differential. Of course many other methods are used to elicit attitudes. For instance, projective techniques may give

useful information about attitudes, but they describe rather than measure attitudes.

#### 8. The Relationship of Attitudes to Behaviour

Attitude may be defined as behaviour at a covert level. For example, Osgood defines attitude in terms of representational mediational process. Or attitude may be defined as the manifest behaviour. For example, Guttman conceives of the item checking behaviour of an individual as a subset of the total set of behaviour related to an attitude object, which is defined as his attitude. As the definition of attitude may vary, ambiguity can arise in the discussion of the relationship of attitudes to behaviour. It is more precise to examine the relationship between attitude statement checking behaviour, and other types of relevant behaviour or action.

There are three levels of activity involved. At the genotypic level there is the underlying attitude which cannot be observed. At the phenotypic level there is the manifest behaviour which takes the form of item checking behaviour. Then, there is the behaviour which is to be predicted. This also has its phenotypic and genotypic components, though it is assumed that, at the genotypic level, the same attitude or variable underlies both the scale checking behaviour and the external behavioural criterion, or there would be no point in trying to predict the one from the other.

LaPiere (1934) provided some discouragement for the concept of attitude measurement. He travelled in the United States with two

Chinese companions, and observed the reaction his companions met in restaurants and hotels. Later, when he wrote, or telephoned, the proprietors asking if they would accept members of the Chinese race as guests in their establishment, he found that the vast majority (who had already done so) said that they would not do so. From this he concluded that the questionnaire is of minimal value.

Guttman (1950), also feels that attitude questionnaires are not easily related to behavioural criteria. He feels that prediction cannot be an integral part of attitude measurement, and the evaluation of techniques should be in terms of internal rather than external validity. A man may, for example, have a negative attitude towards his employer and yet act politely towards him. But this is perhaps an oversimplification. More than one attitude comes into play in any situation. The man's attitudes towards authority, losing his job, acting politely and so on, are also involved in this case.

Tittle and Hill (1967) suggest that the degree of correspondence between an attitude, as measured, and a behavioural criterion is at least a function of:

- "1. The measurement technique used
2. the extent to which the criterion behaviour constitutes action within the range of common experience, and
3. the degree to which the criterion behaviour represents a repetitive behavioural configuration (p.199)".

Tittle and Hill provide a summary of studies on the correspondence between measured attitude and behavioural patterns, which

suggests there is some correspondence to be observed if the investigator is careful to use appropriate procedures. This is reproduced (with slight alteration to provide fuller details for purposes of reference) as Table I. The table is fully reproduced because it provides a summary of research into the relationship between attitude and behaviour. A similar approach would be helpful in the study of the relationship between the semantic differential and Guttman scales, as each may take varying forms, and varying correspondences might be discovered, but unfortunately there seem to be fewer studies available.

The work of Tittle and Hill is also of particular interest because, although these authors do not directly aim to examine the relationship of scalogram analysis and the semantic differential, they do examine the relationship of these to a behavioural criterion, giving some evidence of the relative reliability and validity of the two techniques.

TABLE I : SUMMARY OF STUDIES OF CORRESPONDENCE

BETWEEN MEASURED ATTITUDE AND BEHAVIOURAL PATTERNS

<u>STUDY</u>	<u>ATTITUDE MEASURE</u>	<u>CRITERION</u>	<u>CIRCUMSTANCES</u>	<u>CORRESPONDENCE</u>
		Single Act (SA) or Patterned Behaviour (PB)	Unusual (U) or Normal (N)	
La Piere (1934)	Hypothetical single question	SA	U	Low
Kutner et al. (1952)	Single question	SA	U	Low
La Piere (1936)	Stereotypical single question	PB	N	Low
Bray (1950)	Summated rating scale	Set of SA's	U	Low
Corey (1937)	Thurstone-Likert scale	PB	N	Low
Zunich (1961)	Summated rating scale	Set of SA's	U	Low
De Fleur & Westie (1958)	Summated differences scale	SA	U	Moderate
Linn (1965)	Intuitive scale	SA	U	Moderate
Pace (1949)	No indication	PB	N	Low to Moderate
Rogers (1935)	Battery of single questions	PB	N	High
Murphy et al. (1937)	1 Thurstone scale	PB	N	High
	2 No indication	PB	N	High
	3 No indication	PB	N	High
Nettler & Goldington (1946)	Thurstone scale	PB	N	High
Poppleton & Pilkington (1964)	Thurstone, scored 4 ways	PB	N	High

Source: Tittle, C. and Hill, R. "Attitude Measurement and Prediction of Behaviour: An Evaluation of Conditions and Measuring Techniques". Sociometry, 1967, 30, p.203.

## CHAPTER II

### SCALOGRAM ANALYSIS

#### 1. Guttman's Definition of a Scale

Guttman defines a scale as follows:

"For a given population of objects the multivariate frequency distribution of a universe of attributes will be called a scale if it is possible to derive from the distribution a quantitative variable with which to characterize the objects such that each attribute is a simple function of that quantitative variable. (1944, p.142)".

By "attribute" he simply means a qualitative variable which has various values or categories. He gives the example of religion, which is an attribute which may have as its categories values such as "Catholic", "Buddhist" or "Jewish". By universe he simply means the population, or totality. However he calls the totality of subjects the population, and the totality of items, or the attributes which delimit them, the universe, to avoid confusion. There are two sampling problems involved in constructing an attitude scale. The population of possible subjects is sampled, and the universe of content, or the unlimited number of questions that could be asked about the attitude object must also be sampled. There are different sampling problems involved as the population can, at least in theory, be listed, while the number of possible questions that could be asked cannot be listed, and hence cannot be precisely defined. The universe is the concept whose scalability is being investigated, and consists of all the

attributes that define the concept. That is, it consists of all the attributes which are of interest to the investigation, which have a common content, so they are classified under a single heading which indicates the content. Thus he sometimes talks about "the universe of content".

If  $x$  is a quantitative variable, and  $y$  is an attribute or qualitative variable with  $m$  values,  $y$  is a simple function of  $x$  if the  $x$  values can be divided into  $m$  intervals which have one-to-one correspondence with the values of  $y$ . That is the  $m$  classes, which are delimited by the attribute  $y$ , and have no a priori ordering, can be mapped onto a continuum  $x$  to establish their order. If they cannot be mapped onto a continuum in this way, the classes cannot be ordered, and symbols are only nominal. The logical process of scaling is set out in a simple manner in Table 2.

Scale analysis is a method which examines qualitative data to see if it can be adequately characterised as a function of a quantitative variable. It is a formal analysis and

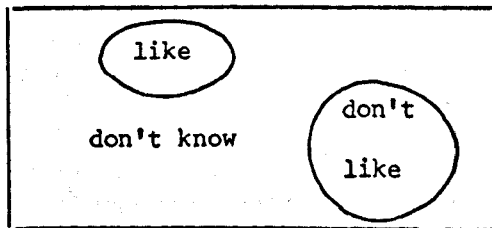
"hence applies to any universe of qualitative data of any science, obtained by any manner of observation (1944, p.142)."

Shapiro (1948) for example used the geographical southern states of America as the units of scale analysis. Podell and Perkins (1957) employed behaviour patterns rather than attitude statements to yield their scale data and Wohlwill (1968) used scalogram analysis to study the development of number concept in children. Scalogram analysis provides a general test of the dimensionality of data.

TABLE 2.

Methodological Note I : The Process of Scaling

1. A population of subjects is divided into classes according to whether they say they like, don't like, or don't know if they like the British.



The class definitions can be called values of the attribute according to which the population is classified.

$y_1$  = saying "I like the British",

$y_2$  = saying "I don't like the British",

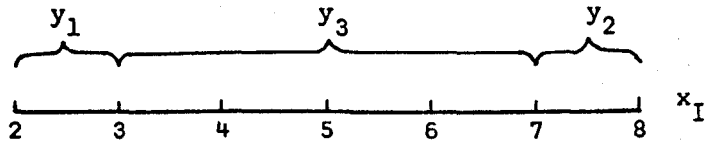
$y_3$  = saying "I don't know whether I like the British or not".

$y_1$ ,  $y_2$  and  $y_3$  form a nominal scale only. There is no ordering of elements: the property of = or  $\neq$  holds between any two elements but not  $<$ .

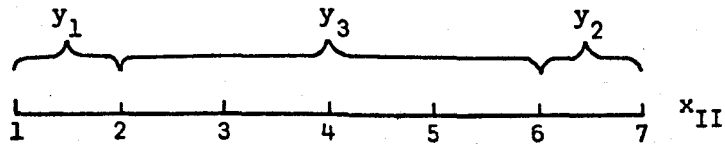
2. If  $y_1$ ,  $y_2$  and  $y_3$  can be mapped onto a quantitative variable they are said to scale. That is, the relation  $<$  or  $>$  holds between classes, in relation to the attribute. All the quantitative variables which might be chosen are identical in that they can be mapped onto one another. The unit of measurement is arbitrarily

TABLE 2 (continued)

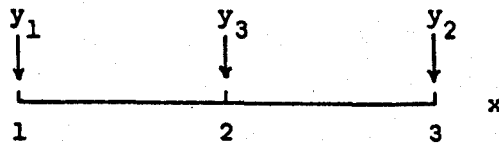
chosen, and there is no absolute zero. For example,  $y_1$ ,  $y_2$  and  $y_3$  may be mapped onto a quantitative variable  $x_1$ :



or mapped onto a quantitative variable  $x_2$ :



In practice rank order is used as the scale variable as this gives one-to-one correspondence between scale values and class labels.



The values  $x_1$ ,  $x_2$  and  $x_3$  can now be used to label the classes.

The classes can be ordered according to their scale scores.

$y_1 < y_3 < y_2$  because  $x_1 < x_2 < x_3$ . The items used to differentiate classes are also ordered by this process.

TABLE 2 (continued)

3. If this second step is possible the data are said to scale.

Scalogram analysis is the method by which one investigates the possibility of this mapping process, for any given set of data.

The importance of testing the dimensionality of data is underlined by M<sup>C</sup>Nemar (1946).

"Measurement implies that only one characteristic at a time is being quantified. The scores on an attitude scale are most meaningful when it is known that only one continuum is involved. Only then can it be claimed that two individuals with the same score or rank order are quantitatively, and within limits qualitatively, similar in their attitude towards a given issue (p.298)".

## 2. Cumulative Scales

The Guttman scale is a cumulative scale. That is, the stimuli are monotone. The following questions about probability illustrates what is meant by this property.

Suppose there is a population of voters in which 60% are Democrats and 40% are Republicans.

1. What is the probability that one person chosen at random will be a Democrat?
2. What is the probability that two people chosen at random will both be Democrats?
3. What is the probability that out of ten people chosen at random at least one will be a Democrat?

Responses can be right, or wrong. There are theoretically then 8 possible response patterns or scale types. However, looking more closely at the questions it can be seen that a person who is unable to answer the first question will be unlikely to be able to

proceed to the more difficult problems. Four response patterns, or scale types, could be expected to occur in practice. The first scale type would be represented by an individual with all the questions incorrect. The correct solution of question one only would be the second type. The third scale type would be the response pattern of a person with questions one and two right, but question three wrong. The correct answer to all three questions would provide the fourth scale type. The terms "more" and "less" have definite meaning. A person of the third type, for example, knows everything that a person of the second type knows, plus something else, with relation to this test. Also the response patterns are reproducible. If it is known how many correct answers a person gave to the questions, it is possible to say which questions he got right and which questions he got wrong. That is, the answers he gave to all the questions can be reproduced from the scale score.

The properties of cumulative scales were discussed by David Walker (1931), but Guttman developed, probably independently, their application to attitude measurement by generalizing to cases where answers cannot be categorized as right or wrong. To return to the above example. Suppose the possible answers below are given to students completing the test.

1. 0.5
2. 0.36
3. 0.42

Students completing the test are asked to say if the given answers are true or false. Assuming for the sake of the example that students get no questions correct by chance the four scale types can be set out, as in Table 3.

A scale type, then is the response pattern which distinguishes each class of students. It is perfectly reproducible from the scale score which is assigned to a class, so as to order the class in relation to other classes. That is, if it is known which class a particular student should be assigned to, it is also known what his responses to each question must have been.

Only the theory of perfect scales has been discussed. The perfect scale is the formal system: the properties of and relationships between elements follow logically from definitions, and the system is circular. It is the model, into which observed data are fitted. It is only when imperfect scales are considered that difficulties arise.

### 3 Methods of Scale Analysis

To determine whether or not a complex set of questions is cumulative, or forms a scale for a given population, it is necessary to test to what extent the individual responses are reproducible from scale scores, which denote scale types or classes of subjects. That is, subjects are divided into classes according to their response patterns. Subjects who do not fit into classes are assigned to the most similar class. Every response which does not coincide with the ideal class pattern, or in Guttman's terms, the scale type, is counted as an error.

TABLE 3

Example I : A Cumulative Scale

1. A number of students are asked the following questions:

Suppose there is a population of voters in which 60% are Democrats and 40% are Republicans.

1. What is the probability that one person chosen at random will be a Democrat?
2. What is the probability that two persons chosen at random will be Democrats?
3. What is the probability that out of ten people chosen at random at least one will be a Democrat?

Assuming no students get answers right or wrong by chance; no student who finds the first question too difficult will be able to proceed to the second; no student who finds the second question too difficult will be able to proceed to the third and so on.

The items divide the students into four classes. The first class of students is comprised of those failing all items. The second class of students is comprised of those passing one (the first) item. The third class of students is comprised of those passing two (the first two) items and the fourth consists of those who pass all items.

The data can be set out as follows:

TABLE 3 (continued)

	No. Correct	Class	Questions					
			1		2		3	
			P	F	P	F	P	F
Students	3	4	x		x		x	
	2	3	x		x			x
	1	2	x			x		x
	0	1		x		x		x

... (1)

2. Students are given possible answers to the questions:

1. 0.5

2. 0.36

3. 0.42

and asked to say if these are true (T), or false (F). It is then possible to arrange the data, without knowing the correct answers.

Question	3	2	1	3	2	1
Response	F	T	F	T	F	T
Student	4	x	x	x		
class:	3		x	x		
	2			x	x	
	1				x	x

... (2)

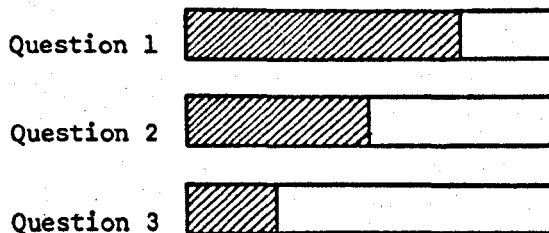
TABLE 3 (continued)

3. If the original items are monotonic stimuli, certain properties can be observed, and a characteristic parallelogram pattern appears when the data are arranged in this way.

Now a logical obversion is made, and it is deduced that when data can be arranged in this way the original items must be monotone stimuli. This is a valid step, as it is inherent in the definition of "monotone". However, the same stimulus does not always produce the same response. The stimuli are only monotone when they produce responses with the appropriate properties.

In other words it is a semantic convenience to speak of unidimensionality as being a property of the stimuli: it is the responses which generate a continuum.

Or the data can be tabulated. Assuming there are equal numbers of students in each class, for the sake of example, responses to each question can be represented as bar charts:





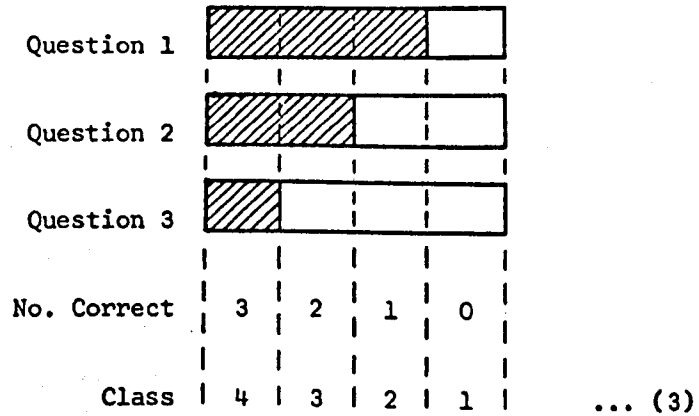
where  = proportion of students answering question correctly  
 = proportion of students answering question incorrectly.

TABLE 3 (continued)

Because the items form a cumulative scale, the bar charts can be connected, to give the response patterns of the classes of students:



The proportion of error represents the deviation from perfection, and is incorporated in the coefficient of reproducibility which is

$$1 - \frac{\text{number of errors}}{\text{number of responses}}$$
$$=$$
$$1 - \frac{\text{number of errors}}{\text{number of questions} \times \text{number of subjects}}$$

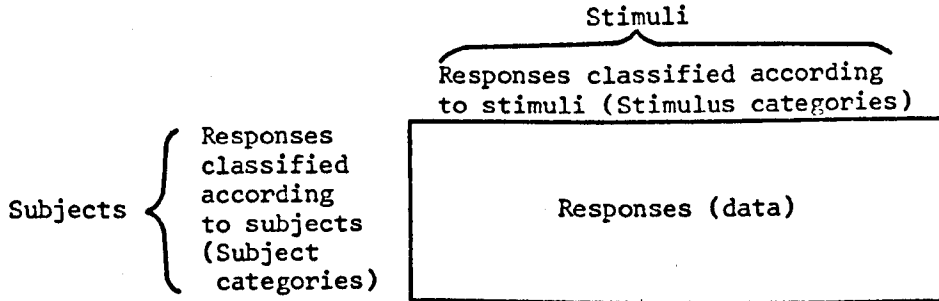
If one error occurs, the data does not fit the formal system, and the inference that the stimuli are monotone cannot be made on purely logical grounds. An attempt has been made to define statistically the probability that the inference is correct by Shuessler (1961), but the probability that a randomly selected domain of data would fit, or almost fit by chance, the perfect scale pattern is not known. For this reason Guttman urges that other criteria in addition to the coefficient of reproducibility, should be taken into account when deciding if the data form a scale pattern.

A general model of the scalogram matrix of data is given in Table 4. Of course, this is not the most general model that could be constructed. It is in terms of stimulus and response because the primary concern here is with the measurement of psychological variables. Scale analysis could of course be applied to other types of data. Similarly the use of horizontal categories for items, and vertical categories for respondents is conventional rather than necessary.

Four basic forms of scale analysis, based on the same theory, have been used by Guttman and his associates. They are the least squares method, the scalogram board technique, the Cornell trial

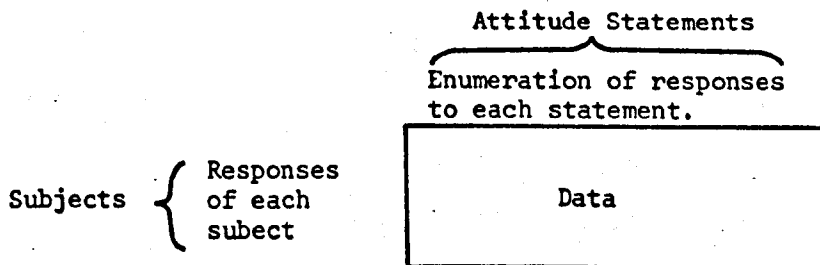
TABLE 4

Methodological Note II : The Matrix of Data in Scalogram Analysis



The stimuli are selected to produce responses which are thought to be indicative of the attribute to be investigated. The scalogram is the pattern of responses. The aim is to arrange this to it as nearly as possible the perfect scale model, to infer unidimensionality.

A more specific case is involved in attitude measurement. The stimuli are usually statements about an attitude object. The responses are in the form of agreement or disagreement, or some degree of these. The responses are the check marks made, by the subjects, on the questionnaire.



scoring and graphic technique and the tabulation technique.

The least squares method was described by Guttman (1941) in "The Quantification of a Class of Attributes". At that time it was thought to be of theoretical rather than practical value.

It involves

"rather lengthy, though simple, numerical calculations and it can often be usefully approximated by simpler and even intuitive - procedures (p.323)",

but, he continues

"it is of primary importance to define first a "best" answer so that one can know what it is that is being approximated (p.323)".

Since this time the method has been adapted and developed so that computer techniques can be applied, for example by Slater (1956), and it is a practical method. The least squares method is still, however, a useful exposition for the understanding of Guttman's scale theory.

The matrix of data may be thought of as being infinitely large. In practice the responses of only a small number of individuals to a small number of questions, is known. Given that unidimensionality exists, or in Guttman's words that the responses (or the class of attributes) are a single class of behaviour, it is possible to make predictions about the missing portions of the matrix.

"We should like to know, for example, on the basis of our knowledge of the behaviour of an individual on a sample of the items, what his behaviour would be like on the totality of items. Or we would like to

know, on the basis of our knowledge of the behaviour of a sample of individuals on a particular item, what the behaviour of the population of individuals would be with respect to that particular item ...

We propose, then to quantify the class of attributes to attempt to facilitate the achievement of these ends. We shall derive a single set of numerical values for the items that will tend to be best for these purposes (p.322)".

It may be of value to ask here to what extent this quantification is a process of measuring attitudes. A measure of attitude is constructed from the observed item checking behaviour. There are two sampling problems. The first is the problem of sampling individuals. The second is the problem of sampling items. The attitude can be conceived as an infinite number of responses: the responses of all people to all the possible items relating to the attitude. Guttman prefers to call the data in the matrix behaviour, rather than responses, because the behaviour of an individual is not considered to be a single value, but a distribution of the acts he performs. It is consistent behaviour, rather than discrete responses. If there is more than one attitude involved, if the sample has been drawn from more than one of these matrixes, it is not possible to quantify the data as a function of a single variable:

"Now every set of items cannot be usefully quantified for every population. Only if the behaviour of each individual in the given population is sufficiently consistent for the given set of items can a single set of numerical values efficiently reproduce the entire

behaviour of the population. Our method will give a measure of the utility of regarding the set of items as a single class of behaviour for the given population (p.322)".

Some of the characteristics of the method are:

1. That

"no a priori judgements as to whether or not one act should be assigned a higher value than another" are made.

"The desired values" he continues "come out automatically by analyzing the behaviour of all the individuals in the population simultaneously, thus taking the entire configuration into account in one stroke (p.323)"

That is, the items or stimuli are selected a priori only to the extent that they are relevant or not, it is not necessary to judge which responses are indicative of a certain degree or direction of attitude.

2. The behaviour of an individual is not considered to be a single value, but a distribution of the values of the acts he performs. In other words, in scale analysis, a single score is eventually allotted to each individual, but this denotes the distribution of his responses to all the items, which is reproducible from his scale score, within a certain margin of error.

3. From the sample of responses or acts, a method is available "for predicting in a best general manner what his behaviour will be for acts outside the sample (p.324)", and similarly a method is available for

"predicting in a best general manner what the behaviour of other individuals will be with respect to the act (p.324)".

4. For prediction purposes it is useful to assign each individual a single numerical value. This is proportional to the arithmetical mean of his frequency distribution of behaviour, and is merely to fix the position of his distribution with respect to that of the rest of the population. The method then employs addition, which makes certain metric assumptions, but this is a matter of convenience rather than necessity.

5. Item subcategories may be characterized by the frequency distribution of the values of individuals who check them. The categories, then, can be assigned a value proportional to the arithmetic mean of this distribution.

6. The method fulfils the requirements of internal consistency.

The mathematical development of the least squares method is outlined in Appendix 1.

The scalogram board is a device which permits the shifting of both horizontal and vertical rank order. Individuals' responses to each category of each question can be entered onto the board, and subject and question categories can be ranked in a series of successive approximations to the desired parallelogram pattern. It would be possible to perform the analysis without knowing the content of the questions, but in practice it is easiest to make judgements about the "favourableness" or "unfavourableness" of responses. Subjects can then be arranged in descending order of "favourableness". This arrangement

serves to reduce the number of approximations necessary to produce the desired parallelogram pattern. Once this initial ordering of respondents is established, for example by ranking respondents according to the number of positive answers, it is possible to hold the order of respondents and rank the question categories. It is possible to combine overlapping categories so as to reduce the number of errors. This is particularly useful in determining the scoring of neutral categories. The item categories are ordered such that the rank order of positive response categories runs from the lowest positive frequency to the highest positive frequency, while that of the negative response categories runs from the highest negative frequency to the lowest negative frequency. The individuals are then reranked according to their corrected (by means of category combination) scores. Cutting points or class boundaries are decided upon, and responses falling outside these are counted as errors. The cutting points are placed so as to produce as few errors as possible. The cutting points are the class boundaries which distinguish classes of individuals according to their scale type, or distribution of responses.

A number of criteria must be taken into consideration when the scalability of a set of data is being evaluated. The coefficient of reproducibility is one criterion. Guttman (1950) suggests that a coefficient of reproducibility of 0.95 or above is acceptable. However, as the probability that this will occur by chance is not known, and as the coefficient is relative to the marginal frequencies, a high coefficient of reproducibility is not sufficient evidence that the

data approximate the model of a perfect scale. For example if 90 per-cent of the subjects fall into one answer category, the maximum possible error for that question is 10 percent. That is, the coefficient of reproducibility of the question cannot be less than .90. To guard against spuriously high reproducibility it is important to have at least some of the questions with a 50-50 type of marginal distribution. In general

"the more items included in a scale the greater the assurance that the entire universe of which these items are a sample is scalable (Guttman (1950) p.79)".

and

"the more categories that remain uncombined the more credible is the inference that the universe is scalable (Suchman (1950) p.117)".

Scalogram analysis is a trial and error method which may capitalize on chance errors, if only a small number of items are used, and categories are extensively combined. The probability that completely random data could be found to fit the perfect scale pattern is not known, but it is assumed to be less probable if more items are used. Guttman does not seem to think that the problems of sampling items differ very much in kind from the problems involved in sampling a population of people.

The existence of blocks of non-scale types may suggest that more than one dimension is involved. Wallin (1953) for instance found, in constructing a scale for measuring women's neighbourliness, that the behaviour scores of people over 60 could not be classified as scale

types. Age related factors, rather than neighbourliness, restricted contacts with neighbours.

The Cornell trial scoring technique, described by Guttman (1947) has the advantage of being a pencil and paper technique, and the number of subjects and item categories is not restricted to 100 x 100 as is often the case with the scalogram board. However it is a less flexible method, when the rearrangement of item categories is necessitated. Like the scalogram technique the Cornell method begins by making preliminary judgements about the "favourableness" of responses, and consequent analysis verifies, or indicates a need to revise these preliminary judgements. A score is obtained for each individual by summing the weights of question categories and questionnaires are arranged in rank order. A table is prepared which allows one row for each individual and one column, subdivided into the appropriate number of category columns, for each question. A cross is placed in the appropriate cell to mark off the responses of each individual to each question. If the data approximate the perfect scale, and the preliminary order of individuals is the scale rank, then a simple pattern of responses should emerge. A simple example of this type of pattern is illustrated in Figure (1) in Table 3. In a more complex instance some combination of categories is usually required to produce the appropriate degree of reproducibility. This is attributed to the influence of verbal habits. Some people are thought to say "strongly agree" where others with the same attitude say "agree". Though they have the same position on the basic continuum

TABLE 5

Example II : The Scalogram Board Technique

The attitude statements A, B and C, with possible responses

A: 1, 2 or 3;

B: 1, 2 or 3;

C: 1, 2, 3 or 4

are presented to ten subjects.

Step I. Responses are entered on the board, and individuals are provisionally ranked.

S's	Statements									
	A			B			C			
	1	2	3	1	2	3	1	2	3	4
1	x			x			x			
2	x				x		x			
3		x			x			x		
4	x				x			x		
5		x			x			x		
6			x		x		x			
7			x		x			x		
8		x				x			x	
9			x			x			x	
10		x				x			x	

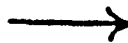
TABLE 5 (continued)

Step II. Categories may be combined to reduce error. Responses in the categories of item A are seen to contain error.

Combining  $A_1$  and  $A_2$  would leave 2 errors. Combining categories  $A_2$  and  $A_3$  leaves 1 error. Categories  $A_2$  and  $A_3$  are therefore combined. Category  $C_4$  is dropped as it contains no responses.

Item categories are then ranked:

S's	Statements							
	A		B			C		
	1	2+3	1	2	3	1	2	3
1	x		x			x		
2	x			x		x		
3		x		x			x	
4	x			x			x	
5		x		x		x		
6		x		x			x	
7		x		x				x
8		x			x			x
9		x			x			x
10		x			x			x



S's	Statements							
	B	C	A	B	C	A	C	B
	1	1	1	2	2	2+3	3	3
1	x	x	x					
2		x	x	x				
3				x	x	x		
4			x	x	x			
5		x		x		x		
6				x	x	x		
7				x		x	x	
8					x	x	x	
9					x	x	x	
10					x	x	x	

TABLE 5 (continued)

Step III. The combination of categories may alter the rank order of individuals. Intervals 3 and 4 are now incorrectly ordered.

This is corrected:

S's	Statements								Scale Types
	B	C	A	B	C	A	C	B	
	1	1	1	2	2	2+3	3	3	
1	x	x	x						1
2		x	x	x					2
4			x	x	x				3
3				x	x	x			4
5		(x)		x		x			4
6				x	x	x			4
7				x		x	x		5
8					x	x	x		6
9					x	x	x		6
10					x	x	x		6

Errors are ringed and counted, and the coefficient of reproducibility calculated.

the behaviour differs at the phenotypic level. As Marshall (1963) colourfully expresses it "one man's 'blast' is another man's 'damn'". Although the importance of verbal habits is often recognized, it is extremely difficult to show what position a person should be assigned to on the basic continuum, unless his manifest behaviour is accepted as being a manifestation of his position on the basic continuum. However categories can be combined without using this particular psychological interpretation: responses can be reclassified. For example, if two categories  $A_2$  and  $A_3$  are combined there is no need to say that this is the result of a verbal habit. If the item A has three categories it can be preproduced from the scale score whether an individual's response was  $A_1$  or ( $A_2$  or  $A_3$ ). That is,  $A_1$  or not  $A_1$ . Categories are combined so as to minimize error. Questionnaires are rescored and reranked, and responses are tabulated on a new form. The cutting points, which separate respondents of different scale types are located, and responses which fall outside the correct cutting points are counted as errors. The error of reproducibility is computed and interpreted with reference to marginal frequencies, the pattern of errors, the number of questions and categories.

The tabulation technique, developed by Goddenough (1944) utilizes the type of data presentation illustrated in Figure (5) in Table 3. Goodenough suggests that the method is more rigorous than the scalogram board technique because it does not rise chance errors in the same way. It is capable of dealing with large numbers of cases

while preserving the flexibility of the scalogram board technique.

Its limitation is that:

"for practical use, the relative rank order of categories for each item must be assumed in advance, which is not necessary with either the least squares or scalogram board technique (p.179)".

The tabulation technique is initiated by the construction of an ideal scale, as determined by the marginal frequencies. The expected frequencies of the ideal scale types are calculated. The data are then sorted into the observed scale types or response combinations. Adjacent categories may be combined to increase reproducibility. The deviation from perfection is expressed by Goodenough as the percentage of non-scale types which occur: this is not directly related to the coefficient of reproducibility which incorporates the proportion of errors, not of non-scale types. However a coefficient of reproducibility can be calculated from the data if this method has been used.

Other methods of scale analysis have recently been developed. The scalogram board technique has been described by Suchman (1950), Trenaman (1960) and Oppenheim (1966). The main drawback with this technique is the limitation placed on the size of the sample. Kahn and Bodine (1951) describe a method of machine analysis which is similar to the scalogram board technique, which has the added advantages of providing a record of operations, and of being applicable when a large sample is considered. This method would probably be the best method to use if IBM equipment was available. Slater (1956) also

TABLE 6

Example III : The Pattern of Errors in Scalogram Analysis

1. The coefficient of reproducibility is relative to marginal frequencies: some items may not scale, even though the coefficient of reproducibility is high.

S's	Statements					
	A		B		C	
	1	2	1	2	1	2
1	x		x			x
2	x		x		x	
3	x		x		x	
4		x	x		x	
5		x	x		x	
6		x	x		x	
7		x	x		x	
8		x	x		x	
9		x		x	x	
10		x		x	x	

The visual pattern of statement C responses indicates that it does not scale with the other two items (unless the direction of categories is reversed). The coefficient of reproducibility of statement C is:

$$1 - \frac{\text{number of errors}}{\text{number of responses}}$$

TABLE 6 (continued)

Hence coefficient =  $1 - 1/10 = 0.90$

The pattern of errors must be considered.

2. Classes of subjects may be non-scale types. Items may not scale for some groups of subjects.
3. The pattern of errors should be random, and the marginal frequencies should be fairly even.

describes a method of machine analysis, but this is based on the least squares solution, and treats the variables as quantitative. The typewriter notation method described by Waisanen in "Individuals, Groups and Economic Behaviour" by Hickman and Kuhn (1956) is quick, simple and cheap, but tends to oversimplify the structure of the data. Other methods such as the method of linear segments, Mander(1952), and a method utilizing summary statistics, Green (1956), may also be employed. Ford (1950) employs the property of geometric progressions, that the sum of the series is uniquely obtained, to maintain reproducibility of scale scores. These methods tend to be variations of the methods outlined, or adaptations of these to allow the use of more advanced means of computation. They have not been fully outlined here as the main interest is not so much with the characteristics of each method, as with the theory of scale analysis which underlies all these procedures.

#### 4. The Universe of Attributes

There are really two distinct uses of scale analysis. The first is descriptive. Given a set of data it may be found that this can be adequately represented as a function of one variable: individual response distributions can be described by a single scale score, and an order between questions, and between classes of subjects is generated. The structure of the data given has been investigated, but the investigator has not gone beyond the given data. The second use is inferential. The investigator gives the variable a name, such

as "attitude towards something" and uses the data to measure a latent variable. Guttman describes this in terms of the "universe of attributes". The term "attribute" can be a little ambiguous as he seems to use it for both the variable, if there is found to be one, and the value it takes. The universe of attributes is the concept whose scalability is being investigated; it is:

"all the attributes of interest to the investigation which have common content, so that they are classified under a single heading which indicates that content (Guttman (1944) p.141)".

If an attitude questionnaire is constructed, it is evident that there may be an indefinitely large number of possible questions. This indefinitely large number of possible question wordings and forms, from which the questionnaire items are drawn, is the universe of attributes for any particular investigation. The universe is defined in terms of its content: an attribute belongs to a universe by virtue of its content. The investigator

"indicates the content of interest by the title he chooses for the universe, and all attributes with that content belong in the universe (Guttman (1944) p.141)".

For example the universe of content of the statistics test mentioned previously (see Table 3) would presumably be all possible questions which could have been asked about probability. The universe of attributes is presumably all the responses which could have been made to these.

As the data, which relates items and individuals, are responses it may be more clear to think of a universe or population or large class of responses. The sampling problem can then be represented visually as in Table 7. The universe of content then becomes all stimuli, thought to be related in terms of some dimension and able to be ordered by the relations ( $>$ ) and ( $=$ ) in relation to this. The universe of attributes is the total number of responses which can be made to these stimuli, thought to be indicative of the dimension, or to be behavioural manifestations of it. The population is, of course, the total number of people that the investigator is interested in, and wishes to order in relation to their responses to the stimuli, and hence in relation to the dimension that the responses to stimuli are thought to be indicative of. This approach may seem a little unwieldy but it is important to define exactly what scale analysis aims to do, if the criticisms of scale analysis are to be countered. Critics sometimes argue that scale analysis is not a measure of unidimensionality, or that it is a too stringent test of unidimensionality, while holding different ideas of what constitutes unidimensionality.

If the universe of data is scalable, that is, if the universe of responses can be adequately represented as being generated by, or being a function of one underlying variable, the ordering of persons based on a sample of items will be essentially the same as that based on the universe. Adding further items serves to differentiate further scale types, but not to interchange the order of types already in the

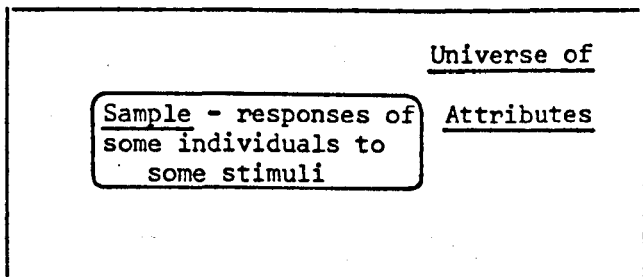
TABLE 7

Methodological Note III : The Universe of Attributes

Responses in Universe

(all possible items or stimuli that could be used for the investigation, related by relevance to the investigation; all stimuli, responses to which are thought to be indicative of some dimension).

Responses  
of Population  
(all possible  
people that  
could be  
investigated).



The universe of data, or responses, is sampled: the population of people is sampled and the universe of items is sampled. Different sampling problems are involved as the people can, at least in theory, be enumerated, while the universe of content is of unknown size.

sample. The converse of this process is seen when questions are deleted: the order remains intact but the number of scale types, or classes of respondents, diminishes. If the universe is scalable the ordering of individuals is not dependent on a particular section of items.

It remains to be demonstrated, however, that given the sample data forms a scale, the universe can be inferred to be scalable. Guttman suggests that

"at present it seems quite clear that in general the problem of finding a sample of attributes to form a scale by chance for a sample of individuals is quite negligible, even if there are as few as three dichotomous items in the sample and as many as one hundred individuals (1944, p.148)".

But he does not give the basis for his comment. As the universe of data is unlimited, and cannot be enumerated, the probability of drawing data, from a non-scale domain, such that they can be arranged to fit the scale pattern, is not known. Torgerson (1958) for example, states that one of the limitations of Guttman's method of scaling resides in its incapacity to test whether the observed scale data is a good fit. This is slightly different criticism, because it relates to the evaluation of observed data, rather than to the relationship of the observed data to the universe of data. Although this is put in different terms it is a familiar problem. Scale analysis can be used to examine a set of data to see if it can be adequately represented as a function of one variable, but to go beyond the observed data and attempt to name

the variable, and say what it is that is being measured, requires further empirical research or non-mathematical inferences about the content of stimuli. As Guttman puts it, the validity of an attitude scale constructed by means of scale analysis has internal rather than external validity. He concentrates on internal validity as he is sceptical of the possibility of finding external criteria. However, as it was suggested in the first chapter, a correspondence between attitude and behaviour may be observed if the behavioural criterion is not too naïvely selected.

##### 5. Statistical Test of "good fit" for Scalogram Data

Concentrating on the observed data, then, at what point can it be decided that the data do fit the scale model well enough to be adequately represented as functions of one variable.

Shuessler (1961) suggests three approaches for testing the significance of results. The third does not treat data as purely qualitative, and so tends to assume what scale analysis is trying to prove. Therefore, only the first two will be described here. The procedures suggested by Schuessler are meant to supplement rather than replace Guttman's criteria. They would normally be applied, he suggests, in advance of scale analysis to rule out the null hypothesis: that the data is randomly distributed. In some cases the significance of the scalogram pattern may appear obvious, and a null test of chance unnecessary, but the methods he sets out do give a framework of statistical inference to the investigator using scale analysis.

TABLE 8

Example IV : Statistical Test of Significance of Scale Findings

Suppose the response distributions are as follows:

1. Marginals:

Item	Response	
	+	-
1	0.80	0.20
2	0.60	0.40
3	0.35	0.65

2. Distribution of response patterns:

Number of positive responses		Item			Frequency (observed)
		1	2	3	
(x)	3	+	+	+	30
	2	+	+	-	25
	2	+	-	+	3
	2	-	+	+	1
	1	+	-	-	22
	1	-	+	-	4
	1	-	-	+	1
	0	-	-	-	14
					<hr/> 100

TABLE 8 (continued)

3. Test I.

This compares the number of scale types expected (on the hypothesis that responses to successive items are statistically independent) and the number of scale types observed. Scale types are designated. The highest expected frequency is used to choose between response patterns representing the same number of positive responses. In the event of this, the highest observed frequency is chosen. Expected frequencies can be tabulated as follows:

x	Response Pattern	$P_{1j}$	$P_{2j}$	$P_{3j}$	$P_{1j} \cdot P_{2j} \cdot P_{3j}$	Expected Frequency	Selected Scale Types
3	+ + +	0.80	0.60	0.35	0.168	16.8	*
2	+ + -	0.80	0.60	0.65	0.312	31.2	*
2	+ - +	0.80	0.40	0.35	0.112	11.2	
2	- + +	0.20	0.60	0.35	0.042	4.2	
1	+ - -	0.80	0.40	0.65	0.208	20.8	*
1	- + -	0.20	0.60	0.65	0.078	7.8	
1	- - +	0.20	0.40	0.35	0.028	2.8	
0	- - -	0.20	0.40	0.65	0.052	5.2	*
$\Sigma = 1.000 \quad \Sigma = 100.0$							

$P_{ij}$  = proportion of persons giving  $j^{\text{th}}$  response to  $i^{\text{th}}$  item.

Because there are three questions there are three possible ways of scoring, say 2: + + -, + - +, - + +. Of these the permutation + + -

TABLE 8 (continued)

is selected as the scale type as it has the highest expected frequency of 31.2. The correct permutation of other combinations is similarly chosen. The extreme scores must always be scale types as they are necessarily reproducible.

The individual probabilities of scale types are added to give the probability of a scale type on a single trial:

$$\begin{aligned}\text{Probability} &= 0.168 + 0.312 + 0.208 + 0.052 \\ &= 0.74\end{aligned}$$

Expected frequency of scale types, if questions are independent, is

$$0.74 \times n = 0.74 \times 100 = 74$$

This gives a  $\chi^2$  value of 15.02 which, with one degree of freedom, exceeds the value expected by chance.

But this does not give the direction required: there may be fewer scale types than expected by chance, but this is simply checked as there were 91 observed scale types and 74 expected scale types.

It is not possible to test the hypothesis of perfect scalability by probabilistic methods, but weaker alternatives can be tested.

The chance hypothesis can be tested and rejected as above. Varying degrees of stringency can also be tested. For example, it would be possible to test the hypothesis that the probability of a scale type occurring is equal to, or less than 0.8. The rejection of this hypothesis leads to the acceptance of the alternative, that the probability that items are not independent is greater than 0.8.

TABLE 8 (continued)

4. Test II.

The calculation of expected frequencies and designation of scale types is as before, but the observed and expected frequency of scale types is compared within score intervals. This procedure gives a more refined statistical interpretation of the matrix of responses as it indicates at what points the divergence from chance is negligible. It is necessary though to determine visually if the observed scale frequencies exceed or fall below those expected by chance.

	x	Observed	Expected	O - E	$\frac{(O-E)^2}{E}$
*	3	30	16.8	13.2	10.37
*	2	25	31.2	-6.2	1.24
(non-scale)	2	4	15.4	-11.4	8.44
*	1	22	20.8	1.2	0.07
(non-scale)	1	5	10.6	-5.6	2.97
*	0	14	5.2	8.8	14.90
		<hr/> 100	<hr/> 100.0	<hr/> 0.0	<hr/> 37.98

The probability of  $\chi^2$  being greater than 37.98, with five degrees of freedom, is less than 0.1 so the hypothesis that the items are independent is rejected.

Source: Shuessler, K.F. A note on the statistical significance of the scalogram. Sociometry, 1961, 24, 312-318.

On a practical level there are certain difficulties limiting the routine application of these methods. They require the dichotomization of response categories. Dichotomization is not always necessary to produce a highly reproducible scale, and in any case the basis for dichotomizing data is usually established as part of the scale analysis procedure. However this method could be adapted for use with complex distributions. Furthermore as the technique enumerates and establishes probabilities for all the observed response patterns, the technique would be too unwieldy to employ with a large number of items and respondents unless a machine technique was developed to handle the operations. However it is important that

"it is possible to construe the matrix of observations (scalogram) in such a way that statistical hypothesis may be tested, and inferences drawn in respect to the presence of basic dimensions which could account for the configuration of observed responses. (Schuessler (1961) p.312)".

The tests of significance of scale data allow the rejection of the hypothesis that the items are independent, and the acceptance of the hypothesis that items are related, or in Guttman's terms, drawn from the same or related universes of content. Scale analysis produces high internal consistency, and gives a highly reliable scale, but the problem of validity remains. The investigator must use other criteria for deciding if he is sampling the universe he wishes to investigate, and not sampling, consistently, another universe. That is, it must be assumed that the stimuli produced responses which were manifestations of the attitude which was to be studied, not consistent

manifestations of say, some other attitude. For example a questionnaire full of generalizations about racial groups may measure the individual's attitude to generalizations, or his tendency to agree with generalizations rather than his attitude to the racial groups.

It is common to use a split-half reliability coefficient to evaluate the reliability of an attitude scale. This is because attitudes may change quite quickly over time. If there is an attempt to measure the same attitude, of the same individual, several times a series of discrepant results may be due either to the instability of the measuring instrument, or to genuine changes in the individual's attitude. Therefore it is customary to try and rule out the time factor by comparing the result based on one half of the questions (the even questions) with the other half of the questions (the odd questions) on a single questionnaire. By definition scale analysis tests the reliability of every question and the more perfect the scale pattern is the higher the reliability of the whole scale. The perfect scale is perfectly reliable in that each score, and hence each response to every question, is a function of the scale variable. It is a property of perfect scales that the ordering of individuals is invariant, whether odd or even questions are used to generate the order. As the scale deviates from perfection, so its reliability is reduced. But the reliability of a Guttman scale cannot be measured by the split-half reliability coefficient as this is based on correlation methods. The correlation between items, or sets of items may range from 0 to +1, when the data form a perfect scale, given that the categories of items

which scale in a reverse direction have been reordered, otherwise from -1 to +1. The items are functions of the scale variable, but need not be correlated with one another.

#### 6. Further Components of Scales

Guttman (1954) has recently become interested in the functions of content scales which he identifies with principle components. The intensity component was first discussed as a solution to the problem of bias; an attempt to locate an absolute zero on the attitude continuum. As well as asking a number of questions about the attitude (content) the investigator asks a number of questions about how strongly the respondent feels about the issue expressed in the items. The intensity responses can be scaled like the content responses, and plotted against them to give a J or V shaped curve. The lowest part of the curve is said to divide the population into those "for" and those "against" the issue in a manner which is invariant whether the questions are favourably or unfavourably biased. The intensity curve can also be derived from the content scale, by means of weighting scores in a different manner. This led Guttman to try to give a psychological interpretation of further components of scales. As data from other sources has the same components, provided it is found to fit the cumulative scale model, these psychological interpretations should be considered with caution. For example the second component of the statistics test given as an example in Table 3 can hardly be interpreted as an intensity function. However Guttman is to be praised for

clarifying the distinction between intensity and content and suggesting a possible curvilinear relationship, as intensity and content questions may be merged in an attitude questionnaire, without their relationship being considered. For example "strongly agree", used for example in Likert scales merges content and intensity.

#### 7. Criticisms of Scale Analysis

Some of the most well known investigators in the field of attitude measurement have criticized scale analysis. Festinger (1947) unfortunately wrote his comments on scale analysis before the main theoretical development was published in Stouffer et al. (1950). Festinger says that the technique of scalogram board analysis

"is not very rigorous. Outside of possible hunches the investigator may have concerning how the questions will scale themselves the method is largely one of trial and error and inspection (p.153)".

The method is one of successive approximations, which will converge into the proper scale ordering, if the scale exists. If the data do not fall into the scale pattern with ease, and a great deal of trial and error is needed then it is probable that the data do not fit the model of the perfect scale. Scalogram board analysis could be performed blindly, hunches serve only to reduce the number of iterations needed to produce the final result. Festinger feels that the tabulation and Cornell techniques are more rigorous, even though the former cannot be performed blindly. It is important to remember that

"the basic theory of scale analysis is not to be confused with particular techniques for carrying out such an analysis in various kinds of situations (Guttman (1947) p.462)".

Festinger feels that many existing measuring instruments may not meet Guttman's criteria for unidimensionality. This seems a justified conjecture, but it does not necessarily imply that scale theory is inadequate. It may be that the existing measuring instruments have utilized less stringent criteria. For example, a later study by Clark and Kriendt (1948) failed to establish unidimensionality for the Rundquist-Stetto Scale of Economic Conservatism, and a study by Eysenck and Brown (1949) found the reproducibility of their anti-semitic scale to be less than 0.85. Guttman (1951) suggests that in the latter case the items belong to distinct, though possibly inter-correlated, universes. Some items, such as "The Jews will stoop to any kind of deceit to gain their own ends" are related to an allocation of Jews along a good-bad dimension, while others are relative and compare Jews with other people. For example "There is no reason to believe that innately the Jews are less honest and good than anyone else". These may function separately. Jews may be regarded as different from others because they are better or because they are worse, and they may be regarded as being the same as other people because they are good, like everyone else, or bad, like everyone else. If items are divided into groups of scales, and the relationships between these empirically examined, the resultant structure will clearly be more meaningful than that given by an aggregation of correlated items.

Festinger questions the utility of scale analysis suggesting that few scales will be found in practice. Time has not proved him correct, and scale analysis has been used quite extensively, not only for the construction of attitude scales, but also for demonstrating the unidimensionality of other data. For example, Riley et al. (1954) detail some applications to complex sociometric data. Wohlwill (1968) studies the development of number concept in children using scalogram analysis and Kofsky (1968) similarly studies the development of logical operations, Wallin (1953) constructed a scale of neighbourliness and Tittle and Hill (1967) employ scale analysis to index behavioural criteria in studying the validity of attitude scales. Festinger also says

"It would appear futile to insist upon unidimensional scales or to make very much of a distinction between scales which possess different 'degrees of unidimensionality' such as the distinction between scales and quasi scales (1947, p.159)".

This failure to appreciate the distinction between scales and quasi-scales is echoed in Eysenck's (1951) criticism, and explains the difficulties encountered by Clarke and Kriendt (1948). A scale is a set of data which closely approximates the perfect scale pattern. The properties of the perfect scale are attributed to the data and discrepancies are regarded as a source of error. The quasi-scale is a set of data which to some extent approximates the structure of a perfect scale, but the properties of the perfect scale cannot be attributed to it without introducing so much error that the procedure

is not worthwhile. Guttman (1951) gives an analogy to clarify the idea of the perfect scale as a limiting structure. Consider a circle. One way of arriving at a circle is to picture a series of equilateral, closed, polygons, each having one more edge than the preceding one. The first is an equilateral triangle, the second a square and so on. If all the polygons have the same area they have a limiting form which is called a circle. It would be possible to call the limit a perfect circle, and the polygons "quasi-circles". As the theory of perfect circles is more simple than that of polygons it would be convenient for practical purposes to treat a million sided polygon as if it were a circle. But this does not mean that the triangle or square should be treated as if it were a circle. These forms deviate too much from the limiting structure for this to be of value. Their properties are entirely different from those of a circle, and they are worth studying in their own right. If a number of people were working with these forms it would be convenient to agree upon the point at which figures should cease to be regarded as quasi-circles, and begin to be regarded as circles, though this would be arbitrary.

The point at which a set of data can usefully be regarded as fitting the perfect scale model is similarly arbitrary. The notion of reproducibility was introduced to help the investigator to make this decision. Guttman suggests that

"Festinger seems to have misunderstood the definition of a quasi-scale for he seems to believe that it differs from a scale only with respect to reproducibility. In a quasi-scale there is one dominant

factor and an infinitely large number of minor factors: the quasi-scale differs from the scale in that it does not have an intensity function or further components of that kind (1947, p.462)."

Exact cutting points cannot be established on a quasi-scale, it is merely possible to say that a person above a certain point will probably have answered a question in a particular way.

Clark and Kriedt (1948) applied scale analysis to data collected by administering the Rundquist-Stetto Scale of Economic Conservation to 306 psychology students. Questionnaires were randomly divided into two groups and each writer worked independently on a group of questionnaires. They experienced some difficulty with the analyses.

"The writers found the process of assigning cutting points for computing reproducibilities to be difficult and arbitrary. Guttman's description of methods for computing reproducibilities were found to be inadequate for ensuring that each of the writers would use the same criteria for determining these points (p.217)".  
and

"Although Guttman implies that it is relatively easy to determine how to combine categories the writers found it difficult with the present data, and often disagreed with each other (p.222)".

This could be expected from the final result: the data was not found to scale. Their analysis tends to confirm what Guttman says about quasi-scales.

Eysenck (1951) gives a number of criticisms of Guttman's work, which are similar to those put forward by Festinger. First, scale analysis does not seem to fit into Eysenck's theoretical attitude hierarchy except insofar as it can be used to measure attitudes at the lowest stratum:

"It is clear that the closer we come to such a type of scale the more trivial will be the results achieved until triviality is complete in the ideal case (1951, p.99)".

He suggests that if data are found to scale perfectly they must be trivial. Let us consider an example from the field of mental testing, where factor analytical methods, which Eysenck prefers, are extensively used. The Binet IQ test consists of an aggregate of items, testing primary mental abilities. On the basis of empirical observations these are grouped into age levels. It is felt that although some subjects excel at some types of items a general score can be allotted to an individual which will be indicative of, or a measure of, one psychological variable: his intelligence. The scoring system is as follows. A subject must begin (even if this entails working backwards through the items) by passing all the items at a particular age level. He then works through items at each age level, passing some items and failing others, until he reaches an age level at which he fails all items. He is then stopped. Even if he could pass items higher on the test these would not contribute to his total score.

The Binet test is generally accepted as an instrument for measuring intelligence. That is, it is generally accepted as being,

at least in one sense, unidimensional. It is also possible to show that the Binet test satisfies the criteria for unidimensionality set out by scale analysis. The scoring system must be altered. Age levels rather than individual items are passed or failed. Suppose age levels are dichotomized into pass or fail. A pass on one or more items constitutes a pass on the age level. A failure of all items at an age level constitutes a fail on the age level. The respondent then must pass all age levels until he fails one. Then he has failed all age levels beyond that point. If the maximum age level passed is known, his total distribution of passes and failures to all age levels is known: he passed all those below it and failed all those above it. Of course scale analysis cannot be used to test the unidimensionality of the data, as it was built into the scoring system, but it could be so used if the subjects were given the opportunity of passing or failing all age levels. The point is that this data is data which, presumably, Eysenck would not consider trivial, and it can fit the perfect scale model.

In any case, the hierarchical structure of attitudes is not accepted as fact by all writers. Guttman does not accept the definition of attitude as a construct to explain the consistencies of human behaviour. He begins with the observed behaviour and then, if it can be adequately as a function of one variable, infers that it is a unitary class of behaviour or delimiting totality of acts, or an attitude.

Eysenck also objects to the arbitrariness of the choice of 0.95 as an acceptable coefficient of reproducibility. He also feels the coefficient of reproducibility is of reduced value insofar as it is relative to the marginal frequencies and cannot be used to compare different scales. The chi square coefficient Guttman (1951) points out is similar to the coefficient of reproducibility in this respect. These coefficients cannot be used to compare different distributions: their proper function is to test a hypothesis about distributions separately. The chi square metric is in part a function of the marginals, like the coefficient of reproducibility and this does not reduce its usefulness. Of course the sampling distribution of chi square is known whereas that of the coefficient of reproducibility is not. That is the probability of obtaining a value of 0.95 for the coefficient of reproducibility by chance is not known. However, any set of data can be examined in terms of chi square to rule out the null hypothesis that the items are independent. Eysenck's criticisms of the coefficient of reproducibility are justified to some extent, and of course, he could not have known about the statistical tests of scalogram data as these were published by Schuessler in 1961.

Eysenck suggests that Guttman and his colleagues show

"a lack of critical awareness of the difference between data which are genuinely qualitative and data which are quantitative but presented in a qualitative form for the sake of convenience, or because of the peculiarities of the measuring instrument (1951, p.97)".

Eysenck seems to believe that a qualitative observation can be assumed to be an expression of a hypothetical underlying continuum. Guttman considers the qualitative data in their own right and examines the usefulness of regarding them as functions of an underlying continuum, and the degree of error involved in such a conception.

Eysenck adds (in a footnote):

"it is easy to show that the zero point, as defined by the second component, is not invariant (1951, p.99)".

Although a certain degree of scepticism about the further components of scale analysis, or at least of the value of giving these a psychological interpretation, has been expressed, Eysenck's comment is not of great value as he gives no evidence to support this assertion.

#### 8. Sampling the Universe of Content

Eysenck's most important criticism is one which is also made by Festinger (1947) and by Edwards and Kilpatrick (1948). This relates to the sampling of the universe of content to establish that the universe is scaleable. Guttman says

"The argument is essentially that a highly reproducible universe must yield highly reproducible samples - regardless of how the items are selected, randomly or not - and something like the ordinary theory of confidence intervals can be used to infer population characteristics from a sample, as is done in any sampling theory. The process of inference is not different in kind - though it does differ in detail - from the same process in other branches of statistical estimation (1951, p.111)".

TABLE 9

Methodological Note IV : Categories of the Scalogram Matrix of Data

1. Horizontal Categories

(a) Classes of responses (items).

Are assumed to be manifestations of the attribute such that they can be ordered with the relations  $>$ ,  $<$  or  $=$  in their possession or indication of the attribute.

(b) Elements (response categories).

Within a class responses can be ordered by  $>$ ,  $<$  but  $\neq$  in their indication of the attribute (approximately equal categories are combined).

2. Vertical Categories

(a) Elements (individual subjects).

Pairs of elements are ordered by the relation  $>$ ,  $<$  or  $=$ .

Individuals who are equal with respect to their response distributions, and hence in their possession of the attribute, are placed in the same class. This process generates:-

(b) Classes of subjects.

(i) Within classes elements are related according to the relation  $=$  in possession of the attribute.

(ii) Between classes the relation  $>$  or  $<$  holds and is generated by the relations between elements, with respect to how much of the attribute they possess.

TABLE 9 (continued)

If the items do not produce responses which are indicative of one attribute, the data do not scale and individuals cannot be ordered. Scale analysis is intended to show that only one attribute is involved, but it cannot show that the one the investigator chose to investigate is that which is being measured.

However the process does differ in kind because the size of the universe is not known, and its component items are in no sense equal units. He is assuming the two sets of categories which define the matrix of data do not differ in kind and can be selected in similar ways. However these do differ in respect to the properties that hold between them. For example item subcategories are inter-related by the fact that an individual may choose only one of these for each item whereas, until a scale is inferred to exist, categories of subjects are not related. This is set out more fully in Table 9.

It seems more useful to take a conventional approach to the observed data and test the possibility that the observed configuration is due to chance factors. If it is not it is possible to accept, with varying degrees of assurance, that the items are inter-related. Then it is possible to relate these items with other items, external criteria and so on to study their validity in terms of empirical research, rather than in terms of their relationship to a universe of content which cannot be defined.

## 9. Unidimensionality

Nunnally (1967) is also critical of scale analysis and it will be of value to discuss his criticisms so as to define more clearly the concept of unidimensionality.

In a perfect scale the probability of a response alpha where items are dichotomized is either 1.0 or zero. Up to a point the probability of response alpha is zero (and the probability of the

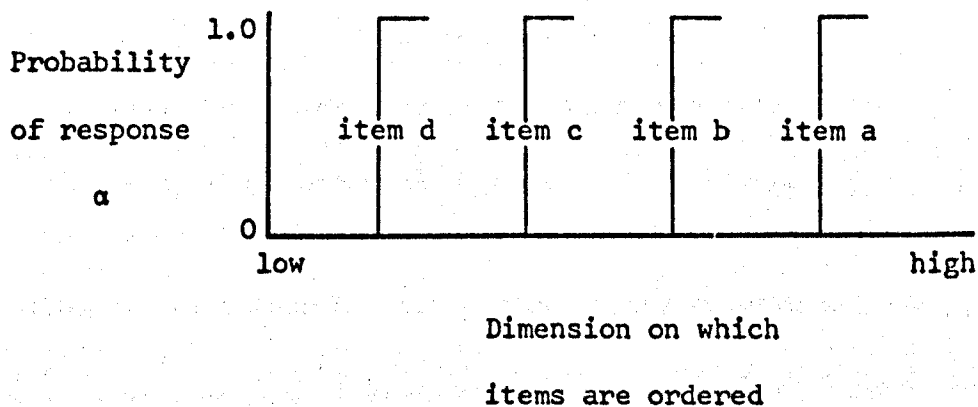
opposite response is 1.0), and beyond that point the probability of response alpha is 1.0. Each item has a biserial correlation of 1.0 with the attribute, though the correlations between items may vary, and consequently each item perfectly discriminates at a particular point of the attribute. Despite its intuitive appeal, this model is felt by Nunnally to be less useful than less deterministic models, such as those described by Lazarsfeld (1959), which allow for some error on the trace lines. His argument is that it is highly unrealistic to think that items should have trace lines such as those shown in Table 10. He suggests also that the observed scalogram pattern (which he calls a triangular pattern as he marks only positive responses) in the data does not necessarily indicate that the items will have these trace lines. He gives an example of a set of items which he thinks satisfies Guttman's scale criteria but which does not, he thinks, have these trace lines, and therefore he feels is not unidimensional. Although the choice between probabilistic and deterministic models is not really an either/or one (as the latter may be a special case of the former) and will depend on which fits the observed data or is more useful in a particular case, his example is of interest as it illustrates what is meant when a scale is said to be unidimensional. To minimize error the formal system which best fits the data is chosen to describe and perform operations on the data. Guttman's criteria for unidimensionality form a precise and simple scale model, but if the data do not fit the model there is no reason why another approach should not be tried.

TABLE 10

Methodological Note V : Trace Lines of Scale Models

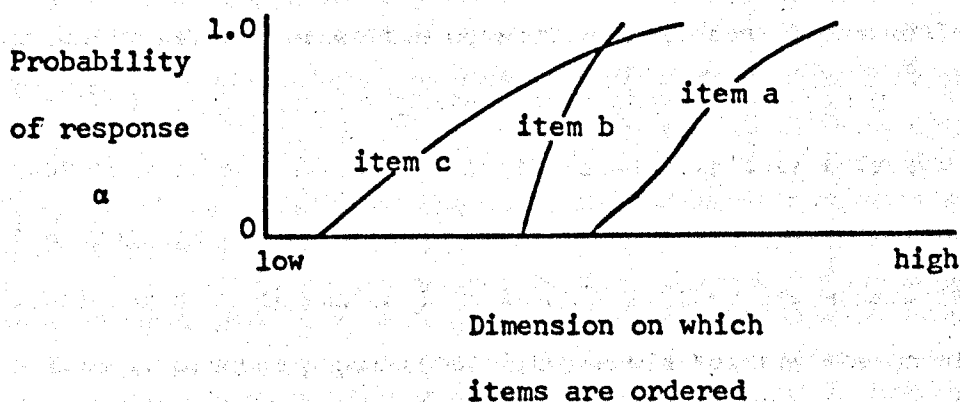
1. Trace Lines of Items on a Perfect Scale

Deterministic scale model



2. Trace Lines of Items on a Non-Perfect Scale

Probabilistic scale model



Trace lines may vary. There is some evidence that item a is, say, more difficult than items b and c but the items do not discriminate perfectly.

Nunnally gives the following example:

"If items are spaced far enough in difficulty (popularity on nonability items), the triangular pattern can be obtained even if the trace lines are very flat rather than vertical. This is illustrated in the following four items.

- a. Solve for  $x$  :  $x^2 + 2x + 9 = 16$
- b. What is the meaning of the word "severe"?
- c. How much is  $10 \times 38$ ?
- d. When do you use an umbrella? (given orally)

Although the author has not performed the experiment, the above four items administered to persons ranging in age from six to sixteen probably would form an excellent Guttman scale. Any person who got the first item correct probably could get the others correct.

Any person who failed the first item but got the second correct would probably get the other two correct. Those four items would produce the required triangular pattern even though there is good evidence that they do not all belong to the same attribute ("factor", in the language of factor analysis). The reason they apparently fit the model for a unidimensional scale is that they are administered to an extremely diverse population. They would not fit the model if they were investigated within one age group only (1967, p.65)".

Now, is the scale unidimensional or not? In relation to the population Nunnally describes the trace lines that can be constructed. The probability that a person who passes a also passes b, c and d is 1. The probability that a person who passes a fails b, c or d is 0.

This can be graphed as in Table 11. The trace lines, based on this population, fit the unidimensional model. The scale types can also be set out as in Table 11. Scale analysis indicates that the scale is, for this population, unidimensional, and the trace lines, for this population, are such that the probability of a response alpha is either 1.0 or zero and are vertical rather than flat. Scale analysis indicates that the scale is unidimensional and that a simple order can be generated between the individuals in the population according to their responses to the items. It indicates that the items are related according to some attribute, when responded to by this population. But scale analysis cannot name the attribute according to which the items and individuals are ordered any more than factor analysis does. Inspection of the items indicates that they are ordered according to difficulty, and hence the individuals are ordered according to their general ability to solve these items. A large number of abilities are required in solving any problem. First the subject must be able to see (or hear) before he can answer the problems. Given that subjects are equal in their possession of other attributes the items could be used, with other populations, to distinguish between those who are blind (pass item d, fail items a, b, c) and those who are not (pass items a, b, c, d), or those who are deaf (pass items a, b, c) and those who are not (pass items a, b, c, d) or those who speak English (pass items items a, b, c, d) and those who speak only other languages, or do not yet speak at all (a, b, c, d failed). Or we might use the scale to distinguish between those who have a knowledge

TABLE 11

Example V : Example of a Scale (Hypothetical)

Nunnally (1967) gives the following example of a series of questions which satisfy scale criteria. He contends that they are not unidimensional.

Children aged 6 to 16 years answer the following questions:

- (a) Solve for  $x$ :  $x^2 + 2x + 9 = 16$ ;
- (b) What is the meaning of the word 'severe'?
- (c) How much is  $10 \times 38$ ?
- (d) When do you use an umbrella? (given orally);

and their responses form a scale pattern.

1. Trace Lines of Items

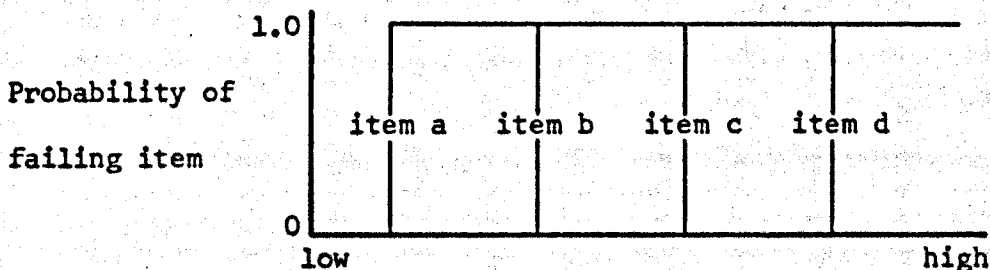
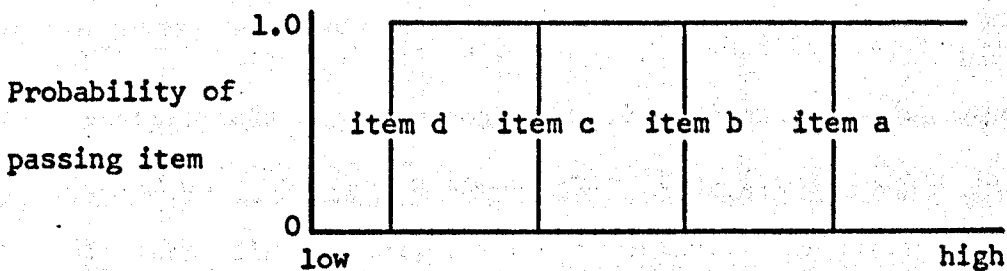


TABLE 11 (continued)

2. Scale Pattern of Responses

Items	Pass				Fail			
	a	b	c	d	a	b	c	d
Scale 1	x	x	x	x				
types 2		x	x	x	x			
3			x	x	x	x		
4				x	x	x	x	
5					x	x	x	x

3. The items do, then, form a scale and are, when responded to by these subjects, unidimensional. Nunnally suggests the items are not unidimensional because they have different factorial associations in other contexts. Factor analysis is, however, based on responses to items rather than items themselves and factors, like scales, are not invariant. The items are unidimensional in this case although they may not belong to the same factor, or scale, in another context.

of mathematical notation and arithmetical operations (pass items a and c) and those who do not (fail items a and c). The subject must be capable of responding. A child may know the answer to d, but not be capable of speaking the answer. If the question was rephrased "It is now raining. What do I need to keep the rain off? (whether answers or not) Go and get it for me" which presumably is as difficult in relation to how much knowledge of the relationship between rain and umbrellas is possessed, children who are capable of responding verbally could be distinguished from children who are not. Similarly subjects passing items a b and c must be able to write, so given that they are equal in their possession of other attributes we could use the items to distinguish the subjects who can write (pass items a, b, c) and subjects who cannot (pass item d, fail items a, b, c). That is, unidimensionality is not inherent in items, it is a relationship between items with respect to some attribute. It is not the items that impart unidimensionality: it is the responses to items which are indicative of some attribute. If the population of respondents is altered, the responses form a different set of data. What does Nunnally mean when he says the trace lines for these items are flat rather than vertical? When administered to whom? If they were administered to a population of foreign students, most of whom were unable to speak English, the trace lines would be very flat indeed, the probability of a pass possibly being lower on verbal items than on mathematical items as mathematical language is more universal. If they are administered to the described population the trace lines are vertical. If they were administered

to a population of deaf and blind subjects the trace lines might be completely flat, the probability of a pass being zero at every point of the scale.

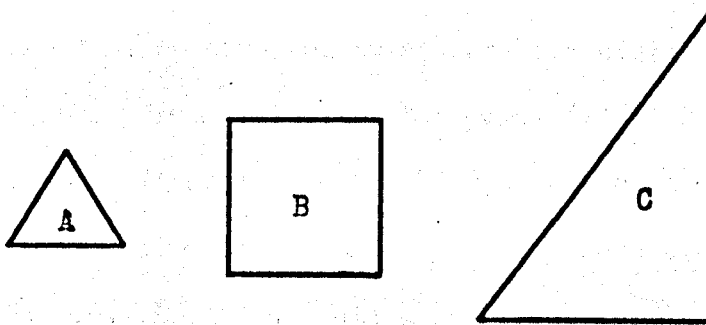
Nunnally's reference to factors is interesting as factors are invariant in exactly the same way as scales are. Factor analysis groups responses together and indicates that these can usefully be regarded as being of a unitary character, but it cannot indicate that the attribute it identifies is the one the investigator had in mind. For instance in one study item c may be grouped with other items of arithmetic as being indicative of arithmetical ability, and differentiated from verbal items. In a study of a more specific kind it might be grouped with other multiplication items, and differentiated from items involving division. In yet another more specific study it may be grouped with other items which involve multiplying by 10, and distinguished from items which involve more difficult multiplication such as multiplying by 7 or 9. The item itself has many attributes, or the response to it may be indicative of a number of attributes or abilities, and it is necessary to consider the context in which it occurs to identify what the response is indicative of. A simple way to conceive this is to remember that it is responses that are studied, and it is from the responses that unidimensionality is inferred to exist. Unidimensionality is a relationship between responses, and hence, it is inferred, of the stimuli which produce these responses. For verbal simplicity a set of stimuli is referred to as unidimensional, and this means that they may, or usually do, within a defined population,

TABLE 12

Methodological Note VI : The Meaning of Unidimensionality

Unidimensionality is not a property of objects, but a description of the relationship between them, with respect to some attribute.

For example, consider the following set of objects:



Each object has a large number of properties. The relationship between the objects cannot be described, until one property or attribute is selected.

1. In relation to the attribute height  $A < B < C$
2. In relation to the attribute width  $A < B < C$
3. In relation to the attribute area  $A < B < C$
4. In relation to the property circumference  $A < B < C$

These objects then form a ratio scale (in relation to these first four attributes).

5. In relation to the attribute size  $A < B < C$ . As there is not a unit of size this attribute is not always a useful standard of comparison. If, for example, shapes of equal area but different

TABLE 12 (continued)

height were included the set of objects could no longer be fully described in terms of one dimension. Two dimensions, such as width and height may be needed to adequately describe the objects.

In relation to the attribute size the objects form an ordinal scale.

6. The attribute location could be considered. For example, ratio scales could be constructed to compare the distances of the bases of the figures from the top of the page. ( $A < B < C$ ). Or the nearest point of each figure to the top of the page could be taken as a measure of location ( $A = B < C$ ). Other points of reference could be considered. Location could be measured from the side of the page, or both horizontal and vertical placement could be taken into account in a two dimensional reference scheme.

An indefinite number of nominal scales could be used to describe the objects. For example:

7. In relation to the attribute of being triangular  $A = C \neq B$ .
8. In relation to the attribute of having four equal sides  
 $B \neq C, B \neq A$ .
9. In relation to the attribute of being polygons  $A = B = C$ .
10. In relation to the attribute of being visual stimuli  $A = B = C$ .
11. In relation to the attribute of being non-circular  $A = B = C$

and so on. The objects have in common the attribute of not being an unlimited number of things, such as wave motions, books, household furniture.

TABLE 12 (continued)

Returning to the question of unidimensionality, it is obvious that the items are unidimensional in respect to some attributes, and not unidimensional in respect to others.

Items may be called unidimensional (with respect to some attribute) if they can be ordered on a dimension or continuum according to whether they possess more or less of this attribute.

The attribute 'size' provides a good analogy with psychological variables. It can be observed that one object has more, or less, size than another and objects can be ordered on a continuum. However, it may not always be possible to order objects according to size. One object may be taller while another is broader, and two dimensions will be needed to adequately describe the relationship between them.

In the same way an attribute such as 'intelligence' may adequately describe the relationship between people. However, if intelligence test items of a similar level are selected, and people of similar intelligence are considered, the unidimensionality may break down. The relationship between people will have to be expressed in terms of a number of dimensions, such as verbal skill, reasoning, sensori-motor coordination and so on.

These dimensions in turn break down when subjects and items are selected closer together on the continuum. For example closer study of verbal skill will show that some people are more fluent than others, who are in turn better at reading and comprehending verbal instruction.

TABLE 12 (continued)

In Example V (Table 11), it was suggested that items of varied content may form a scale, provided the items are widely spaced in difficulty. This scale measures "difficulty" (of items) and "ability to solve problems" (of subjects), in a very general way. The fact that each item could be used to measure something else does not invalidate the scale: the simplest relation that holds between them, under the described conditions, can be described in terms of one dimension.

Neither factor analysis nor scale analysis uncovers an absolute dimension. The idea of a property is itself an abstraction, to describe the similarities and differences between observed things. If a set of observed things can be described in terms of having more, or less of a property, they are unidimensional in respect to this property. However, if further objects are added to the set, the relationships between objects may not be able to be expressed in terms having more or less of this property.

Factors, and scales are not invariant because unidimensionality is an expression of the relation between objects rather than an inherent property of the objects themselves.

produce responses which can be ordered in a simple way according as to whether they possess more or less of some attribute. But the same stimuli, producing the same responses, can be ordered in other ways, in relation to other attributes which may or may not suggest unidimensionality.

#### 10. Evaluation of Scale Analysis

Nunnally also criticizes scale analysis because it produces an ordinal scale.

"If psychology were to settle only for ordinal measurement it would so limit the usable methods of mathematics that the science would be nearly crippled (1967, p.65)".

It must be agreed that Guttman scales must be dealt with at a low order of measurement, but the data evolve the most suitable or fitting level of measurement. It is not necessarily methodologically better to arbitrarily impose a stronger level of measurement on the data. It will then be possible to perform more sophisticated operations in the formal system, but these are at the formal level and may not be able to be applied to the data with validity.

The main advantage of scale analysis is that it does not impose an arbitrary system onto the data, but allows the data to unfold the most fitting or suitable level of measurement. It may suggest that responses to items form a unitary class of behaviour such that items, and people responding to them, can be ordered in relation to how much of the attribute is possessed, but it cannot show which

attribute is measured and the problem of validity remains.

The validity of the scale may be inferred from the content of the items, or evidence for validity may be established by means of further research. The outlook of the investigator will determine which approach he takes to establish the validity of the scale he constructs.

### CHAPTER III

#### THE SEMANTIC DIFFERENTIAL

##### 1. Description of the Semantic Differential

The semantic differential was developed by Osgood (1952, 1954, 1957) to measure the meaning of a concept. The subject is asked to encode, or make a series of successive judgements about, a concept in relation to a number of scales which are assumed to be representative of the major ways in which meaning can vary.

Osgood et al. (1957) used concepts such as ME, MOTHER, GOD, BABY, THE NEGRO and CAPITAL PUNISHMENT, but any stimulus could be successively rated on a number of dimensions in this way. The scales or dimensions are usually presented as pairs of bipolar adjectives such as good/bad, weak/strong, happy/sad with seven intervals between them. For example:

good    (1)    (2)    (3)    (4)    (5)    (6)    (7)    bad

The subject is asked to check interval (1) if he feels the concept is very closely related to that end of the scale that is defined by the adjective "good", interval (2) if he feels that the concept is quite closely related, and interval (3) if he feels that the concept is only slightly related to that end of the scale that is defined by the adjective "good". If the subject considers that the concept is neutral



That is, the positive intervals may not be equal to the negative intervals. Similarly the subject may be tempted to check both a positive and negative interval, meaning "I would like to have one of these (provided I didn't have to buy it myself)" and "I would not like to have one of these (if I had to buy it myself)". That is, the scale might be non-linear. If the subject were allowed to check more than one interval a continuum could be inferred (or not inferred) from the data, as for example by Coombs (1953) pick 2 method, but as the subject must check only one interval the assumption of linearity is built into the scale a priori. Osgood feels that familiar and common adjectival opposites do define linear continua with equal intervals, and a neutral or zero midpoint.

The use of seven intervals has its basis in empirical findings. Osgood et al. say:

"Over a large number of different subjects in many different experiments it has been found that with seven alternatives all of them tend to be used with roughly, if not exactly, equal frequencies (1957, p.85)".

Stagner and Osgood (1946) used scales with 9 and 5 intervals. College students used as subjects tended to use the discriminative intervals (i.e. intervals 2 3 4, 6 7 8) less frequently than the polar and neutral intervals (1, 5, 9) when 9 intervals were used, and expressed irritation at not being able to discriminate when only 5 intervals were used. However it is probable that some types of people would find a lesser number of intervals less confusing as this simplifies

the task. Osgood et al. (1957) for example suggest that children seem to work better with a five step scale. However, the criterion set for deciding the optimum number of scale intervals (that is, that intervals should be selected such that they tend to be used with roughly, if not exactly, equal frequencies) is rather vague. The frequency with which a particular interval is used is related to the concept being differentiated and the people who are encoding the concept or sign. For example if the concepts chosen are NAZI, SIN, EVIL and DICTATORS a high frequency of responses on the negative intervals can be expected. If the concepts chosen are MOTHER, ICE CREAM, BABY and HAPPINESS a high frequency of responses on the positive intervals can be expected. The problem of sampling concepts, to study the distribution of responses, is similar to that discussed by Guttman (1950) as the problem of sampling the universe of content. A list of all the concepts that could have been used cannot be constructed, and even if it could it could not be sampled in the way a population of people is sampled as the units are in no sense equal. For instance it is difficult to tell if the concepts A NAZI, THE NAZI, NAZI are the same concept, or different concepts, without measuring them by means of the semantic differential (which is circular). Similarly the frequency of responses on a particular interval will be related to the group of people differentiating the concept. For instance, many people were Nazis, and had they been chosen as subjects, NAZI chosen as a concept would probably have produced many positive responses. This is less of a problem than the sampling of

concepts as populations of people can be defined, listed (at least in theory), and representative samples drawn out and studied with a defined margin of error. However many of the studies quoted by Osgood et al. (1957) use college students as subjects, and the findings cannot necessarily be generalized as being applicable to other populations.

Osgood et al. (1957) postulate a semantic space, Euclidean in character, and of unknown dimensionality. Each bipolar scale is assumed to represent a straight line function that passes through the origin of this space. Research, for example the study by Stagner and Osgood (1946), indicated that certain scales are correlated and that their replication adds little to the definition of semantic space. For maximum efficiency, then, a least number of orthogonal (or uncorrelated) dimensions which exhausts the dimensionality of the space is sought to define that space. In practice, factor analysis is applied to the scales, and those dimensions which can be reliably identified are used to define the semantic space. For example, factor analysis may establish two orthogonal factors of which the scales good-bad and strong-weak are most representative, and the concepts can be plotted in a two dimensional space, as in Table 13. The location of a concept in a semantic space can be represented by means of solid geometrical models if three dimensions are needed to describe the space, or, if more dimensions are needed to describe the space, algebra must be used.

TABLE 13

A subject differentiates the concepts FATHER, MOTHER and BABY on the scales good/bad, happy/sad, kind/cruel, weak/strong, active/passive and quick/slow.

A concept that is rated "good", is also rated "happy" and "kind" in related degrees. A concept that is rated "strong" is also rated "active" and "quick" in related degrees. Factor analysis will then define two (rather than the original six) dimensions.

The results may be as follows:

FATHER							MOTHER									
good	<u>X</u>	-	-	-	-	-	bad	good	<u>X</u>	-	-	-	-	-	-	bad
happy	<u>X</u>	-	-	-	-	-	sad	happy	<u>X</u>	-	-	-	-	-	-	sad
kind	<u>X</u>	-	-	-	-	-	cruel	kind	<u>X</u>	-	-	-	-	-	-	cruel
weak	-	-	-	-	-	<u>X</u>	strong	weak	-	-	-	-	-	<u>X</u>	-	strong
active	<u>X</u>	-	-	-	-	-	passive	active	-	<u>X</u>	-	-	-	-	-	passive
quick	<u>X</u>	-	-	-	-	-	slow	quick	-	<u>X</u>	-	-	-	-	-	slow

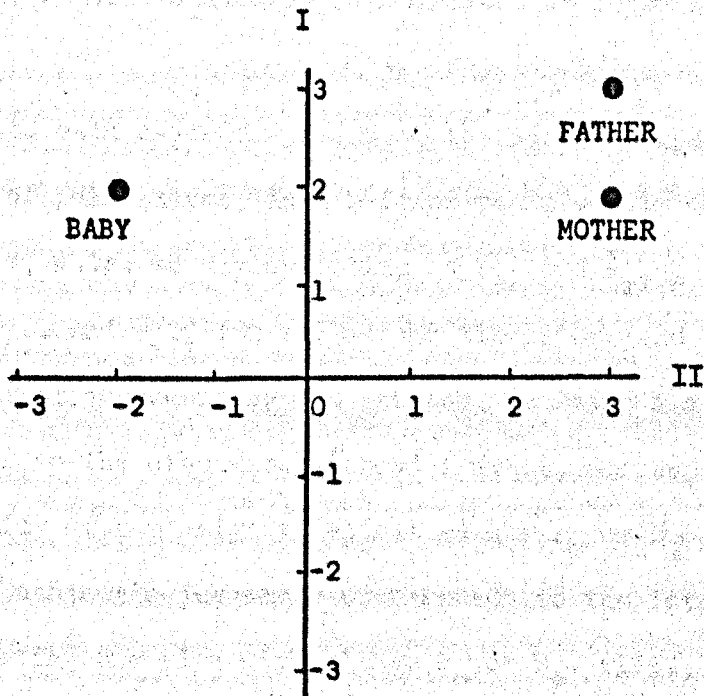
BABY							
good	<u>X</u>	-	-	-	-	-	bad
happy	<u>X</u>	-	-	-	-	-	sad
kind	<u>X</u>	-	-	-	-	-	cruel
weak	<u>X</u>	-	-	-	-	-	strong
active	-	-	-	-	-	<u>X</u>	passive
quick	-	-	-	-	-	<u>X</u>	slow

TABLE 13 (continued)

Response intervals can be scored +3 to -3. The positive score represents the good, happy, kind; strong, active quick poles. The data can then be summarized, as follows, in terms of the mean factor scores:

		Concept		
		FATHER	MOTHER	BABY
Dimension	I	+3	+3	+2
	II	+3	+2	-2

and plotted in a semantic space defined by the two dimensions:



Osgood et al. (1957) try to establish that there are invariant factors which define a semantic space such that the location of a concept in the semantic space serves as an operational definition of the meaning of the concept. A point in the space, which represents the location of the concept, has two properties: direction and distance from the origin. These are identified with the quality (direction) and intensity (distance) of psychological meaning. The direction from the origin is thought to depend on the alternative polar terms selected, and the distance is thought to depend on the extremity of the scale positions which were checked.

## 2. Osgood's Learning Theory

The use of bipolar scales is rationalized in terms of Osgood's learning theory.

"Corresponding to each major dimension of the semantic space, defined by a pair of polar terms, is a pair of reciprocally antagonistic mediating reactions, which we may symbolize as  $r_{mI}$  and  $\bar{r}_{mI}$ , for the first dimension,  $r_{mII}$  and  $\bar{r}_{mII}$  for the second dimension, and so forth. Each successive act of judgement by the subject, in which a sign is allocated to one or to the other direction of a scale, corresponds to the acquired capacity of that sign to elicit either  $r_m$  or  $\bar{r}_m$  and the extremeness of the subject's judgement corresponds to the intensity of reaction associating the sign with either  $r_m$  or  $\bar{r}_m$ . (1957, p.27)".

At the lowest level of behaviour there is a certain stimulus patterns which produce certain sequences of responses. Other patterns

of stimuli have, by constant pairing with the original stimuli, acquired this capacity. For instance, if the sound of a buzzer is constantly followed by the taste of food, the sound of the buzzer alone may initiate digestive processes appropriate to the taste of food. That is, the buzzer comes to be a sign, which acquires its meaning from its association with the food, or the original pattern of stimulation, which Osgood terms the significate. The significate is

"any stimulus which, in a given situation, regularly and reliably produces a predictable pattern of behaviour (1957, p.6)".

The predictable pattern of behaviour may be innate or acquired.

"Whenever some stimulus other than the significate is contiguous with the significate, it will acquire an increment of association with some portion of the total behaviour elicited by the significate as a representational mediational process (1957, p.6)".

A sign is defined as follows:

"A pattern of stimulation which is not the significate is a sign of the significate if it evokes in the organism a mediating process, this process (a) being some fractional part of the total behaviour elicited by the significate and (b) producing responses which would not occur without the previous contiguity of non-significate and significate patterns of stimulation (1957, p.7)".

The word SPIDER is an example of the type of sign which may be studied by the semantic differential. The stimulus object or

significate  $\dot{S}$  is

"the visual pattern of hairy legged insect body often encountered in a threat context provided by other humans (1952, p.205)".

The stimulus elicits a complex pattern of behaviour  $R_T$ , which in this instance includes a heavy loading of autonomic fear activity for some individuals. Portions of this total behaviour to the spider become conditioned to the word SPIDER, embodied in an  $r_m$ . As the sign is used in isolation from the actual object the mediation process becomes minimal, but still includes these autonomic reactions. The  $r_m$  produces a distinctive pattern of self stimulation  $s_m$  which may elicit a variety of overt behaviours  $R_X$ . These might, for example, be

"shivering and saying "ugh", running out of the room where a spider is said to be lurking, and even refusing a job in the South, which is said to abound in spiders (1952, p.205)".

There are a number of signs associated with a significatesuch as the spider. For example, the visual stimulus of the word SPIDER, the sound of the same word spoken, and pictorial representations of spiders are all associated with the original object. It is not necessary to actually encounter a spider to assign meaning to these signs, as the meaning of a sign may be learnt from other signs. Few men have, for example, seen the craters of the moon, but meaning can be given to these words by pictures and words, or signs.

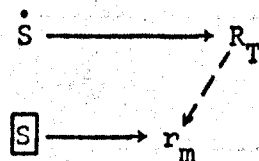
The most important feature of Osgood's definition of meaning in terms of representational mediational processes is that although it is behavioural it is not necessarily in terms of observable things.

TABLE 14

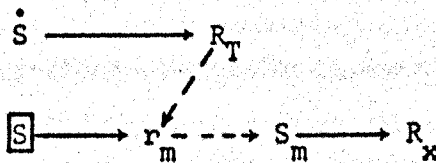
Methodological Note VII : Paradigm of the Sign Process

1. The Sign

The sign  $\boxed{S}$ , through constant association with its significate  $\dot{S}$ , comes to elicit behaviour  $r_m$  which is representative of the total behavioural reaction  $R_T$  to the object which is the significate.



But the organism is not a passive receptacle. The end result of the sign process is instrumental or goal directed action which takes account of the significate. The behaviour  $r_m$  is, Osgood suggests, self stimulating to produce these instrumental acts,  $R_x$ .



2. The Assign

A series of signs  $S_1$ ,  $S_2$ ,  $S_n$  have meaning, and so can be represented

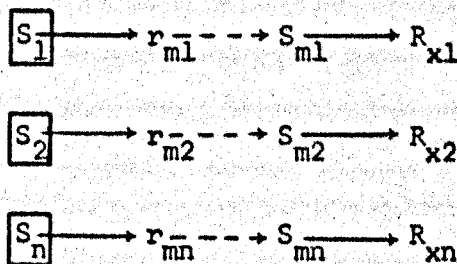
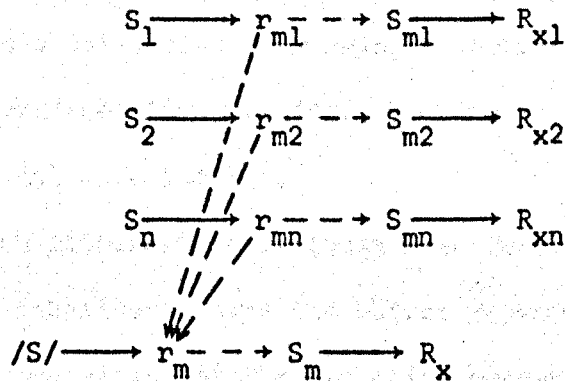


TABLE 14 (continued)

Another sign, /S/, can acquire meaning through being associated with these. .



Source. Osgood et al. The Measurement of Meaning. Urbana, Ill.:  
University of Illinois Press, 1957, p.7.

Mediational processes must be inferred from the instrumental acts which take account of the significate. The semantic differential elicits responses  $R_x$  from individuals, and it is hoped to infer from these the nature of the underlying processes.

Also, Osgood's definition of meaning is concerned with conotative rather than denotative meaning. Meaning is individual and contingent upon personal associations.

"The meanings which different individuals have for the same signs will vary with their behaviour toward the object represented. This is because the composition of the mediation process, which is the meaning of a sign, is entirely dependent upon the composition of the total behaviour occurring while the sign process is being established (1952, p.206)".

### 3. Definition of Meaning

Osgood, then, defines meaning in two ways. Operationally meaning is defined as a point in semantic space. Theoretically meaning is defined as a representational mediational process which relates the observed behaviour produced by a sign to the original significate. The usefulness of the theoretical definition is open to doubt, as is its relationship to the operational definition. The speculative isomorphism between the two systems rests on a number of assumptions. Harper (1963) suggests the following conditions must hold.

1. "that there is a finite number of representational mediating reactions available to the individual (p.38)".

This is a reasonable assumption in that the experience of individuals is finite, but the number of mediating reactions available to the individual may be indefinitely large.

2. "that the number corresponds to the number of dimensions or factors in the semantic space (p.38)".

This is a crucial assumption; that the mediational processes can in fact be represented as dimensions in a semantic space.

It is impossible to prove, as the mediational processes are inferred from the semantic differential data, and the structure of the semantic differential is based on certain assumptions about the nature of the mediational processes.

3. "that linguistic quantifiers used to identify scale positions have been associated with more or less equal degrees of intensity for all ( $r_m - s_m$ ) processes (p.38)".

That is, that extremely hot represents the same degrees of hotness as extremely sweet does of sweetness. It is also assumed that the intervals are equal. That is, that the distances between scale intervals are the same within and between scales, and that the central point of all these scales coincides, so that they can be located with relation to one another in a common space.

Fortunately the empirical validity of the semantic differential is not dependent on its connection with Osgood's learning theory, but on the evidence of research. His theories about the

dimensionality of semantic space can also be held in abeyance until further research findings have been considered. For the moment the meaning of a concept will be defined as the responses it produces or ratings it receives on a set of semantic differential scales. The data which serve to define a concept are relative to the scales chosen and to the people responding.

### The Dimensionality of Semantic Space

Osgood et al. (1957) describe a number of factor analytic studies, the purpose of which is to

"discover the 'natural' dimensionality of the semantic space, the system of factors which together account for the variance in meaningful judgements (p.31)".

There are three sources of variance: subjects, scales and concepts used. For the major factorial studies college undergraduates are the subjects. The sampling of scales is a more elaborate procedure, as it is the dimensionality of the scale system which is of interest. Scales are sampled in terms of external criteria: the frequency of usage of certain descriptive terms, and subsequently Roget's Thesaurus were used to select terms which defined scales. The sampling of concepts is less clearly defined, but in one study concepts were eliminated entirely.

In the first study 40 nouns were taken from the Kent-Rosanoff list of stimulus words for free association. 200 undergraduate students were asked to write down the first descriptive adjective that occurred

to them in response to these. These data were then analyzed for the frequency of occurrence of all adjectives. The 50 most frequently appearing adjectives were selected and made into sets of polar opposites.

"As might be expected the adjectives good and bad occurred with frequencies more than double those of any other adjectives. Perhaps less expected was the fact that nearly half of the 50 most frequently appearing adjectives were also clearly evaluative in nature (p.33)".

Three further scales, pungent/bland, fragrant/foul and bright/dark were inserted to give a total of 50 pairs of polar opposites. That is, although the adjectives were on the whole selected empirically, the selection of opposites and matching of pairs such as "good" and "bad" was decided upon by the investigators on other criteria. The bipolarity was built in to the semantic space a priori.

The sampling of concepts was also on a priori grounds. The criteria used were that firstly, they should not be the same nouns that had been used to select the sample of adjectives, secondly, that they should be familiar to the subjects, and thirdly, that they should be as diversified in meaning as possible so as to augment the total variability in judgements. The concepts LADY, BOULDER, SIN, FATHER, LAKE, SYMPHONY, RUSSIAN, FEATHER, ME, FIRE, BABY, FRAUD, GOD, PATRIOT, TORNADO, SWORD, MOTHER, STATUE, COP, AMERICA were used. The semantic differential was set out in the following form

LADY	good	-	-	-	-	-	-	-	bad
SIN	clean	-	-	-	-	-	-	-	dirty
BOULDER	poor	-	-	-	-	-	-	-	rich

and so on until every concept had been matched with every scale (1000 items). Forty random items were repeated as a reliability check. 100 students were used as subjects. Thus the study generated a 50 (scales) x 20 (concepts) x 100 (subjects) cube of data. Responses were assigned a number from 1 to 7 to indicate the scale interval chosen. Values were summed over subjects and concepts to give a set of generalized scale values, and a matrix of the 50 x 50 intercorrelations of the scales was computed. Thurstone's Centroid Factor Method (1967) was applied to this matrix of intercorrelations. Four factors were extracted and rotated into simple structure, maintaining orthogonality. The fourth factor accounted for less than two percent of the variance, and is interpreted by the authors as being a residual.

The first factor is labelled an evaluative factor by the authors, as it has high loadings on the scales good/bad, beautiful/ugly, sweet/sour, clean/dirty, valuable/worthless and so on. These scales are interpreted as being "purely" evaluative in the sense that the extracted variance is almost entirely on this factor.

The second factor is identified as a potency factor. The scales which have the highest and most restricted loadings are large/small, heavy/light and thick/thin. Other scales have considerable evaluative meaning as well: hard/soft, loud/soft, deep/shallow,

brave/cowardly are instances of scales with loadings on both factors.

The third factor is identified, by the authors, as an activity factor. The most distinctively loaded scales being fast/slow, active/passive and hot/cold.

The evaluation factor accounts for 68.55% of the common (extracted) variance.

The second study was designed to check the possible relativity of the factor structure to the sample of concepts used. The same pairs of adjectives were used, and 40 subjects were drawn from the same undergraduate college population. Osgood et al. (1957) describe the method as follows:

"The method used involves a forced choice between pairs of polar terms as to the direction of their relationship. Given the following item, for example

SHARP - dull; relaxed - tense.

The subject is asked to simply encircle that one of the second pair which seems closest in meaning to the capitalized member of the first pair. There is no restriction on the concept (if any) that may be used. Some subjects might think of "people" concepts, others of "object" concepts, and yet others of "aesthetic" concepts. Introspectively (and as judged from the comments of subjects) there is usually no particular concept involved. If 100 percent of the subjects select "tense" as might happen in this case, it would indicate that sharp-with-tense vs. dull-with-relaxed is an appropriate parallelism or association

over concepts in general; if subjects divide randomly (e.g. half one way, half the other) on an item such as

FRESH - stale; long - short

it would appear that either the multitude of conceptual contexts in which these qualities might be related are random with respect to direction or that subjects differ randomly in their judgements of the relation - in either case, no particular concept or set of concepts is forcing the direction of relation (p.40)".

Every scale was compared with every other scale and a symmetric matrix of percentages, analogous to the usual matrix of inter-correlations was tabulated. It is argued that if two scales are equivalent their percentages of agreement with all the other scales will be equivalent. The matrix was factored by the diagonal method, as given by Thurstone (1947), selecting any one of the scales as a pivot to begin with, and continuing factoring until the residuals are zero. The method is called the D method as it can be applied to raw score matrices of semantic differential data, and when this is done the distances (D) between variables can be reproduced. Osgood's description of the method is given, in full, in Appendix II.

Five dimensions were extracted and rotated graphically, maintaining orthogonality between dimensions, to maximize the similarity between this structure and that obtained with the centroid method in the previous study. Osgood et al. (1957) conclude from the results:

"There is thus no question about identification of the first dimension of the semantic space - an evaluative factor is first in magnitude and order of appearance in both analyses (p.44)."

The potency determinant, however, gives a poorer correspondence, but Osgood et al. feel the evidence of correspondence is fairly satisfactory. The three variables having the highest coordinates on dimension II, strong/weak, large/small and heavy/light are the three variables having the highest loadings on factor II.

The third dimension is similarly identified with factor III and the variables sharp/dull, active/passive and fast/slow have high coordinates on dimension III and high loadings on factor III. The fourth and fifth, and the other dimensions which could have been extracted, are not discussed by the authors, but they are aware of the possibility of further specific factors. The original procedure for sampling scales resulted in a large number of evaluative terms, which made the number of other types of terms relatively small, too small to permit other types of additional factors to appear clearly.

Roget's Thesaurus (1941 edition) was used as a source of adjectives.

"The task set by Roget and his subsequent editors was precisely to provide a logically exhaustive classification of word meanings, and this source had the added advantage that most categories were already arranged in terms of polar opposition (p.48)."

The most familiar and representative pair of terms was selected from each adjective list. This resulted in a list of 289

adjective pairs. This list had to be reduced, or broken up into groups for separate analyses, if the computer was to be able to handle the results. The authors decided on the former approach. A group of 18 students from an advanced class in advertising copy writing was asked to sort the terms, written on cards, into 17 piles in terms of similarity of meaning. No restrictions as to the size of the piles or the criterion of meaning to be used was placed on the subjects. The grouping together of two pairs of terms, by five subjects was used as a criteria of similarity, as this is significant at the 1 percent level. Of similar terms, the least familiar, and least bipolar in appearance were rejected. This left 105 pairs of adjectives. An additional 29 pairs were also discarded by the authors, using the same criteria, to leave 76 terms, representing the number of variables that could be handled by the computer.

100 students acted as subjects, and 20 concepts were chosen to represent a variety of the types of concepts which could be used with the semantic differential. Five of the concepts were the same as those used in the first study. The concepts represented various categories. Personal concepts, for example, MY MOTHER, FOREIGNER, physical objects, for example, KNIFE, BOULDER, abstract concepts, such as MODERN ART, TIME, event concepts, such as, BIRTH, SYMPHONY and institutions such as, HOSPITAL and AMERICA were sampled.

Concepts were not rotated against scales, as before, for although this is methodologically more sound and helps to guarantee independent judgements on each scale, it is less satisfactory from the

subject's point of view than making a series of judgements while the one concept is held in mind.

A centroid factor analysis extracted eight factors.

Factoring was then discontinued as the eighth factor accounted for only about 1 percent of the variance. The three first factors could be identified with the evaluative, potency and activity factors.

The centroid structure was rotated "blind". This Quartimax Rotation (from Neuhaus and Wrigley, 1954) seemed to retain the original factors, although these appeared in a different order. Factor I was still the dominant evaluative factor but other factors were less clearly defined and identified from the terms which generated the original continua.

The same data was then analyzed by the Square Root method of factoring (from Wrigley and McQuitty, 1953). The pivotal scales for the first three factors were arbitrarily selected: good/bad, hard/soft and active/passive were selected as the first, second and third pivots. Factoring was discontinued after eight factors had been extracted.

Osgood et al. (1957) quote two other studies in which non-student populations were sampled. Solomon (1954) had trained Navy sonar men, rate sonar signals, and Tucker (1955) had groups of artists and non-artists rate paintings against adjectival scales. Both studies identified evaluative, potency and activity factors. Osgood et al. (1957) conclude from these studies that there is a semantic space with a generalized factor structure:

"To test the generality of the factor structure obtained, we have in our several studies (a) varied the subject populations, (b) varied the concepts judged (and in one case eliminated specific concepts entirely), (c) varied the type of judgmental situation used in collecting data, and (d) varied the factoring method used in treating data. Since the same primary factors keep reappearing despite these modifications, we conclude that the factor structure operating in meaningful judgments is not dependent upon these variables, at least. (p.33)".

#### Osgood's Evaluation of the Semantic Differential

The criteria set down by Osgood et al. (1957) for evaluating a measuring instrument such as the semantic differential are objectivity, reliability, validity, sensitivity, comparability and utility.

The semantic differential, as usually administered, is objective insofar as the results obtained with it are not in any way dependent on the idiosyncrasies of the observer, although, of course, the interpretation of results might vary from investigation to investigation.

Osgood et al. (1957) define reliability as

"the degree to which the same scores can be reproduced when the same objects are measured repeatedly (p.126)".

As part of the first study 40 items were repeated, and the test and retest scores were correlated, to give a coefficient of 0.85.

The authors study the test-retest reliability over time intervals ranging from a few minutes up to three weeks and suggest that the average error of measurement is less than one scale unit. The evaluative factor score was judged slightly more reliable than other factor scores, and some concepts showed a greater variation, from test to retest, than others. The nature of these concepts, such as MY MOOD TODAY, suggested that the difference was due to an instability in the concept itself, rather than the instability in the test (such a difference, in fact, attests to the sensitivity of the instrument).

The validity of the semantic differential is more difficult to establish as there is not a commonly accepted criterion of meaning. Validity is usually established by correlating the results of a new test with those of an older accepted test, or some external standard such as school success in the case of intellectual tests. Osgood concentrates on face validity. Face validity is really the 'plausibility' of a test. WHITE ROSEBUDS, GENTLENESS and SLEEP tend to cluster together in a semantic space determined by the differential. This is what association technique or commonsense would suggest.

Osgood and Luria (1954) interpreted the semantic differential scores obtained from the differentiation of a number of concepts by the three personalities of Eve White. They achieved considerable agreement with the descriptions of her personalities provided by her psychiatrists.

Osgood et al. (1957) also report a study by Dr. A. Solarz, conducted at the University of Illinois in 1953 in which subjects learn

to associate a set of verbal signs with either approach or avoidance movements. The verbal signs were evaluative adjectives. Subjects found it easier to learn to move a word like "sweet" towards them, than away from them, or to learn to move a word like "sour" away from them than towards them. That is, the negative evaluative terms were easily associated with an avoidance reaction, and the positive evaluative terms were easily associated with an approach reaction, but learning took longer if the direction of the evaluative term and reaction required conflicted. Other experiments related extremeness of judgement to the time taken to respond to items.

The semantic differential was also used to predict the voting behaviour of 'don't know' voters in the 1952 Presidential election. Of 18 subjects who were undecided as to how they would vote, the vote of 14 was correctly predicted from whether their meaning profiles for a set of concepts clustered with the profiles of Stevenson supporters or with the profiles of Eisenhower supporters.

The sensitivity of the semantic differential is illustrated by its ability to distinguish between concepts such as GOOD and NICE.

The comparability of semantic differential data across subjects is, Osgood et al. (1957) feel, shown in cross cultural studies. These indicate the same factors of judgement which appear in the English language may operate when Japanese and Korean subjects work in their own languages.

The comparability of concepts is less straight forward. Ideally this requires that individual scales, or less strictly at least

factor scores, should maintain constant intercorrelations regardless of the concept being judged. This is not so in the case of individual scales; strong may equal good in evaluating HERO, but not when differentiating SYMPHONY. There is also a shift of meaning in regard to factor scores. An evaluative factor was found in each of the concepts used in the first study, when concepts were analyzed separately, but scales contributing to it were found to vary from concept to concept. The factor structure may alter if single concepts are studied. For instance, the evaluative and potency factors may be correlated in judging LEADER. Results suggest that the more evaluative a concept, for example SIN or FATHER compared to less evaluative concepts such as METHODOLOGY or BOULDER, the more the meaning of other scales tends to shift towards the evaluative factor.

If Osgood's theory of the generality of semantic space is accepted, the utility of the semantic differential as a measuring instrument is great. Any type of stimuli can be rated on semantic differential scales, and related to one another by means of the ratings given. A symphony could be compared to a practical joke in terms of its psychological meaning. A framework of judgement is provided in which any stimuli can be placed: political issues, products, people, programmes, semantic relationships, sights, sounds, aesthetic objects, or anything which can be seen, heard, smelt or felt, or symbolized in some way such that the symbol can be a stimulus. Further, this framework of judgement is thought to be cross-cultural. As well as this, people responding to stimuli can be studied. The meanings people attach to

stimuli might be indicative of their political affiliations or personal adjustment. However, in comparing the semantic differential with Guttman's scale analysis, it is hoped to show that the semantic differential assumes certain metric properties, and that the idea of a semantic space is built into the data a priori, rather than necessarily having a psychological basis, reflecting a unitary frame of judgement which people apply when differentiating the meaning of concepts.

## CHAPTER IV

### THE DIMENSIONALITY OF DATA

#### 1. Levels of Measurement used by the Semantic Differential and Scale Analysis

Social and psychological measurement involves both collecting data, and ordering data. Scale analysis is primarily concerned with establishing procedures for ordering data. The semantic differential is concerned with collecting and ordering data, and certain properties of the data are assumed to hold, contingent upon the method of collection. As both scale analysis and the semantic differential try to describe the configuration of responses in terms of one or more dimensions it may be of value to compare the aims of these two methods in terms of the way in which each method orders responses.

The data are the responses of individuals to each stimulus. In scale analysis the stimuli would, as a rule, be attitude statements such as "The government is too powerful" or "The young people of today do not know the value of money". Responses might be "agree", "disagree" and so on. In the semantic differential a concept such as THE GOVERNMENT or YOUNG PEOPLE might be evaluated, and responses would take the form of checking a scale interval indicating an association with powerful rather than powerless, bad rather than good and so on.

A matrix of the data obtained can be constructed as in Table 15.  $a_1$  is the response of subject or person 1 to stimulus A.

TABLE 15

Methodological Note VIII : The Matrix of Data in Scalogram Analysis and the Semantic Differential

	Stimuli		
	A	B	C
Subjects: 1	$a_1$	$b_1$	$c_1$
2	$a_2$	$b_2$	$c_2$
3	$a_3$	$b_3$	$c_3$

The matrix of data shows the responses of subjects 1, 2 and 3 to the stimuli A, B and C. The response of subject 1 to stimulus A is denoted  $a_1$ , the response of subject 2 to stimulus B is denoted  $b_2$  and so on.

Scale Analysis may indicate that the response distributions are such that  $a_1 b_1 c_1 > a_2 b_2 c_2 > a_3 b_3 c_3$ , and that  $a_1 a_2 a_3 > b_1 b_2 b_3 > c_1 c_2 c_3$  in relation to an attribute. This implies that  $A > B > C$ , and that  $1 > 2 > 3$ .

The Semantic Differential assumes values can be assigned to responses.

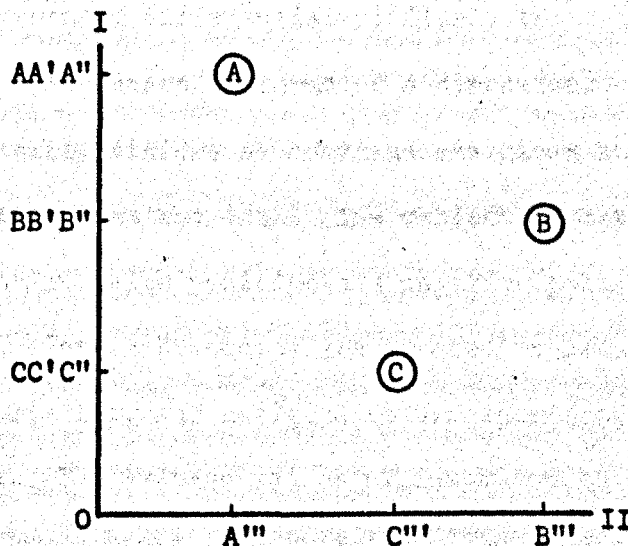
The values of  $a_1 + a_2 + a_3$ ,  $b_1 + b_2 + b_3$  and  $c_1 + c_2 + c_3$  locate the stimuli A, B and C on a continuum. The values of  $a_1 + b_1 + c_1$ ,  $a_2 + b_2 + c_2$  and  $a_3 + b_3 + c_3$  locate the subjects 1, 2 and 3 on a continuum.

Responses to further related stimuli A'B'C', A''B''C'' and A'''B'''C''' are then obtained. By means of the ordering of responses, the stimuli can be located on continua. For example:

TABLE 15 (continued)

A	A'	A''	B'''
B	B'	B''	C'''
C	C'	C''	A'''
0	0	0	0

Now, the first three continua are replications, so they can be represented by one dimension (I). A second continuum (II) is needed to describe the stimuli A'''B'''C'''. If ABC, A'B'C' and A''B''C'' and A''' B''' C''' are regarded as being facets of the concepts (A), (B) and (C), these concepts can be described economically in terms of two continua. Because a common zero is assumed the two continua can be related and the concepts can be related in a two-dimensional space.



Scale analysis can be used to indicate if the response distributions can usefully be considered as being ordered according to some variable, or along some continuum. The response distribution  $(a_1 \ b_1 \ c_1)$  may, for example, appear to be greater than the response distribution  $(a_2 \ b_2 \ c_2)$  with respect to some attribute. For example scale analysis may order the response distributions  $(a_1 \ b_1 \ c_1)$ ,  $(a_2 \ b_2 \ c_2)$ ,  $(a_3 \ b_3 \ c_3)$  and suggest that  $(a_1 \ b_1 \ c_1) < (a_2 \ b_2 \ c_2) < (a_3 \ b_3 \ c_3)$ . These can then be represented on a continuum. This also orders the stimuli  $A < B < C$  and the subjects  $1 < 2 < 3$ . The data can be described as being generated by one attribute: the stimuli are said to possess the attribute, the responses are said to be indicative of it, and the subjects are said to be ordered according to it. The continuum is evolved from the data. There is no attempt to map the data onto a defined continuum, on which the unit of measurement and the location of the zero is known.

The semantic differential similarly tries to describe the configuration of responses in terms of a dimension. The criterion of unidimensionality differs as continua are given a priori, rather than being evolved from the data. The subject places his responses on built in continua with positive and negative poles, defined intervals and neutral midpoints. Responses are treated as being functions of quantitative variables, and assigned values according to their location on the continua. The average values of responses to a stimulus serves to locate, and hence order, the stimulus. The average value of responses given by an individual to the series of stimuli, serves to

locate, and hence order, him in relation to other responding individuals. This is quite simple if only one continuum (or scale) is used, but the semantic differential tries to measure the stimuli and subjects in relation to a number of different attributes simultaneously. As well as obtaining the responses to the stimuli A, B, C the stimuli A' B' C', A'' B'' C'', A''' B''' C''' , for example, are also obtained. These can be located by means of average response values, to establish four continua of stimuli.

To make this development less abstract: these might all represent the concepts THE GOVERNMENT, YOUNG PEOPLE and HAPPINESS. A might be the stimulus THE GOVERNMENT good/bad, B might be the stimulus YOUNG PEOPLE good/bad, and C might be the stimulus HAPPINESS good/bad. A' then might be THE GOVERNMENT kind/cruel, B' then would be YOUNG PEOPLE kind/cruel and C' HAPPINESS kind/cruel and so on. The problem, as in scale analysis, is to represent the data in terms of a least number of dimensions. If, for example, the continua A B C, A' B' C', and A'' B'' C'' are identical, (and the criterion for their being considered identical will be further discussed) these can be represented in terms of one dimension. A further dimension may, however, be needed to describe the stimuli A''' B''' C'''. The concepts A B C , can then be plotted in a two dimensional space.

It would not be possible to build up multidimensional frameworks of measurement in this way with scale analysis, as the relationship between any two continua is not assumed to be known. But, of course, scale analysis could be used to study the dimensionality of

semantic differential data, to test whether one, or more continua are necessary to describe the data. Theoretically A B C, A' B' C' and A'' B'' C'' or any subset of these should produce responses which scale. A B C, A' B' C', A'' B'' C'' and A''' B''' C''' should produce responses which do not scale as a whole (though selected subsets should do so) provided the assumptions made in the analysis of the semantic differential data, about the scale continua, are correct.

Of course, subjects could be plotted in a multidimensional framework, just as validly as stimuli, but Osgood is concerned with the nature of the stimuli rather than the nature of the subjects as he aims to index meaning.

The criteria of dimensionality set out by the two methods thus differ. In scale analysis a configuration of data is given and the investigator tests the usefulness of regarding the data as being generated by a single variable. Any dimension that preserves the order of responses (and hence stimuli and subjects) can be the scale variable. On a graphic semantic differential the continua are given, as part of the stimuli, and the problem is to discover a least number of dimensions to describe the responses on these continua. To relate continua to one another the distances between responses must be known. For instance in the example, to equate the dimensions defined by the stimuli A B C, A' B' C' and A'' B'' C'' it was necessary to know the location of the zero, and the distances from the zero of all the points representing stimuli. For example the distance between zero and A must be equal to the distance between zero and A', and zero and A'' and so on. Or to put it

another way, the value of responses must be equal. Of course, factor analysis allows for some error: an unrealistic case was described for the sake of example. In practice factor analysis will not give a set of dimensions which perfectly describe the data unless there are as many factors as there were variables.

The semantic differential, then, employs a higher level of measurement and hence more sophisticated analyses of data are possible. However, many assumptions are involved in treating scales as quantitative variables.

## 2. Metric Assumptions of the Semantic Differential

To plot a concept in semantic space a number of assumptions are made about the data.

Concept A							Concept B										
(+)	<u>a<sub>1</sub></u>	<u>a<sub>2</sub></u>	<u>a<sub>3</sub></u>	<u>a<sub>4</sub></u>	<u>a<sub>5</sub></u>	<u>a<sub>6</sub></u>	<u>a<sub>7</sub></u>	(-)	(+)	<u>b<sub>1</sub></u>	<u>b<sub>2</sub></u>	<u>b<sub>3</sub></u>	<u>b<sub>4</sub></u>	<u>b<sub>5</sub></u>	<u>b<sub>6</sub></u>	<u>b<sub>7</sub></u>	(-)
(+)	<u>a'<sub>1</sub></u>	<u>a'<sub>2</sub></u>	<u>a'<sub>3</sub></u>	<u>a'<sub>4</sub></u>	<u>a'<sub>5</sub></u>	<u>a'<sub>6</sub></u>	<u>a'<sub>7</sub></u>	(-)	(+)	<u>b'<sub>1</sub></u>	<u>b'<sub>2</sub></u>	<u>b'<sub>3</sub></u>	<u>b'<sub>4</sub></u>	<u>b'<sub>5</sub></u>	<u>b'<sub>6</sub></u>	<u>b'<sub>7</sub></u>	(-)

To plot concept A in a semantic space it is assumed that

- 1)  $a_1 a_2 a_3 a_4 a_5 a_6 a_7$  can be arranged on a linear continuum  
 $a'_1 a'_2 a'_3 a'_4 a'_5 a'_6 a'_7$  can be arranged on a linear continuum.
- 2)  $a_4 = a'_4$ . That is, they both have the same value because they are indicative of neutrality, and can be represented by an absolute zero which is common to all scales.
- 3)  $a_1 = a'_1$  and not  $a'_7$ . That is the direction of scales is known.

- 4) the distance between  $a_1$  and  $a_4$  is equal to the distance between  $a_4$  and  $a_7$ , and these distances each equal the distances between  $a'_1$  and  $a'_4$ , and  $a'_4$  and  $a'_7$ .
- 5) the distances between  $a_1$  and  $a_2$ ,  $a_2$  and  $a_3$ ,  $a_3$  and  $a_4$ ,  $a_4$  and  $a_5$ ,  $a_5$  and  $a_6$ ,  $a_6$  and  $a_7$  are equal, and these all equal the distances between  $a'_1$  and  $a'_2$ ,  $a'_2$  and  $a'_3$ ,  $a'_3$  and  $a'_4$ ,  $a'_4$  and  $a'_5$ ,  $a'_5$  and  $a'_6$ ,  $a'_6$  and  $a'_7$ .

The concept can then be represented by coordinates on two axes. The axes can be rotated to give coordinates on two orthogonal axes, or, if responses are equal (such that a subject who chooses  $a_1$  also chooses  $a'_1$ , a subject who chooses  $a_2$  also chooses  $a'_2$ , a subject who chooses  $a_3$  also chooses  $a'_3$ , a subject who chooses  $a_4$  also chooses  $a'_4$  and so on, and vice versa) or have the same value the concept can be represented as a point on one axis.

To plot both A and B in a common semantic space it is assumed that

- 1)  $a_1 a_2 a_3 a_4 a_5 a_6 a_7$  can be arranged on a linear continuum  
 $a'_1 a'_2 a'_3 a'_4 a'_5 a'_6 a'_7$  can be arranged on a linear continuum  
 $b_1 b_2 b_3 b_4 b_5 b_6 b_7$  can be arranged on a linear continuum  
 $b'_1 b'_2 b'_3 b'_4 b'_5 b'_6 b'_7$  can be arranged on a linear continuum.
- 2)  $a_4 = a'_4 = b_4 = b'_4$ .
- 3)  $a_1 = a'_1$  and not  $a'_7$   
 $a_1 = b_1$  and not  $b_7$   
 $a_1 = b'_1$  and not  $b'_7$ .

- 4) the distance between  $a_1$  and  $a_4$  is equal to the distance between  $a_4$  and  $a_7$ , and that these distances each equal the distance between  $a'_1$  and  $a'_4$ ,  $a'_4$  and  $a'_7$ , and  $b_1$  and  $b_4$ ,  $b_4$  and  $b_7$ ,  $b'_1$  and  $b'_4$ ,  $b'_4$  and  $b'_7$ .
- 5) the distances between  $a_1$  and  $a_2$ ,  $a_2$  and  $a_3$ ,  $a_3$  and  $a_4$ ,  $a_4$  and  $a_5$ ,  $a_5$  and  $a_6$ ,  $a_6$  and  $a_7$  are equal, and these are all equal to the distances between  $a'_1$  and  $a'_2$ ,  $a'_2$  and  $a'_3$ ,  $a'_3$  and  $a'_4$ ,  $a'_4$  and  $a'_5$ ,  $a'_5$  and  $a'_6$ ,  $a'_6$  and  $a'_7$ ,  $b_1$  and  $b_2$ ,  $b_2$  and  $b_3$ ,  $b_3$  and  $b_4$ ,  $b_4$  and  $b_5$ ,  $b_5$  and  $b_6$ ,  $b_6$  and  $b_7$  and  $b'_1$  and  $b'_2$ ,  $b'_2$  and  $b'_3$ ,  $b'_3$  and  $b'_4$ ,  $b'_4$  and  $b'_5$ ,  $b'_5$  and  $b'_6$ ,  $b'_6$  and  $b'_7$ .
- 6) the continuum underlying  $a_1 a_2 a_3 a_4 a_5 a_6 a_7$  can be identified with that underlying  $b_1 b_2 b_3 b_4 b_5 b_6 b_7$  (they are defined by the same terms) and the continuum underlying  $a'_1 a'_2 a'_3 a'_4 a'_5 a'_6 a'_7$  can be identified with that underlying  $b'_1 b'_2 b'_3 b'_4 b'_5 b'_6 b'_7$  (they are defined by the same terms).

The concepts can be represented by coordinates on two axes which can be rotated to give coordinates on two orthogonal axes, or if the responses are identical such that  $a_1 = a'_1$ ,  $a_2 = a'_2$ ,  $a_3 = a'_3$ ,  $a_4 = a'_4$ ,  $a_5 = a'_5$ ,  $a_6 = a'_6$ ,  $a_7 = a'_7$  and  $b_1 = b'_1$ ,  $b_2 = b'_2$ ,  $b_3 = b'_3$ ,  $b_4 = b'_4$ ,  $b_5 = b'_5$ ,  $b_6 = b'_6$ ,  $b_7 = b'_7$  for all subjects the concepts can be represented as points on a single continuum.

To plot the responses of two people in a common semantic space the same assumptions are made. The previous example can be rewritten:

Rating of concept A by person 1	Rating of concept A by person 2
(+) <u>a<sub>1</sub></u> <u>a<sub>2</sub></u> <u>a<sub>3</sub></u> <u>a<sub>4</sub></u> <u>a<sub>5</sub></u> <u>a<sub>6</sub></u> <u>a<sub>7</sub></u> (-)	(+) <u>b<sub>1</sub></u> <u>b<sub>2</sub></u> <u>b<sub>3</sub></u> <u>b<sub>4</sub></u> <u>b<sub>5</sub></u> <u>b<sub>6</sub></u> <u>b<sub>7</sub></u> (-)
(+) <u>a'<sub>1</sub></u> <u>a'<sub>2</sub></u> <u>a'<sub>3</sub></u> <u>a'<sub>4</sub></u> <u>a'<sub>5</sub></u> <u>a'<sub>6</sub></u> <u>a'<sub>7</sub></u> (-)	(+) <u>b'<sub>1</sub></u> <u>b'<sub>2</sub></u> <u>b'<sub>3</sub></u> <u>b'<sub>4</sub></u> <u>b'<sub>5</sub></u> <u>b'<sub>6</sub></u> <u>b'<sub>7</sub></u> (-)

The same assumptions that were made in plotting two concepts, are made in plotting two persons in a semantic space.

The identity of responses was given as a criterion for identifying two continua. In practice, correlations may be used as the data of factor analysis. If a series of responses such as  $a_1$ ,  $a'_1$ ,  $a''_1$  does not have this perfect identification (such that when a person chooses A, he also chooses  $a'_1$  and  $a''_1$ , or if he chooses  $a'_1$  he also chooses  $a_1$  and  $a''_1$  and so on) there may nonetheless be some correlation. For instance when a person chooses  $a_1$  he tends to choose  $a'_1$  or  $a'_2$  and so on. Factor analysis then indicates that the responses can, with a certain margin of error, be indicated on one continuum, but more than one factor will be needed to describe the data completely.

In practice individual response distributions are not studied, as they are in scale analysis, but responses are summed across individuals to give a generalized score for the concept on a number of scales. The factorial structure of this generalized data does not necessarily represent the factorial structure of individual protocols. For example, if some subjects evaluate a concept highly, and some subjects evaluate it as being extremely negative, the average value may be somewhere near the origin. If data are analysed in the normal way it will be concluded

that the scale is not particularly evaluative in relation to the concept. The values are not reproducible as in scale analysis and this entails some loss of information.

In summary: although the semantic differential procedure gives a higher level of measurement this is at the cost of making certain assumptions about the nature of the data, which are not made by scale analysis, and which may or may not be correct. The semantic differential procedures do not allow for a check on these assumptions, so external evidence of their correctness must be furnished if the semantic differential is to be regarded as a useful measuring instrument.

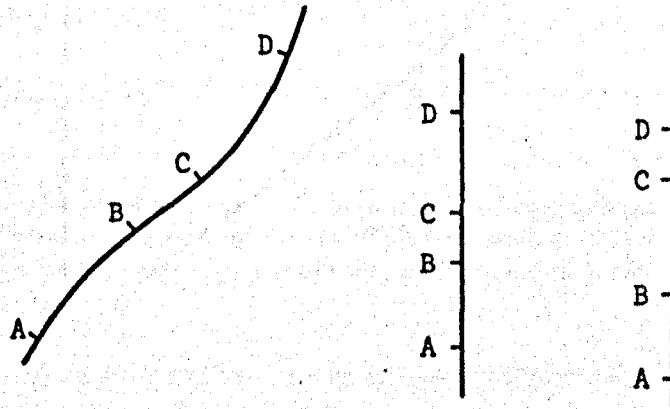
TABLE 16

Example VII : The Identity of Continua

1. Scale Analysis

Continua are identical if the order between elements is maintained.

For example:



But not:

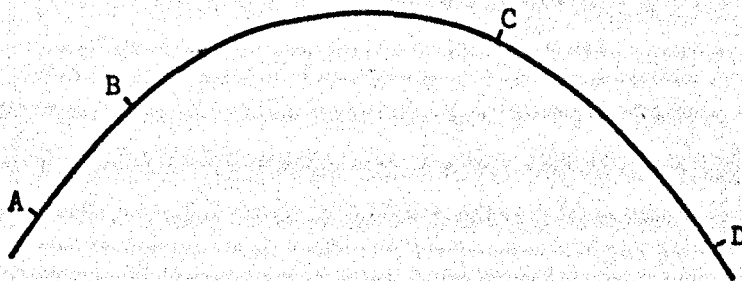
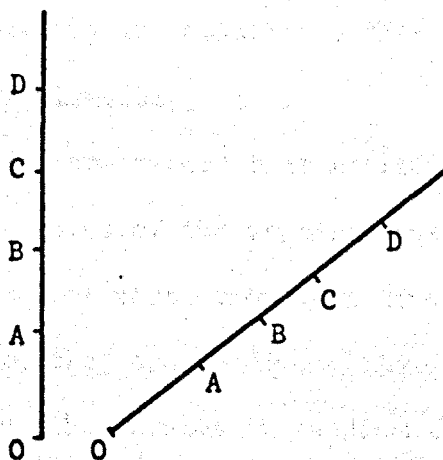


TABLE 16 (continued)

2. Semantic Differential

Continua are identical only if the distances of the elements from the origin are maintained. For example:



## CHAPTER V

### THE SEMANTIC DIFFERENTIAL AS AN ATTITUDE SCALE

#### 1. Attitude as a Component of Meaning

Osgood et al. (1957) identify the evaluative component of meaning, as measured by the semantic differential, with attitude. Osgood begins very plausibly,

"Most authorities are agreed that attitudes are learned and implicit - they are inferred states of the organism that are presumably acquired in much the same manner that other such internal learned activity is acquired. Further, they are predispositions to respond, but are distinguished from other states of readiness in that they predispose towards an evaluative response (p.190)."

He continues:

"Thus attitudes are referred to as "tendencies of approach or avoidance", or as "favourable and unfavourable", and so on. This notion is related to another shared view - that attitudes can be ascribed to some basic bipolar continuum with a neutral or zero reference point, implying they have both direction and intensity and providing a basis for the quantitative indexing of attitudes (p.190)."

He goes on to identify attitudes with the internal mediational activity which operates between most stimulus and response patterns, and to define an attitude towards an object as its projection onto the evaluative dimension in semantic space.

Now this is not a commonly accepted definition of attitude as it is concerned with affective rather than cognitive reactions. On the basis of every day usage, attitudes are usually thought of as having both affective and cognitive components. A statement such as "I hate the Jews" is largely affective, and attributes an emotion to the subject rather than a property to the object. A statement such as "A lot of Jews live in Israel" is largely cognitive; it is a statement which, if the terms are defined, can be checked against statements of fact, such as Census records. A cognitive statement can provide a basis for discussion about the object, an affective statement provides a basis for discussion of the subject and his emotions only. Most attitudes seem to lie somewhere between these two extremes. The subject attributes a property to an attitude object, so the attitude statement has some content, and hence bears some semblance to a statement of fact, while displaying some bias of emotional overtones. Osgood defines attitudes as being largely affective and almost entirely ignores the cognitive aspect of attitudes, as he does the denotative aspect of meaning. This is part of his behaviourist standpoint. Not all writers in the field of social attitudes would accept his view of the reducibility of complex human behaviour to simple stimulus-response models. So while it is possible to agree that attitudes are learned and implicit and are inferred states of the organism that are presumably acquired in much the same manner that other internal learned activity is acquired, it is not necessary to agree that conditioning is the only possible form of learning.

Osgood's suggestion that attitudes are tendencies of approach or avoidance similarly seems to oversimplify the situation. For instance, people are not normally said to have an attitude to an electric shock, but they do avoid an electric shock, not so much because it is good or bad, but rather because it is painful.

Similarly, the conception of attitudes as favourable or unfavourable may be a useful one when it is desired to rank statements or respondents on such a continuum, but it is not an adequate definition of attitude. Oppenheim (1966) suggests

"Our thinking on the nature of attitudes has been rather primitive. Most of the time we tend to perceive them as straight lines, running from positive, through neutral, to negative feelings about the object or issue ... There is no proof, however, that this model of a linear continuum is necessarily correct, though it does make things easier for measurement purposes. For all we know attitudes may be shaped more like concentric circles or overlapping ellipses or three dimensional cloud formations (p.107)".

And even if it is accepted that some attitudes can be usefully perceived as being linear, this does not mean that they are necessarily bipolar. For instance attitudes towards loneliness in old age might range from "I think loneliness is the worst problem of old age" to "I don't think loneliness is an important problem of old age", from a high rating of the importance of loneliness to a neutral zone.

Osgood's definition of attitude as the projection of the concept onto the evaluative dimension suggests that every concept has

an attitudinal component, even though this may be of zero magnitude. This is intuitively appealing but does not accord with normal usage, and does not seem to be very useful operationally.

## 2. The Measurement of Attitudes with the Semantic Differential

The semantic differential may, then, be used to measure attitudes.

"to index attitude we would use sets of scales which have high loadings on the evaluative factor across concepts generally and negligible loadings on other factors, as determined from our various factor analytic studies. Thus, scales like good/bad, optimistic/pessimistic and positive/negative should be used rather than scales like kind/cruel, strong/weak or beautiful/ugly because the latter would prove less generally evaluative as the concept being judged is varied. (Osgood et al. 1957, p.91)".

and

"For purposes of scoring consistency we have uniformly assigned the unfavorable poles of our evaluative scales (e.g., bad, unfair, worthless etc.) the score "1" and the favorable poles (good, fair, valuable) the score "7" - this regardless of the presentation of the scales to subjects in the graphic differential, where they should be randomized in direction. We then merely sum over all evaluative ratings to obtain the attitude "score" (p.191)".

The advantage of this procedure is its simplicity. The direction of the attitude, favourable or unfavourable, is indicated

by the subject's selection of polar terms. The neutral point is known. The intensity of the attitude is indexed by how far along the evaluative dimension, from the origin, the score lies. The unidimensionality of the attitude scale is believed to be established by the factor analytic procedures from which the scales are selected.

Furthermore, the semantic differential is a generalized attitude scale. The same set of scales can be used to differentiate a diverse array of concepts or attitude objects. This means that a scale exists to measure the attitude towards some new policy or event, or that a number of diverse attitudes can be directly and quantitatively compared with one another. The ideal conditions for using the semantic differential as a completely generalized instrument have not yet been fulfilled. One set of master scales which maintain a high loading on the evaluative factor, regardless of the concept being judged, has not yet been constructed. Osgood is however optimistic,

"we need to test the generality of these scales by comparing them with a battery of varied, specific attitude-measuring instruments, demonstrating (a) that these scales maintain high intercorrelation among themselves across the objects being evaluated and (b) that the summation scores derived from them jointly display high and roughly equal correlations with the various specific attitude-measuring instruments used as criteria. The evidence we have collected so far indicates that this will be a likely conclusion (1957, p.197)."

This does not seem to be a probable conclusion. The semantic differential is designed to measure affective reactions to signs or

concepts. Other attitude scales are designed to measure to a greater extent the content or cognitive component of attitudes. Words like FASCISM, COMMUNISM and BOSSES are usually avoided because respondents are likely to react to the negative terms, rather than to the content of the attitude statements. Osgood does not feel this is a criticism. Subjects may have an unfavourable attitude to the concept FASCISM, and yet agree with many of the beliefs of the Fascists. This, he says, is not a fault of the measuring instrument. The concept should be compared with other related concepts such as CENTRALIZATION OF POWER IN THE HANDS OF A STRONG LEADER. This is quite reasonable. The subject has more than one attitude towards FASCISM and towards Fascist policies and these need not be in the same direction. But this makes it very difficult to compare a content scale with a semantic differential. No rules are laid down for the selection of concepts. It might be possible to hypothesize that if a content scale included the word FASCIST in all questions, then it would largely measure the subject's reaction to the negatively biased label, and the results should agree highly with a semantic differentiation of the concept FASCISM. If on the other hand, the content scale had been constructed so as to exclude this type of label, it would not necessarily be related to the subject's reaction and differentiation of the concept FASCISM. Further a subject who agreed with many Fascist ideas, as studied on the concept scale, might dislike, in a personal way, various personalities such as HITLER, SENATOR M<sup>C</sup>CARTHY, FRANCO, or the opposite might apply and a subject who disagreed on the whole with

Fascist policies might nonetheless admire Franco as a person. The relationship between a content scale and a semantic differential could be expected to be contingent upon the selection of concepts, and no operational rules have been laid down. This is not a criticism of the semantic differential, as it may be extremely useful to study the inter-relationships between the various ideas a person has, and his reaction to certain labels, but to identify the two because they are both called attitudes, seems to be premature.

### 3. Relationship to other Attitude Scales

Osgood and his associates found a fairly simple correspondence between semantic differential and other attitude scores. Suci (1952) was able to differentiate between high and low ethnocentrics as determined by their attitude scores on the E scale. Osgood et al. (1957) report a study which compared the ratings of THE NEGRO, THE CHURCH and CAPITAL PUNISHMENT against a series of scales with Thurstone (1931) scales designed to gauge attitudes towards these things. As both scales attempt to place these concepts, and the subjects, on a favourable-unfavourable continuum a high relationship could be expected. A high correspondence was found and Osgood et al. (1957) conclude that

"whatever the Thurstone scales measure, the evaluative factor of the semantic differential measures just about as well (p.194)".

But it is important to remember that this set of questions corresponded highly with this set of semantic differential stimuli. There is nothing in the techniques that guarantees that only this set of questions could

have been selected, or only this set of semantic differential concepts and scales could have been used, to study the same attitude.

Osgood et al. (1957) also report a study in which the concept CROP ROTATION as evaluated on the scales good/bad, fair/unfair, valuable/worthless, was compared with a Guttman type scale. 28 subjects were used. The rank order correlation was highly significant ( $\rho = .78$ ;  $p < .01$ ) and the conclusion is drawn that

"we may say that the Guttman scale and the evaluative scales of the differential are measuring the same thing to a considerable degree (p.194)".

The questions used for the scale are not reported, though once again the relationship is contingent on the content of questions. For example, if the questions were something like "I think crop rotation is bad", to which the subject responds agree, disagree, or don't know, a high relationship could be expected whereas a more complex statement which avoided the terms used on the differential may have produced a less clear relationship. That is, Osgood shows that the instruments can be used to measure the same thing, but he does not define the conditions under which they do measure the same thing.

Tittle and Hill (1967) evaluate a number of attitude measures, including a Guttman scale and a semantic differential, against a behavioural criterion. The techniques were (1) Thurstone successive-interval technique, (2) a semantic differential procedure, (3) a summated rating (Likert) technique and (4) a Guttman type scale. A simple self rating was also obtained and the efficiency of each

technique was assessed in terms of its correspondence with five criteria of behaviour.

The Guttman scale was derived using the same responses utilized in the summated rating scale. A random sample of 95 questionnaires was selected from the 213 obtained. The Cornell technique was used to scale responses, and ten items were found to form a scale with a coefficient of reproducibility of 0.928. The items were retested for scalability after being administered to the main sample, and further category combination was needed to produce a coefficient of reproducibility of 0.930. Guttman warns against the use of scale analysis to select items (see Chapter II), and the rejected items may have been more relevant to the behavioural criterion than those selected. One would like to know the content of all the questions.

A semantic differential was constructed using the scales good/bad, valuable/worthless, clean/dirty, pleasant/unpleasant, wise/foolish, fair/unfair, complex/simple, active/passive and deep/shallow, to rate the concepts (1) voting in student elections, (2) discussing student political issues, (3) holding student political office, (4) helping in a student political campaign, (5) keeping informed about student politics.

The behavioural criteria consisted in part of self-reported political activity, but this was checked to some extent by access to voting records.

The interrelationship of techniques was found to be as follows:

	Guttman Scale	Thurstone Scale	Semantic Differential	Self-Rating
Likert Scale	•796	•588	•619	•511
Guttman Scale		•445	•523	•476
Thurstone Scale			•432	•337
Semantic Differential				•387

The highest degree of association, then, is between the Likert and Guttman scales. This is not surprising as they were not independently constructed. The degree of association between the semantic differential and the Guttman scale is considerably lower than that of 0.78 reported by Osgood et al. (1957). The relativity of the relationship between techniques is shown by the fact that although Osgood reports an instance of high correlation of both the Thurstone scale and the semantic differential, and the Guttman scale and the semantic differential, in this instance the Guttman and Thurstone scales are not highly correlated with one another.

The association between each scale and each behavioural index was measured by the Goodman-Kruskal gamma. The mean association of the scales with the behavioural criteria were (1) the 15 item Likert scale 0.543, (2) the 10 item Guttman scale 0.419, (3) the self-rating scale 0.396, (4) the semantic differential 0.339 and (5) the Thurstone scale 0.255. Tittle and Hill (1967) conclude

"In this instance the Likert scale was clearly the best predictor of behaviour. It was most highly associated with every one of the five

behavioural indices (p.211)".

Split-half reliability coefficients, based on the Spearman-Brown formula were calculated. The authors acknowledge that the procedure makes interval assumptions, but say they know of no alternative ordinally-based procedure. The usefulness of this approach in relation to the Guttman scale can be questioned (in a perfect scale the split halves are each functions of the scale variable, but need not correlate with one another, see Chapter II). The coefficients of reliability were (1) Likert scale 0.95, (2) the semantic differential 0.87, (3) the Guttman scale 0.80, (4) the Thurstone scale 0.67. The differential reliability, then, seemed to account for some of the differential predictive ability of the techniques. If the Guttman scale, for which the coefficient of reliability is inappropriate, is omitted, the coefficients of reliability give the same ordering of the scales as did the associations with behavioural criteria.

Tittle and Hill (1967) also suggest that the content of scales was related to their predictive ability. The Likert scale contained more personal pronouns than the other scales. Similarly, the Guttman scale contained more self-referent items than did the Thurstone scale. The specificity of scales seemed to be related to their predictive ability.

On the basis of their study Tittle and Hill (1967) are critical of the semantic differential as a measure of attitude.

"The semantic differential as a measure of attitude suffers a serious disadvantage. Subjects tend to respond in a set. They observe that

'desirable' things appear on one side of a continuum and 'undesirable' things appear on the other. The discriminational process then apparently becomes a matter of self-evaluating overall attitude and marking the scale accordingly, with little distinction between adjectival pairs. Interspersing reversed continua probably only serves to make the respondent's task more difficult without fundamentally altering the problem. In this instance the tendency for subjects to adopt a response set probably accounts for the fact that the semantic differential procedure resulted in a measure having high reliability, but low predictive validity (p.213)".

It may be, however, that the semantic differential is valid as a measurement of attitude to these concepts, but not predictive of behaviour in that the behaviour is relative, while the allocation of the concepts to a continuum is irrelative. For example, a student may feel that all the concepts are to be highly evaluative, that it is a good thing to vote in student elections, discuss student political issues and so on, but that he himself must think first of his studies and allocate most of his time to work, rather than to extracurricular activities. The other attitude scales used tended to be more related to the subject by means of personal pronouns. The difficulty involved in selecting concepts for comparing the semantic differential with other attitude scales is apparent.

## CHAPTER VI

### THE SCALE ANALYSIS OF SEMANTIC DIFFERENTIAL DATA

#### 1. Investigation of the Metric Properties of the Semantic Differential

The metric properties of the semantic differential as a multi-dimensional measuring instrument were discussed in Chapter IV. Having isolated one of these dimensions as a possible measure of attitude it will now be valuable to discuss, in a more concrete way, the properties of the semantic differential as a unidimensional attitude scale.

The properties of a scale derived from Guttman's scale analysis have been described at length in Chapter II. Given that the data form a scale, they can be treated conveniently as being functions of a quantitative variable. An ordinal scale is constructed.

The metric properties of the semantic differential are of a different nature as, to some extent, they are built into the data.

As Coombs (1953) points out:

"no property of data can be said to hold unless the methods of collecting and analysing the data permit alternative properties to exhibit themselves (p.487)".

For example, the linearity of a semantic differential scale (such as good/bad, or weak/strong) is not tested as the subject is allowed to check only one scale interval. If, for example, the subject was instructed to check two intervals it may be possible to construct an underlying continuum from the data. If the data were such that

subjects did not choose adjacent intervals the conclusion that the scales was not linear could be drawn.

Scale intervals are assumed to be equal and the origin of each scale is assumed to be known. Methods which assume that variables are quantitative cannot be used to give evidence supporting these assumptions, as this merely reveals the axioms on which the methods were based. In other words unless an investigator uses methods which do not make the same assumptions, he is merely examining the system of formal relations imposed on the data, not the data itself.

Messick (1957) uses the method of successive intervals to examine the scale intervals and the location of the zero of a number of semantic differential scales.

"When an integer score is assigned as a concept's scale position on a particular scale, the property of equal intervals within that scale is assumed. Similarly, when a distance measure is taken over several scales, equal intervals between scales are assumed. In addition the application of factor analytic techniques to the assigned scores involves assumptions concerning the location of the scale origins; i.e., it is assumed that the zero point falls at the same place on each scale, namely at the centroid (p.200)".

Messick finds that the intervals, as determined by the method of successive intervals, are not exactly equal but tend to be fairly consistent between scales. He also finds that the origin falls in approximately the same place on each of the scales studied. By comparing the assumed scale boundaries with the scaled boundary positions

he concludes that the correlations indicate that little distortion would be introduced by using successive integers as category midpoints. He finds that the scaling properties implied by the semantic differential procedures have some basis other than mere assumption.

Green and Goldfreid (1966) examine the linearity of the bipolar scales. The conclusions they draw about the metric properties of the semantic differential differ from those of Messick. The method of successive intervals estimates the widths of scale intervals making up an underlying continuum from the cumulative proportion distributions for a given set of statements by assuming

"that these cumulative proportion distributions are normal for each statement when they are projected on the unknown psychological continuum (Edwards, 1957, p.124)".

Messick, then, makes certain metric assumptions (normal distributions, linearity) to establish others (equal intervals, common zero). Green and Goldfreid (1967) must also make some metric assumptions. Scores are treated as functions of a quantitative variable and factor analysis is employed, so equal intervals and a common zero is assumed. However the scales are constructed in such a way that non-linearity can appear in the results.

Osgood et al. (1957) say that

"scales should be linear between polar opposites and pass through the origin ... At present we merely assume that the scales defined by familiar and common opposites have these properties, but research on the problem needs to be done (p.79)".

Green and Goldfreid (1967) ask whether these familiar and common opposites exhaust the semantic space and suggest

"we cannot be sure even of these 'familiar and common' word pairs until further research has been done (p.3)".

Bipolarity may not be clear insofar as the opposite of an adjective may imply further dimensions. For example the opposite of 'passive' implies activity, but it may also imply aggressiveness. The opposite of 'active' may be 'quiet' or 'placid' rather than 'passive' which has different evaluative overtones. Or both polar adjectives may imply some similar characteristic. For example rugged-delicate is quoted by Osgood et al. (1957) as a scale which is non-linear with respect to the evaluative dimension, as both of these adjectives possess similar positive evaluative characteristics.

The use of bipolar scales involves the assumption of reciprocal antagonism. It must be supposed, for instance, that a concept cannot be thought of as being both good and bad. As concepts are, as a rule, complex and may have conflicting attributes or vary through time, this is a supposition which is open to question. In practice, of course, the semantic differential does allow for an ambivalent response and something conceived of as having both good and bad attributes can be rated as neutral. This does make the interpretation of neutral responses difficult and suggests that scale linearity may be an artifact of the measuring technique rather than a reflection of the underlying dimensions of judgement.

The bipolarity of semantic space also requires that scales are symmetric around the origin. Defining adjectives are conceived of as being equal (in intensity) and opposite (in direction). Green and Goldfreid (1965) working with unipolar scales translate these requirements into concrete terms. For instance if "good" has a loading of 0.88 on the evaluative dimension, they argue "bad" should have a loading of -0.88 on the evaluative dimension. Their findings are discouraging. Some of the common antonyms show opposite loadings, but they do not necessarily show symmetry. An attempt to match the pairs of antonyms, blind, on the bases of their factor loadings, failed. Green and Goldfreid do locate some bipolar factors but

"even so, the typical outcome is for unimodal factors to appear. When bipolarity does appear it tends to be concentrated on certain concepts (p.21)".

They find that there is not a generalized semantic space in which all concepts are related, but rather a number of semantic spaces which are "averaged" across a number of concepts. They conclude

"Osgood and his associates have in fact imposed an arbitrary and artificial structure in the domain they call generalized semantic space (p.31)".

However,

"It does not follow that the semantic differential is useless, but it does follow that researchers should bear its characteristics in mind when they use it to obtain and interpret data (p.31)".

## 2. The Scale Analysis of Semantic Differential Data

Scale analysis can be applied to semantic differential data. As scale analysis does not make the same assumptions about the scales as does a typical factor analytic approach the same structure will be accorded to the data only if the assumptions underlying the factor analytic approach are justified. Scale analysis treats the data as purely qualitative and does not make assumptions about linearity, normal distributions, scale intervals or the location of the centroid zero, as do the other methods discussed. The linearity of scales, however, cannot be shown as data in the normal form is insufficient in that the subject is instructed not to check the scales more than once. The data that will be analysed was drawn from a survey (Study B, see Chapter VII); in the ideal case unipolar scales might also have been studied.

Before looking at an actual application of scale analysis to semantic differential data it may be of value to consider an ideal case. Suppose seven subjects rate a concept on a number of evaluative scales such as good/bad, nice/nasty, kind/unkind, pleasant/unpleasant, virtuous/wicked. To achieve a perfect factor structure there are seven possible response patterns. If the concept is rated +3 on one scale it should be rated +3 on every other scale. Similarly if the concept is rated +2 on one scale, it should be rated +2 on every other scale and so on, as in the perfect case each scale is an equally good measure of the evaluative factor. If this result emerges the scales can clearly be considered to be unidimensional. Factor analysis would

TABLE 17

Example VIII : Scale Analysis and Factor Analysis (Equal Intervals)

Suppose the responses of seven subjects, on the following scales,

1. good / bad
2. nice / nasty
3. kind / unkind
4. pleasant / unpleasant
5. virtuous / wicked

are as follows:

	Scales				
	1	2	3	4	5
Subjects: 1	-3	-3	-3	-3	-2
2	-2	-2	-2	-2	-1
3	-1	-1	-1	-1	0
4	0	0	0	0	0
5	1	1	1	1	0
6	2	2	2	2	1
7	3	3	3	3	2

Scale analysis of the data will indicate unidimensionality:

TABLE 17 (continued)

S's	Scales																																				
	1							2							3							4							5								
	3	2	1	0	-1	-2	-3	3	2	1	0	-1	-2	-3	3	2	1	0	-1	-2	-3	3	2	1	0	-1	-2	-3	3	2	1	0	-1	-2	-3		
7	x							x							x							x															
6		x							x							x							x														
5			x							x							x							x													
4				x							x							x							x												
3					x							x							x							x											
2						x							x							x							x										
1							x							x														x									

TABLE 17 (continued)

Factor analysis may not indicate unidimensionality, as equal intervals are assumed.

Factoring	Scales				
	1	2	3	4	5
$\Sigma X_{ji}^2$	28.00	28.00	28.00	28.00	10.00
$\Sigma X_{j1} X_{ji}$	28.00	28.00	28.00	28.00	16.00
$C_{Ii}$	5.29	5.29	5.29	5.29	3.02
$\Sigma X_{ji}^2 - C_{Ii}^2$	0	0	0	0	0.86
$\Sigma X_{j5} X_{ji}$	16.00	16.00	16.00	16.00	10.00
$C_{I5} C_{Ii}$	16.00	16.00	16.00	16.00	9.14
$\Sigma X_{j5} X_{ji} - C_{I5} C_{Ii}$	0	0	0	0	0.86
$C_{IIIi}$	0	0	0	0	0.09
$\Sigma X_{ji}^2 - C_{Ii}^2 - C_{IIIi}^2$	0	0	0	0	0

The D-method of factoring (see Appendix II) indicates that, although one dominant dimension is present, a second dimension is needed to describe the data completely. That is, the factor analysis indicates that there is a small margin of error involved if the data are regarded as unidimensional, although scalogram analysis indicates the presence of a perfect scale.

In this example scales were related across subjects, but the responses of one subject to a number of concepts could have been considered to give the same results.

show one factor only, and scale analysis would show a special case of the parallelogram pattern, where the data do not differentiate between items, as these are all equal with respect to the attribute. But suppose "wicked" is considered by subjects to be more extreme than "bad". The subject who then ranks a concept -3 on the other scales may rank it only -2 on the fifth scale. The subject who ranks a concept -2 on the first four scales may rank it only -1 on the last scale and so on. This is set out in Table 17.

It may be wondered why the ideal types for factor analysis have been considered in terms of absolute values, while the factor analysis of correlations is usual. The point is that the continua are so short that scales cannot appear to be intercorrelated unless the absolute values are similar. This is illustrated in Table 19. If data scale factor analysis may indicate that more than one factor is involved because factor analysis assumes equal intervals and a centroid zero while scale analysis does not. The phenotypic data considered in Table 19 scale, although they are not correlated.

### 3. Scale Analysis of Semantic Differential Pilot Study Data

The differentiation of the concept OBEDIENCE by 60 subjects was used as the data for scale analysis. The concept was rated on the scales (1) good/bad, (2) gentle/violent, (3) fast/slow, (4) active/passive, (5) sweet/sour, (6) clean/dirty, (7) weak/strong, (8) large/small, (9) masculine/feminine, (10) useful/useless, (11) unnecessary/necessary, (12) valuable/worthless, (3) unimportant/

TABLE 18

Factor Analysis I : Coordinates of evaluative, potency, and activity scales selected from Osgood, based on the responses of 60 Pilot Study B subjects to the concept OBEDIENCE.

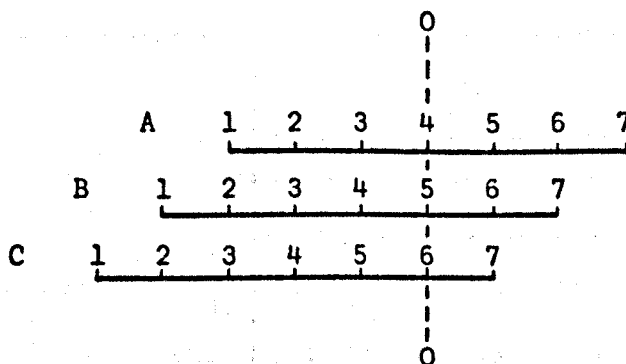
As the D-method of factoring begins with raw scores the loadings or coordinates can be greater than 1.

Dimension	Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
I	11.50	8.75	5.41	8.52	5.63	11.19	7.39	4.74	5.04	12.89	11.63	13.04	11.12	11.63	13.49
II	0.96	0.87	2.46	2.17	2.08	1.55	8.95	1.92	0.11	0.45	1.24	0.19	0.37	0.07	0
III	1.39	0.87	2.45	0.99	7.02	0.45	0	1.01	-1.30	-0.60	-0.84	-0.28	-0.37	-0.08	0
IV	0.89	2.00	1.65	1.87	0.23	1.86	0	0.03	6.47	0.41	-0.76	0.60	-1.17	0.81	0
V	1.84	0.34	-1.25	-0.65	0	0.11	0	0	0	-0.53	1.79	-0.40	6.41	-0.59	0
VI	1.42	0.90	0.27	1.58	0	5.89	0	0	0	2.75	-1.26	0.21	0	0.14	0
VII	0.11	1.82	2.57	0.67	0	0	0	0	0	-0.21	5.44	0.97	0	0.29	0
VIII	4.23	0.88	-1.45	4.77	0	0	0	0	0	-1.47	0	0.34	0	0.13	0
IX	0.39	1.59	0.90	0	0	0	0	0	0	0	0	-0.28	0	4.50	0
X	-1.45	0.04	4.22	0	0	0	0	0	0	0	0	-0.76	0	0	0
XI	1.09	2.88	0	0	0	0	0	0	0	0	0	0.22	0	0	0
XII	0.01	0	0	0	0	0	0	0	0	0	0	2.36	0	0	0
XIII	0.85	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TABLE 19

Example IX : Scale Analysis and Factor Analysis (Common Origin)

Scales which correlate, but are not symmetric about a common zero, can be visualized as follows:



Responses on these scales will tend to be bunched up at the poles.

A concept rated 1 on scale A may be ranked 1 or 2 on scale B, and 1, 2 or 3 on scale C. Possible response patterns might be:

	Scale		
	A	B	C
Subject: 1	1	1	1
2	1	1	2
3	1	2	3
4	2	3	4
5	3	4	5
6	4	5	6
7	5	6	7
8	6	7	7
9	7	7	7

TABLE 19 (continued)

The data form a perfect scale, but the D-method of factoring indicates the need to describe the data in terms of more than one dimension:

	Scales		
	A	B	C
Coordinates: $C_{Ii}$	11.60	13.68	15.43
$C_{IIIi}$	2.72	1.58	-
$C_{IIIIi}$	-	0.30	-

The semantic differential assumes that equal intervals and a common zero are metric properties of the scales that have some type of psychological basis. If the scales have, in practice, these properties, differences between conclusions drawn from the factor analysis and conclusions drawn from a factor analysis will be minimal.

important, (14) desirable/undesirable, (15) wise/foolish, with five scale intervals. The subjects are described in Chapter VII. The data was factor analyzed, using the D method described in Appendix II. The first factor can be identified with the evaluative factor and seems to be the dominant factor. The second factor seems to be general dynamism factor with highest loadings large/small and fast/slow. The third factor is less easily interpreted, having its highest loading on "sour/sweet" while being negatively loaded on other evaluative scales. A number of more specific factors are apparent but, on the whole, the factor analysis gives a structure similar to that reported by Osgood as being general to all semantic differential data. The first factor analysis acts as a check and confirms that the semantic space generated by the ratings of OBEDIENCE is not unique or irregular.

The evaluative scales, selected on the basis of high loadings on the evaluative factor from the first factor analysis, and the scale sweet/sour which is reported by Osgood as having a high loading on the evaluative factor were analyzed by means of scale analysis. The analysis utilized the scalogram board described by Trenaman (1960) and the resultant parallelogram pattern is reproduced in Table 20.

The coefficient of reproducibility, 0.894, falls below scale criterion and the existence of a quasi-scale only may be postulated. The pattern of error is random, except on scale 6, which is clean/dirty. This is not a function of the specific solution adopted, which is to some extent subjective when the scale has low reproducibility, as the ordering of scales strictly according to marginal totals shown in Table 21,

TABLE 20

Scalogram I : Scalogram Analysis of Semantic Differential Data. Evaluative Scales. Study B.

N = 60. (i) Data arranged visually to reduce error.

S.D. Scale	5	14	1	13	11	15	10	12	6	5	14	1	13	11	15	10	12	6
Response	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
	3	3	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3
Subjects 22	x	x	x	x	x	x	x	x	x									
21	x	x	x	x	x	x	x	x	x									
27	x	x	x	x	x	x	x	x	x									
15	x	x	x	x	x	x	x	x	x									
26	x	x	x	x	x	x	x	x	x									
18	x	x	x	x	x	x	x	x	x									
30	x	x	x	x	x	x	x	x	x									
82	x	x	x	x	x	x	x	x	x									
74	x	x	x	x	x	x	x	x	x									
65	x	x	x	x	x	x	x	x	x									
36	x	x	x	x	x	x	x	x	x									
79	x	x	x	x	x	x	x	x	x									
8	x	x	x	x	x	x	x	x	x									
28	x	x	x	x	x	x	x	x										x
71	x	x	x	x	x	x	x	x	x									
17		x	x	x	x	x	x	x	x	x								
85		x	x	x	x	x	x	x	x	x								
48		x	x	x	x	x	x	x	x	x								
47		x	x	x	x	x	x	x	x	x								
73		x		x	x		x	x	x	x			x			x		
80		x	x	x	x	x	x	x	x	x								x
12		x	x	x	x	x	x	x	x	x								x
9		x	x	x	x	x	x	x	x	x								x
6		x	x	x	x	x	x	x	x	x								x
60		x		x	x	x	x	x	x	x								x
5		x	x	x	x	x		x	x	x						x		x
92		x		x	x	x	x	x	x	x								x
56	x		x		x	x	x	x			x		x					x
20		x	x			x	x	x	x	x			x	x				
39	x	x		x		x	x		x			x		x			x	
46		x	x	x	x			x	x	x					x	x		
42	x	x	x			x	x	x					x	x				x
34		x	x	x	x				x	x			x	x				
11			x		x		x	x	x	x			x		x	x		
75			x	x		x	x	x	x	x			x		x			
40			x	x		x	x		x	x				x				
70				x		x	x		x	x				x			x	
29				x		x	x	x	x	x				x				x
94				x		x	x	x		x	x			x				
95		x				x	x	x		x	x		x	x				x
50						x	x			x	x		x	x				x
55		x				x	x	x	x	x			x	x		x		x
38				x		x	x	x	x	x				x				
91				x		x	x	x	x	x				x				
84				x			x		x	x				x	x		x	
58			x	x				x	x	x				x	x	x		
32			x	x				x	x	x				x	x	x		
51			x						x	x				x	x	x		
59				x					x	x				x	x	x		
37									x	x				x	x	x		
63						x				x	x			x	x		x	
66										x	x			x	x	x	x	
87										x	x			x	x	x	x	
24										x	x			x	x	x	x	
99										x	x			x	x	x	x	
97										x	x			x	x	x	x	
98										x	x			x	x	x	x	
90										x	x			x	x	x	x	
72										x	x			x	x	x	x	
64										x	x			x	x	x	x	

does not serve to reduce the error. Scale 5, sweet/sour is the scale with least error. This is possibly related to the fact that it has the most uneven marginal distribution (19 : 41 when the scales are dichotomized).

The scales used in this analysis were factor analyzed, using the D method again, to give a direct comparison between the two analyses. Like the scale analysis the factor analysis suggests the presence of one major factor and a number of more specific factors. The factor loadings, or strictly, dimension coordinates are given in Table 22.

An interesting feature is the second dimension, which takes the scale sweet/sour as a pivot. The scale is interpreted as being described properly in terms of two dimensions, rather than as being adequately represented on the main dimension as in the scale analysis. Scale analysis indicates that the scale sweet/sour is evaluative, but has steeper scale intervals than the other scales. As it was necessary to combine categories (or scale intervals) 2, 3 and 4 to achieve an approximate parallelogram pattern, and as the number of errors indicates the existence of a quasi scale only, any conclusions as to which interpretation is more valid would be tentative. However if the same margin of error is involved in each interpretation the conception of sweet/sour as an evaluative scale with steeper scale intervals would be more parsimonious than the interpretation of this scale in terms of more than one dimension.

TABLE 22

Factor Analysis II : Coordinates of evaluative scales selected from Osgood, as responded to by 60 Pilot Study B subjects differentiating the concept OBEDIENCE.

As the D-method of factoring begins with raw scores the loadings or coordinates can be greater than 1.

Dimensions	Semantic Differential Scale								
	1	5	6	10	11	12	13	14	15
I	11.32	5.70	11.02	12.67	11.47	12.97	10.94	11.47	13.34
II	0.34	7.31	0.71	-0.16	-0.33	-0.13	-0.05	-0.09	0
III	1.14	0	0.24	-0.24	1.99	-0.62	6.28	0.08	0
IV	1.96	0	6.00	0.42	-0.61	0.38	0	0.06	0
V	0.72	0	0	0.06	5.83	0.63	0	0.13	0
VI	1.46	0	0	-0.05	0	0.62	0	4.73	0
VII	4.46	0	0	0.70	0	2.43	0	0	0
VIII	0	0	0	3.12	0	0.61	0	0	0
IX	0	0	0	0	0	0.60	0	0	0

Also the coordinates on the first dimension do not indicate that scale 6, clean/dirty is less evaluative than other scales, although this scale did not appear to scale well with other items.

The analysis of data generated by one concept is of interest as it indicates that scale analysis and factor analysis of the data may give slightly different results, because factor analysis makes different assumptions about the data. This suggests that scale analysis could be used to study the validity of the metric assumptions of the semantic differential. It would be interesting to choose a more controversial concept so as to give a wider dispersion of responses, or to study a variety of concepts which would produce responses on both poles of the scales and apply scale analysis to individuals' responses to the set of concepts.

#### 4. The Relationship of Scale Analysis to Factor Analysis

Guttman suggests that

"From a scale analysis it can be predicted quite well what the factor analysis will show. (1951, p.201)".

But, he adds, factor analysis cannot indicate what scale analysis will show. The relationship between the two techniques is by no means simple as there are many forms of factor analysis, and for most methods there is no unique solution. The D method of factor analysis has been selected for use here as it was recommended by Osgood et al. (1957) for use with semantic differential data. It has the advantage of being objective insofar as there is only one solution.

The process of extracting dimensions can be continued until there are no residuals (i.e. until there are as many dimensions as there were scales), and dimensions are not rotated. In practice some subjectivity is still involved as there would be no point in using factor analysis if all the dimensions were used to plot the concept in semantic space or to determine factor scores. Since there are as many dimensions as there were scales the  $m$  scales could as easily be used to define the  $m$  dimensional semantic space. The purpose of the factor analysis is to reduce the number of dimensions. From Factor Analysis II scales 1, 6, 10, 11, 12, 13, 14 and 15 would probably be selected to represent an evaluative factor, but of these only scale 15 is perfectly described by its coordinates on this dimension. Another investigator might select two, three or even nine dimensions to describe the data.

Some methods of factor analysis may yield results similar to those obtained by scale analysis. Burt (1953) describes applications of factor analysis to qualitative data and indicates that the component weights derived by means of the least squares method of scale analysis, can be derived by means of factor analytic procedures.

Other methods of factor analysis may yield different interpretations of the dimensionality of data. For example:

"If the items themselves have high reproducibility then a factor analysis of tetrachorics will show not one common factor but several common factors (Guttman 1951, p.201)."

When factor analysis is applied to semantic differential data the variables (scales) are assumed to be quantitative. When scale

analysis is applied to semantic differential data the data are treated as being qualitative. If the data are in fact generated by quantitative variables then the results of the two analyses should be congruent. The results of the application of the two methods here were on the whole similar, indicating the presence of one major factor and a number of specific minor factors, though the composition of the factor varied slightly according to the method used. From the scale analysis it could have been predicted what the factor analysis would show, but as Guttman suggests, the results of the scale analysis could not have been predicted from the factor analysis in this case.

## CHAPTER VII

### EMPIRICAL COMPARISON OF GUTTMAN SCALE AND

### SEMANTIC DIFFERENTIAL

An opportunity to obtain data from a Guttman type scale and a comparable semantic differential form, as responded to by a large group of subjects, arose from a series of studies into the effects of ageing on social attitudes, carried out by the writer. As part of one of these studies a Guttman scale was used to investigate attitudes towards obedience, and a semantic differential form was used to evaluate the concept OBEDIENCE.

#### 1. Construction of the Guttman Scale

Two pilot studies (Study A and Study B) were carried out before the Guttman scale was administered to the main sample of subjects. The choice of content was to some extent determined by the need to construct a scale relevant to the main purposes of the final questionnaire. The concept OBEDIENCE was chosen. Two relevant questions were drawn from the F scale of Adorno et al. (1950). It was hoped that the implications of previous work with the F scale might pertain to this investigation. The other questions were developed from the F scale questions in an attempt to clarify, simplify and more clearly delimit the meaning of these statements. An intensity scale was also included. The relevant questions are given in Appendix III (i).

The items were administered to 92 subjects. A cluster technique of sampling was used for reasons of economy. As the main questionnaire was to be used in the Stoke-on-Trent C.B. the subjects were sampled as follows. The names of all the streets in the Stoke-on-Trent C.B. were drawn from a street directory and printed on cards. The cards were shaken in a large box and then a card bearing a street name was drawn out. The name of the street selected in this way was noted and the card replaced. Letters were then delivered to the occupants of each building in the street, explaining the purpose of the research and asking each occupant of twelve or more years of age to act as a respondent. Another street was then selected in the same way. When approximately 100 subjects had been located for interview this process was discontinued. It was not thought appropriate to use a stratified sampling technique as the main questionnaire was to be used with a random sample of residents in the area.

In most cases forms were left with occupants for completion and collected a week later. Occasionally, especially if respondents seemed to be very old, it was considered better to have forms filled in on the spot with the interviewer giving some assistance. This is less objective than allowing the respondent to work on his own, but, as the alternative would probably have been to receive incorrectly filled in, or even unattempted forms, it seemed the best approach. Often more than one call was needed to collect forms as subjects were not at home, or had forgotten to fill in the forms. On the eighth follow up call forms were collected, even if they had not been

completed. People who accepted forms, and agreed to fill them in, but after eight weeks had still not done so, were counted as "non-respondents". People who said they would not fill in forms were classified as "refusals". The response rates can be summarized as follows:

Study A. Number of streets visited : 3

Responding	Unassisted	90	92
	Assisted	2	
Not Responding	Non-Respondents	3	5
	Refusals	2	

Total: 97

Response rates were not calculated in the normal way (as a percentage of the number of people approached) as the sample unit was the street, or cluster, not the individual respondent.

A complete description of the sample cannot be given as some items of background information were omitted by some respondents. The setting out of the background information items was, for this reason, altered on the second study. Also the classification of occupations was not successful because a large number of subjects gave responses such as "housewife" or "retired". For this reason, respondents to the second questionnaire were asked to give their husband's or wife's occupation as well. This proved to be confusing, and not always applicable. The final method adopted was to ask the occupation of the head

of the family. Occupations could then be rated on the Hall-Jones (1950) scale of occupational prestige. Some frequency distributions which describe the sample are set out in Table 23.

The scale analysis employed the scalogram board technique. The scalogram board described by Trenaman (1960) was used. Scale analysis of 10 questions indicated that the data formed a quasi-scale. Question 7 : "No-one is better qualified to say what is right for a young person than his parents", was omitted during the course of the analysis as it obviously would not scale with the other questions. Inspection of the content indicates that other factors are introduced by this question as it is relative. A subject could think that obedience is extremely important and feel that children should not question their parents' authority and yet, quite logically, disagree with this statement. For instance, the subject may feel that a teacher is better qualified than a parent to say what is right for a young person. Scale analysis of these 10 questions, then, indicated that one main dimension was being studied, but that a number of specific factors had been introduced by poor question wording.

A closer inspection of the content suggested that some other statements were relative. As the aim was to study attitudes towards obedience, rather than the evaluation of the relative importance of specific traits and behaviours in children, the irrelative statements were selected for further analysis. These were:

2. Parents should always be obeyed.
3. Even if young people know their parents are in the wrong they should do as they are told.

TABLE 23

Distribution of Sample Variables I : Distribution of (i) age, (ii) sex, (iii) marital status, (iv) education, (v) religion and (vi) place of birth in pilot study samples. Sample A; N = 92: Sample B; N = 99.

Distribution	Percentages	
	Sample A	Sample B
<u>(i) Age</u>		
Under 15	1.1	6.0
15 - 24	19.6	16.2
25 - 34	17.4	14.1
35 - 44	6.5	13.1
45 - 54	16.3	18.2
55 - 64	19.6	12.2
65 and over	13.0	14.1
Information omitted	6.5	6.1
Total	100.0	100.0
<u>(ii) Sex</u>		
Male	51.1	49.5
Female	48.9	43.4
Information omitted	-	7.1
Total	100.0	100.0

TABLE 23 (continued)

Distribution	Percentages	
	Sample A	Sample B
<b>(iii) <u>Marital Status</u></b>		
Single	15.3	21.2
Married	71.7	62.6
Widowed	4.3	7.1
Divorced	-	-
Information omitted	8.7	9.1
Total	100.0	100.0
<b>(iv) <u>Education</u></b>		
Primary only	38.0	31.3
Secondary	47.8	57.6
Technical College	2.2	1.0
University	1.1	2.0
Information omitted	10.9	8.1
Total	100.0	100.0

TABLE 23 (continued)

Distribution	Percentages	
	Sample A	Sample B
<b>(v) <u>Religion</u></b>		
Christian		
Non-conformist	21.7	8.1
Church of England	53.3	61.6
Roman Catholic	8.7	13.1
Fundamentalist	1.1	4.0
Spiritualist	1.1	-
Non-Christian	-	1.0
Atheist, Humanist, no religion	5.4	1.0
Information omitted	8.7	11.2
Total	100.0	100.0
<b>(vi) <u>Place of Birth</u></b>		
Stoke-on-Trent or Newcastle-under-Lyme	84.8	84.9
Other parts of Staffordshire	4.4	2.0
England (ex. Staffs.)	4.3	4.0
British Isles (ex. England)	1.1	-
Outside British Isles	-	2.0
Information omitted	5.4	7.1
	100.0	100.0

4. Do you think young people should always do what their parents tell them to?
5. Young people sometimes get rebellious ideas but as they grow up they ought to get over them and settle down.
6. There are no circumstances under which a parent should be disobeyed.
9. A good parent is one who teaches his child obedience and respect for authority.
11. It is wrong for children to question the authority of their parents.

Questions 5 and 11 were then both judged to be less directly relevant to the issue than the other items and eliminated from the analysis for this reason.

Items 2, 3, 4, 6 and 9 were found to satisfy scale criteria with a coefficient of reproducibility of 0.97. It is interesting that, although Guttman warns against the use of scale analysis to select items (as an appropriate configuration of responses may be largely due to chance factors if this procedure is adopted) the same selection of items could have been achieved with reference to the scale pattern rather than to the content of items. Questions 1, 5 and 10 could have been introduced without increasing the error, but these items had high marginal frequencies and a pattern of error which indicated that their reproducibility could be spurious. The pattern of error also would have suggested the exclusion of item 11, had the items been selected "blind". That is, in this instance, scale analysis could have been used to select the most specific and relevant questions as these best satisfied scale criteria. The results of the analysis of the selected

items are reproduced from the scalogram board in Table 24.

It was not possible to plot the intensity component as the items as a whole formed only a quasi-scale. The intensity component of the selected items was a J curve. As there were only five dichotomized questions, the curve was defined by only six scale types and was too crude to locate, with any precision, an invariant zero.

The high marginal frequencies of the F scale items, statement 1, "Obedience and respect for authority are the most important virtues children should learn" and statement 5, "Young people sometimes get rebellious ideas but as they grow up they ought to get over them and settle down", were unexpected. Strong agreement with F scale items is thought to be indicative of atypical personality adjustment. However strong agreement with these items was the norm for this group, and only one subject actually disagreed with one of these items. Although it is premature to draw conclusions from the use of only two questions it is possible that the F scale would not be a useful tool for differentiating between people from this area.

As the scale constructed from the first pilot study contained only five items it seemed advisable to carry out a second pilot study. The order of selected items was retained and further statements were suggested by ambiguity or vagueness in existing questions. First there was an attempt to establish more exactly what subjects meant by "children". Did they mean children under 5 years of age, children under 12 years of age, teenage children or young people? Similarly, did they mean unquestioning obedience or did they mean that an issue

may be discussed first with children, an appeal to the principle of obedience being something to resort to only when a more flexible approach has failed? The items constructed are given in Appendix III (ii).

The sample was selected as for the first pilot study, and the response was as follows:

Study B. Number of streets visited : 5

Responding	Unassisted	96	99
	Assisted	3	
Not Responding	Non-Respondents	2	7
	Refusals	5	

Total: 106

A description of the sample is given in Table 23.

These items were found to satisfy scale criteria, and to have a coefficient of reproducibility of 0.96. Some items, however, had less evenly distributed marginals than others so marginal frequencies were the main consideration in selecting items for the main study. Items with fairly even marginals were retained, and four of the original five items were retained as these had twice been found to scale.

To examine reliability, the scales of items responded to by the two pilot study groups were compared. Theoretically the items

should yield comparable scale types when responded to by different samples from the same population. Fewer, or more, scale types could be expected in any sample, due to chance differences in sampling, but the ordering of items should not vary. The scales are reproduced from the scalogram board in Table 24 and Table 25. Despite minor changes in wording to give fewer response categories for items on form B, where scale analysis on form A indicated that the degree of differentiation was artificial, the ordering of items is identical. Form B better satisfies scale criteria in that the errors are fewer and more randomly distributed.

The two pilot studies suggested that a Guttman-type scale is a suitable instrument for measuring the attitudes of this type of subject. The people taking part in the study were not used to research and tended to resent the intrusion on their spare time. It seemed preferable to gain a small amount of valid information than to risk a high refusal rate, which tends to be selective and therefore to limit the generality of results, or poor rapport with subjects. The forms were therefore kept as short and unofficial looking as possible. The Guttman-type scale makes brevity possible as a larger number of items can be studied during pilot work, and a small sample of items can be used to represent these in the main investigation. The method of responding is simple. Where subjects cannot complete the questions easily on their own there does not seem to be a high degree of subjectivity involved when the interviewer reads out the questions and marks in the answers for the subject. No subjects said

TABLE 24

## Scalogram III : Scale Analysis of Five Guttman Scale

Items, based on the Responses of 92 Study A Subjects.

[illegible]

Coefficient of reproducibility = 0.96



that the questions were too repetitive, but comments were made indicating that subjects were pleased to be able to differentiate between questions and therefore to delineate clearly the meaning of their answers. Similarly, although the questions seem in general to be very categorical, and many subjects could be expected to disagree with them and point out exceptions, they seemed suitable for this group of people. The observed rates of agreement indicated that items were, if anything, not strong enough to differentiate between these people. The selection of items was to some extent on the basis of marginal totals to isolate items which differentiated between subjects such as to give a 50-50 distribution of agreement and disagreement with each item. This was deemed sufficient for the purposes of this investigation. A more sophisticated design would be the selection of items to represent specific cutting points in terms, for example, of different percentiles of the population.

## 2. Construction of the Semantic Differential

In the second pilot study the semantic differential was introduced. The concept OBEDIENCE was differentiated against the scales:

1. good/bad
2. gentle/violent
3. fast/slow
4. active/passive
5. sour/sweet

6. clean/dirty
7. weak/strong
8. large/small
9. masculine/feminine
10. useful/useless
11. unnecessary/necessary
12. valuable/worthless
13. unimportant/important
14. desirable/undesirable
15. wise/foolish

The scales were drawn from Osgood et al. (1957) to represent the evaluative factor (scales 1, 5, 6, 10, 11, 12, 13, 14, 15), the activity factor (scales 2, 3, 4) and the potency factor (scales 7, 8, 9). The instructions were adapted from Osgood et al. (1957) to make them as simple and brief as possible. The attention of subjects was drawn to the instructions so that they could read them and ask any questions about them before the forms were left with them.

It proved extremely difficult to administer the semantic differential on the spot, or help subjects with the task. Responses seemed to be biased by the presence of the interviewer. Subjects seemed to have difficulty in understanding how to proceed, and to wait for cues from the interviewer's tone before selecting responses. Differential response rates also reflect the difference in difficulty between the two tasks. 99 respondents returned completed Guttman scale forms but only 60 of these were able to go on to complete the semantic

differential. Metaphorical scales such as sour/sweet presented particular difficulty. Some respondents, with whom the task was discussed, became upset because they considered it is nonsense to ask whether OBEDIENCE is sweet rather than sour, masculine rather than feminine or large rather than small. Students, such as those acting as subjects for Osgood et al. (1957) are apparently quite content to use the central scale position to rate what they consider to be irrelevant scales, but with this group of people irrelevant scales seemed to constitute a considerable barrier to rapport and cooperation.

The data was analyzed in the usual way, using factor analysis. The D method described in Appendix II was used. The results of the factor analysis (I) are shown in Table 18. All scales have coordinates on dimension I, and the pattern of coordinates does not separate, blind, the evaluative activity and potency scales. If coordinates over 10 are scored +, and smaller coordinates are scored -, the pattern of loadings on the first factor is as follows:

1. (evaluative) +
2. (activity) -
3. (activity) -
4. (activity) -
5. (evaluative) -
6. (evaluative) +
7. (potency) -
8. (potency) -
9. (potency) -

- 10. (evaluative) +
- 11. (evaluative) +
- 12. (evaluative) +
- 13. (evaluative) +
- 14. (evaluative) +
- 15. (evaluative) +

It seems reasonable, then, to identify this dimension with the evaluative factor. The scale 5, sweet/sour, does not function as an evaluative scale according to this analysis.

The second dimension cannot as easily be interpreted. The pivot scale is weak/strong (in a positive direction, subjects felt obedience to be strong rather than weak), but the coordinates of other potency scales are lower than two of the activity scales. It is tempting, then, to identify this factor with a general dynamic factor. If this interpretation is valid the scale 2, gentle/violent, can be considered to act in a peculiar manner as it groups with the evaluative rather than dynamic scales. Considering the negative overtones of "violent" this is not surprising. Also scale 9, masculine/feminine is not highly loaded on this dynamic factor, making the identification of the factor extremely tentative, but again the association of obedience with masculinity or femininity is tenuous. The more relevant scales are to the concept, the clearer this factorial structure seems to be.

As more than one method of factor analysis could have been used the analysis was checked by means of an elementary linkage analysis

of the matrix of  $D^2$  scores (or squared distances) in the 15 dimensional space. The grouping of distances agrees closely with the factor analytic method. Two groups were isolated. Type I shows a reciprocal association between scales 10. useful/useless, 12. variable/worthless and 15. wise/foolish with a first order association of scales 1. good/bad, 6. clean/dirty, 11. unnecessary/necessary, 13. unimportant/important and 14. desirable/undesirable. This cluster can be identified with the evaluative factor. Scale 5. sour/sweet does not group with these scales but is associated with scale 8. large/small and scale 3. fast/slow. Other dynamic scales cluster with the evaluative scales as second and third order associations. It is interesting that clearest and most meaningful grouping of scales, that is of the highest reciprocal associations (or smallest distances between scales) is evolved by the most relevant scales. In terms of content it seems more logical to judge obedience in pragmatic terms such as useful/useless, valuable/worthless and wise/foolish rather than in terms of a good/bad dimension, or more particularly in terms of an activity or potency dimension. It may be that, for example, that scale 3. fast/slow appears to be related to scale 5. sour/sweet simply because they are both irrelevant to the concept being judged. These three most highly related scales are also the three with highest coordinates on dimension I on the factor analysis, and all have higher coordinates than the scale good/bad which usually defines the evaluative dimension.

The scales wise/foolish and valuable/worthless were selected for use in the main study as these had the highest coordinates on

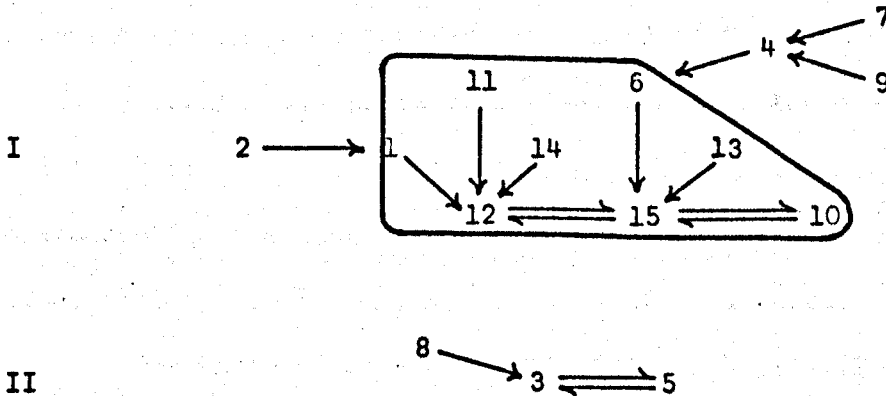
TABLE 26

Table II : Table of squared distances between semantic differential profiles of 60 Pilot Study B subjects on the concept OBEDIENCE.

	Scale														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Scale: 1	-	(51)	99	72	102	44	103	108	101	32	59	25	53	29	26
2	51	-	72	63	97	67	117	86	80	66	74	56	79	53	61
3	99	72	-	71	(53)	95	83	(65)	72	106	104	108	110	87	101
4	72	63	71	-	78	54	(73)	70	(65)	58	88	65	89	60	58
5	102	97	(53)	78	-	110	78	69	113	115	124	117	120	108	113
6	44	67	95	(54)	110	-	104	108	92	45	94	47	83	50	42
7	103	117	83	73	78	104	-	115	131	99	99	116	106	115	103
8	108	86	65	70	69	108	115	-	90	118	132	113	121	103	114
9	101	80	72	65	113	92	131	90	-	96	132	103	126	93	111
10	32	66	106	58	115	45	99	118	96	-	53	20	58	30	13
11	59	74	104	88	124	94	99	132	132	53	-	46	56	65	47
12	(25)	56	108	65	117	47	116	113	103	20	(46)	-	58	(25)	(13)
13	53	79	110	89	120	83	106	121	126	58	56	58	-	57	51
14	29	53	87	60	108	50	115	103	93	30	65	25	57	-	30
15	26	61	101	58	113	(42)	103	114	111	(13)	47	(13)	(51)	30	-

The smallest distance (i.e. highest association) in each column is ringed. Elementary linkage analysis [see McQuitty (1957)] gives the following types:

TABLE 26 (continued)



Scales interpreted as representing the evaluative factor are ringed.

dimension I. The scale good/bad was also selected as this conventionality defines the evaluative dimension or factor and to replicate this the scale evil/virtuous was added. The scales important/unimportant and interesting/uninteresting were also used as these seemed particularly relevant. Although all the selected scales seemed to be evaluative the factorial structure of scales 1. good/bad, 2. wise/foolish, 4. valuable/worthless and 5. important/unimportant only, was taken to be known. The added scales were primarily for the scale analysis of the data.

Unfortunately computer facilities were not available for the analysis of the results of these investigations. This severely limited the scope of the work, and the level of analysis applied. The level of accuracy, however, should not be affected as computations were carried out with all possible care, and the reproducibility of distances in the D method of factor analysis serves to check calculations.

Five, rather than seven scale intervals were used throughout to simplify the task for subjects. Similarly, scales were not rotated against concepts, but presented as successive judgements of the one concept, to give maximum simplicity, and comparability across studies.

### 3. Results of Pilot Studies

The empirical relationship between the two techniques, as administered to the 60 subjects who completed both the Guttman scale and the semantic differential was not clearly defined. Although  $\chi^2$  indicated that they are not independent when subjects were arbitrarily dichotomized according to their scale types on each technique

$\chi^2 = 7.07$  (1 d.f.) :  $p < .01$  , it was not possible to construct a joint scale of Guttman scale and semantic differential responses. Responses across the scales were not consistent enough to indicate the existence of even a quasi-scale.

#### 4. Empirical Comparison of the Two Techniques

The shortened form of the Guttman scale and the brief semantic differential, see Appendix III (iii), were embedded in a questionnaire which was administered to a sample of 1086 residents of the Stoke-on-Trent C.B. The sampling procedures, and methods of collecting data, used for pilot studies were employed. As 78 subjects failed to complete the semantic differential, the comparison of the two techniques was based on the responses of 1008 subjects. The improved response rate to the semantic differential may be attributed to the improved instructions. Improved response rates to the information items are, however, to some extent spurious. Subjects not responding to information items tended to omit the semantic differential, and their questionnaires were not, therefore, included.

Where possible, the sample characteristics were compared with those reported in the Sample Census of 1966, for the Stoke-on-Trent C.B. The census figures are based on a ten percent sample of the estimated population of 269,520 persons. Census figures are available for the variables age, sex, marital status and place of birth. The variables education and religion are not described, and the classification of occupation is for economically active and retired males. This is not

comparable with the classification of occupation used in this investigation to give a socio-economic rating of the subject's family.

The Main Sample is described in Table 27. Frequencies are expressed as percentages of the sample of 1008 persons completing both the Guttman scale and the semantic differential. To compare the Main Sample with the Census Sample category proportions for each variable had to be recalculated, as percentages of the number of subjects responding to each item. That is, the Census procedure does not allow for omissions, so this category of "Information omitted" had to be excluded from the Main Sample classifications to make these statistically comparable. It was considered desirable to retain this category in presenting the material to the reader in Table 27, as this retains all the information.

The Census frequencies were also expressed as percentages and the Main Sample and Census Sample were compared by means of a Chi square test.

In relation to the variables marital status and place of birth the two samples did not differ significantly.

In relation, however, to age and sex, there are differences between the two samples.

In the Main Sample there are proportionately more young people than would have been predicted from the Census Sample. A number of factors relate to this distribution of the variable age. First; the technique used to select the Main Sample was less precise than the procedure employed to select the Census Sample. Also the greater size

TABLE 27

Distribution of Sample Variables II : Distribution of (i) age, (ii) sex, (iii) marital status, (iv) education, (v) religion, (vi) place of birth and (vii) occupational classification of head of family in Main Sample.

N = 1008.

Comparative figures for the Stoke-on-Trent C.B. 1966 Sample Census.

N = 29,952.

Distribution	Percentages	
	Main Sample	Census Sample
(i) <u>Age</u>		
Under 15	9.9	22.9
15 - 24	20.7	15.5
25 - 34	20.9	11.5
35 - 44	17.7	14.0
45 - 54	13.1	13.6
55 - 64	9.3	12.5
65 and over	6.7	10.0
Information omitted	2.6	-
Total	100.0	100.0
(ii) <u>Sex</u>		
Male	41.7	48.9
Female	57.0	51.1
Information omitted	1.3	-
Total	100.0	100.0

TABLE 27 (continued)

Distribution	Percentages	
	Main Sample	Census Sample
<u>(iii) Marital Status</u>		
Single	34.1	40.2
Married	57.5	52.3
Widowed	4.3	7.0
Divorced	0.8	0.5
Information omitted	3.3	-
Total	100.0	100.0
<u>(iv) Education</u>		
Primary only	12.2	-
Secondary	63.0	-
Technical College	18.7	-
University	4.4	-
Information omitted	1.7	-
Total	100.0	

TABLE 27 (continued)

Distribution	Percentages	
	Main Sample	Census Sample
<u>(v) Religion</u>		
Christian		
Non-conformist	17.3	-
Church of England	50.5	-
Roman Catholic	16.2	-
Fundamentalist	11.9	-
Spiritualist	0.2	-
Non-Christian	0.1	-
Atheist, Humanist, no religion	0.8	-
Information omitted	3.0	
Total	100.0	
<u>(vi) Place of Birth</u>		
Stoke-on-Trent or Newcastle-under-Lyme	83.8	-
Other parts of Staffordshire	0.9	-
England (ex. Staffs.)	9.7	-
Total for England	94.4	94.4
British Isles (ex. England)	2.3	3.5
Outside British Isles	1.2	2.1
Information omitted	2.1	-
Total	100.0	100.0

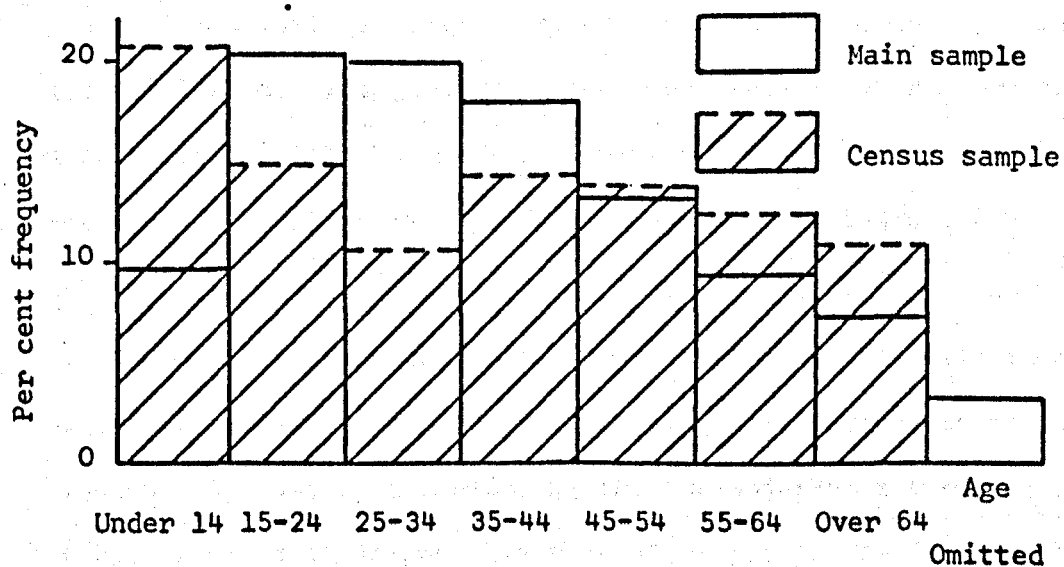
TABLE 27 (continued)

Distribution	Main Sample per cent.
(vii) <u>Occupational Classification of Subjects of Head of Family</u>	
Class 1      Professionally qualified, high administrative	4.5
2      Managerial and executive	10.0
3      Inspectorial, higher non-manual	10.2
4      Inspectorial, lower non-manual	6.1
5(a) Routine non-manual	10.8
5(b) Skilled manual	26.2
6      Manual, semi-skilled	14.5
7      Manual, routine	2.9
Information omitted or too vague to classify	14.8
Total	<hr/> 100.0

TABLE 28

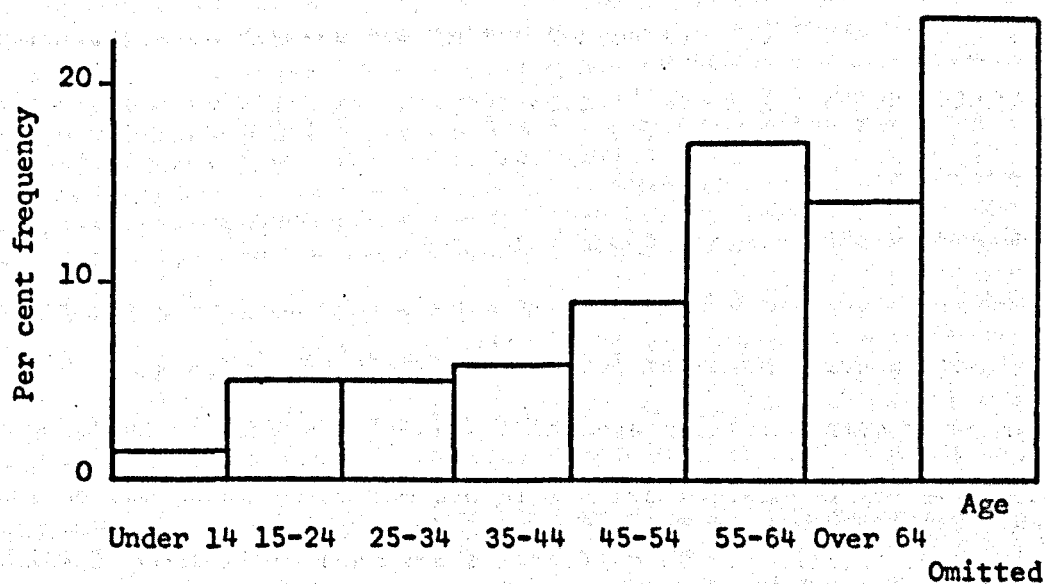
Graph I : Age Distribution of 1008 Main Sample Subjects Completing

Semantic Differential



Graph II : Age Distribution of 78 Subjects Failing to Complete

Semantic Differential



of the latter is a protection against sampling anomalies. Second; refusals may have given a selective bias to the sample. It was more difficult to secure the cooperation of older people. Thirdly; the sample analyzed includes only those subjects who completed the semantic differential form. Though many subjects who failed to complete the semantic differential also failed to give their age, a tally of the semantic differential non-respondents suggests that these are concentrated in the higher age groups. That is, it is probable that the 78 subjects who were unable to complete the semantic differential, and were thus excluded from the Main Sample, were largely in the upper age ranges. The age distribution of the 78 subjects is given in Table 28. Fourth; it is possible that older people are less forthcoming in revealing their ages than young people. If this were the case a misleadingly low number of subjects would be classified in the higher age ranges.

Similarly there are proportionately more females in the Main Sample than there are in the Census sample. This difference may also be due to selective refusal rates. Women seemed to welcome the social contact involved in participating in the research, while men more often regarded the request for cooperation as an intrusion on their spare time.

The Main Sample, then, is not representative of the population of Stoke-on-Trent. This does not invalidate the comparison of the two techniques: the semantic differential and the Guttman scale. While it does limit the generalizability of conclusions about the

content of the attitudes of people in the Stoke-on-Trent area, it does not affect the discussion of the measuring instruments. The design, used for the comparison of the two techniques resembles a matched group one. Each subject acts as his own control in the matching of groups.

The Cornell trial scoring technique was used to establish the best category combinations of the Guttman scale responses. The following combinations were found to minimize error:

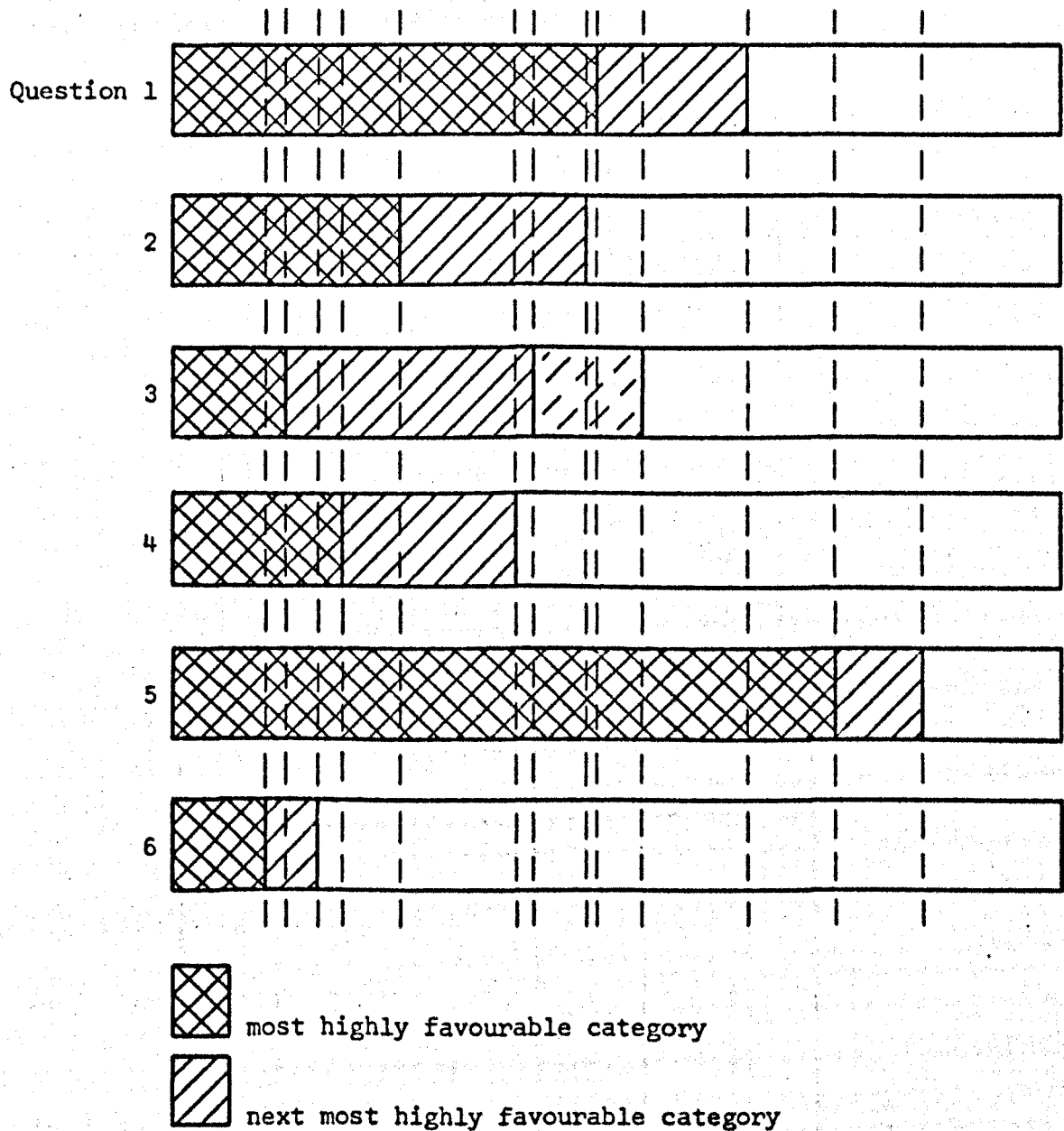
1. (3) (2,1) rescored (1) (0)
2. (3) (2,1) rescored (1) (0)
3. (4) (3,2.1) rescored (1) (0)
4. (3) (2,1) rescored (1) (0)
5. (3,2) (1) rescored (1) (0)
6. (3) (2,1) rescored (1) (0)

To save entering large numbers of repetitious scale types, the scale types were then directly counted and tabulated, as in the Goodenough technique. Errors were noted down as deviant types were classified.

A table of the ideal scale, which differs from the obtained scale pattern through error, and category combination to reduce error, is given in Table 29. The coefficient of reproducibility was found to be 0.96. The marginal frequencies, after category combination, were more extreme than is desirable. However, the pattern of errors was random, and in view of the previous analyses of the items, it seemed reasonable to treat the data as a scale.

TABLE 29

Scalogram V : Scale Pattern of Guttman Scale Data. N = 1008



Errors after category combination: (1), 14; (2), 46; (3), 61

(4), 53; (5), 4; (6), 42.

TABLE 30

Table of Frequency Distribution of Response I : 1008 main sample  
subjects' responses to Guttman Scale items.

(a) Distribution of Responses

	Response Category			
	4	3	2	1
Question 1	-	481	167	380
2	-	260	203	535
3	103	304	128	473
4	-	194	184	630
5	-	746	93	159
6	-	106	47	855

(b) Distribution of Response Types

Items						Score	Frequency
1	2	3	4	5	6		
1	1	1	1	1	1	1	75
1	1	1	1	1	0	2	32
1	1	0	1	1	0	3	57
1	0	0	1	1	0	4	94
1	0	0	0	1	0	5	215
0	0	0	0	0	1	6	424
0	0	0	0	0	0	7	107
							<hr/> 1008

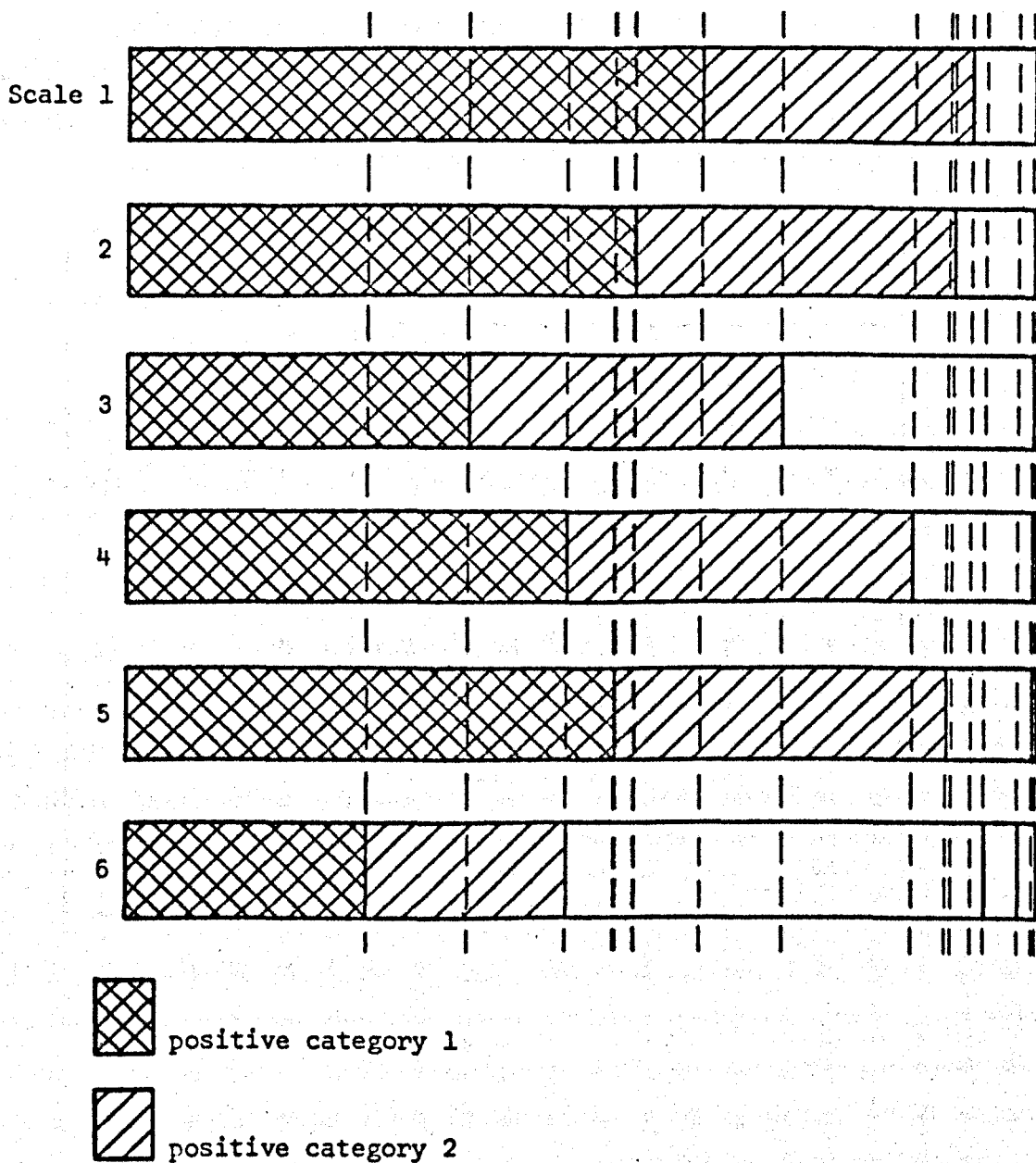
The scale analysis of the semantic differential proved to be more difficult. The large number of cases, and the amount of error involved make decisions about category combinations arbitrary. Categories tended to interlock with adjacent categories on either side. It seemed difficult to reduce the error without collapsing all of the categories of a question into one category, offering no differentiation between respondents. On scale 6, interesting/uninteresting, intervals 2 and 3 were combined, but other categories were left uncombined. As before the Goodenough tabulation technique was used. The ideal scale pattern, and the distribution of errors, is given in Table 31. As the data did not satisfy scale criteria, factor scores based on the four evaluative scales selected from the pilot study were used in the empirical comparison of the two techniques.

The Guttman scale score of each subject was tabulated against the evaluative scale score, based on scales 1, 2, 4 and 5, for each subject. The tabulation of Guttman scores against semantic differential scores is given in Table 33. Inspection of the table indicates that the techniques did not measure the same attitude in the same way, and that the techniques are not, in this case, interchangeable.

To express more precisely the relationship between the two techniques, as derived from the data presented in Table 33, it was necessary to employ only those statistical methods which do not make assumptions about the normal distribution of data. The distributions are extremely skewed. Strenuous efforts were made to normalize the

TABLE 31

Scalogram VI : Scale Pattern of Semantic Differential Data. N = 1008



Errors: (1), 118; (2) 131; (3) 145; (4) 146; (5) 162;  
 (6), 117 (combining categories 2 and 3)

TABLE 32

Table of Frequency Distribution of Response II : 1008 main sample  
subjects' responses on semantic differential scales.

	Scale Interval				
	1	2	3	4	5
Scale number 1	621	296	84	2	5
2	548	348	103	2	7
3	370	340	292	4	2
4	477	375	150	1	5
5	530	364	103	7	4
6	261	217	457	35	38
	2807	1940	1189	51	61

distributions, and all available transformations were applied without positive results. Neither could assumptions be made about the equality of intervals. There are no grounds for assuming that the Guttman scale scores have equal intervals. Both these conditions rule out the use of more powerful statistical methods, such as the use of a Pearson  $r$  to express a linear relationship, or the use of a correlation ratio to express a curvilinear relationship.

The use of a rank-order correlation coefficient, such as is used by Osgood et al. (1957) was not appropriate because of the large number of subjects used in this investigation. As the calculation of a rank-order correlation coefficient requires a minimum of tied scores, and scale analysis groups subjects into scale types, there is reason to suppose that it will be appropriate only under limited conditions. It was not thought desirable to select a small random sample of subjects, to allow the use of  $\rho$ . First; the use of a small number of subjects only would reduce the value of the research. Second; the rationale of random sampling requires that the sample be drawn from a population which is normally distributed.

For these reasons the coefficient of contingency was used as an index of association. The table of data was contracted by means of category combination to reduce the number of zero and near zero cell entries. The resulting tabulation of data is reproduced in Table 34. The Chi square test indicates that the two measures are related  $\chi^2 = 207.67$  (28 d.f.) :  $p < .01$  and that the coefficient of contingency ( $C = 0.41$ ) is significant.

TABLE 33

Table of Frequency Distribution of Response III : Guttman scale scores  
by semantic differential factor scores of 1008 main study subjects.

		Guttman Scale Score							S.D.
		1	2	3	4	5	6	7	Frequency
Semantic Differential	1.00	49	21	32	32	77	71	20	302
Evaluative Factor Score:	1.25	14	5	6	9	30	57	9	130
	1.50	3	4	6	18	34	53	9	127
	1.75	4	4	6	16	25	44	10	109
	2.00	4	0	3	8	24	95	28	162
	2.25	1	2	1	7	9	44	5	69
	2.50			2	2	7	21	4	36
	2.75					3	11	6	20
	3.00			1	2	4	24	9	40
	3.25					1	2		3
	3.50					1	2	2	5
	3.75								
	4.00								
	4.25							1	1
	4.50								
	4.75								
	5.00							4	4
G.S. Frequency		75	32	57	94	215	424	107	1008

TABLE 34

Table of Frequency Distribution of Responses IV : Guttman scale scores  
by semantic differential factor scores of 1008 main study subjects  
(after category combination).

Semantic differential evaluative factor scores	Guttman scale scores					S.D. Frequency
	1-3	4	5	6	7	
1.00	102	32	77	71	20	302
1.25	25	9	30	57	9	130
1.50	13	18	34	53	9	127
1.75	14	16	25	44	10	109
2.00	7	8	24	95	28	162
2.25	4	7	9	44	5	69
2.50	2	2	7	21	4	36
2.75 to 5.00	1	2	9	39	22	73
G.S. Frequency	168	94	215	424	107	1008

$$\chi^2 = 207.66$$

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} = 0.41$$

The results of this investigation suggest that the relation between the two techniques is not as close as was at first assumed. They tend to confirm those of Tittle and Hill (1967). Tittle and Hill (1967) report an association lower than the near identity between the two techniques suggested by Osgood et al. (1957).

To outline more specifically the relation between the two techniques, it was decided to treat both scales as interval scales. Responses to the semantic differential were plotted against responses to the Guttman scale in the same way as an intensity scale is plotted against a content scale. The resulting graph, reproduced in Table 35, suggests that a curvilinear rather than a linear relationship may be found to hold between the two measures. This is presented as a hypothesis to be investigated rather than as evidence for a conclusion. There is not a wide enough range of scale scores to allow irregularities to occur, and the curvature may be an artifact derived from the category combinations used.

In a similar manner, an attempt was made to map the intervals of one scale onto the intervals of the other scale. The deciles of each scale were obtained graphically and these were then plotted on joint axes. The resulting graph, in Table 36, shows the manner in which the scales differentiate at different parts of the continuum. The seven scale scores of the Guttman scale tend to differentiate between subjects marking the first two intervals of the semantic differential scale. This can be seen from the frequency table, Table 33, but is more clearly illustrated graphically.

TABLE 35

Graph IV : Relationship of Guttman Scale and Semantic Differential responses of 1008 main study subjects. The Semantic Differential test is plotted as if it were a further component of the Guttman Scale, on the basis of cumulative frequency of response.

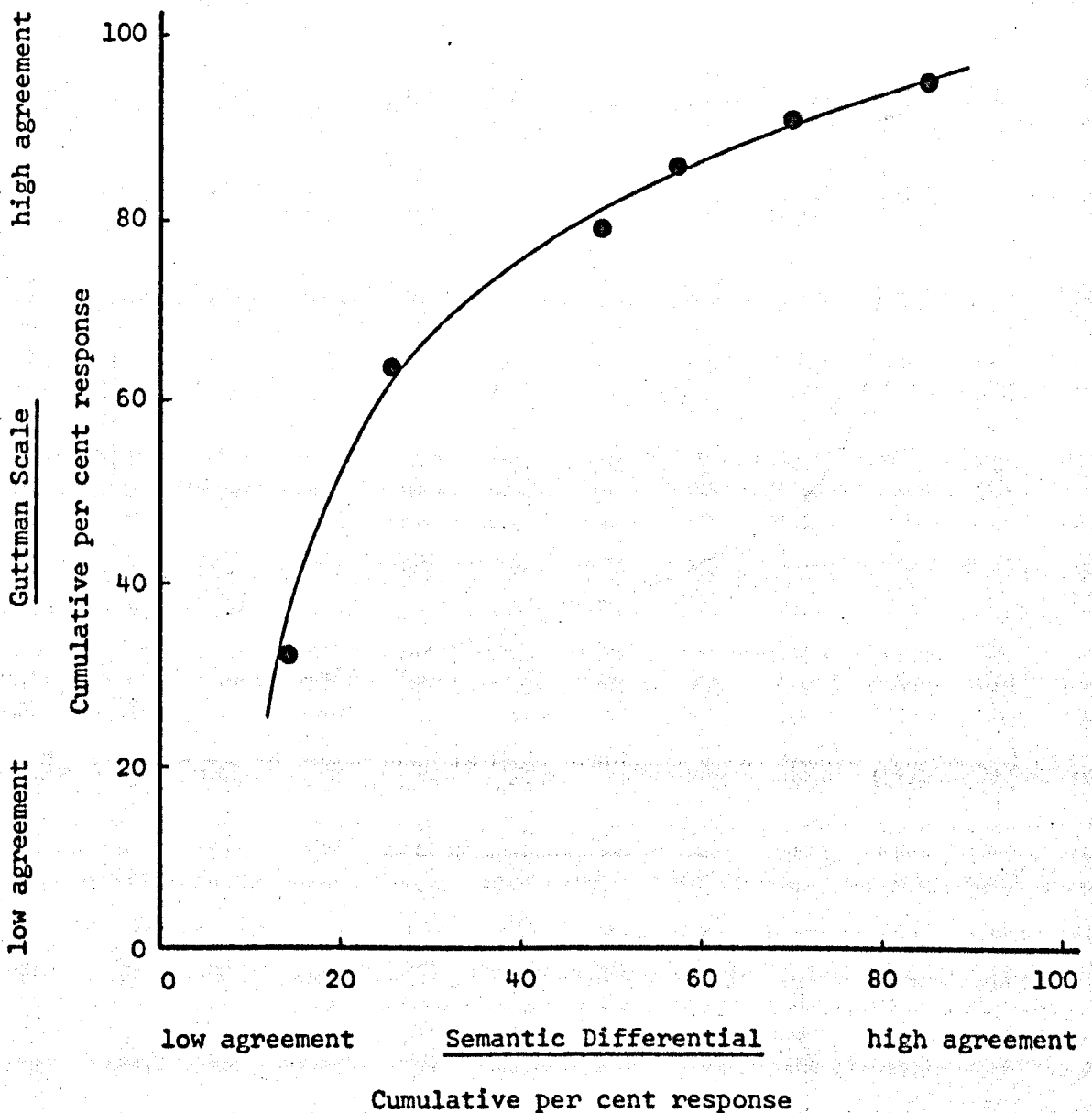
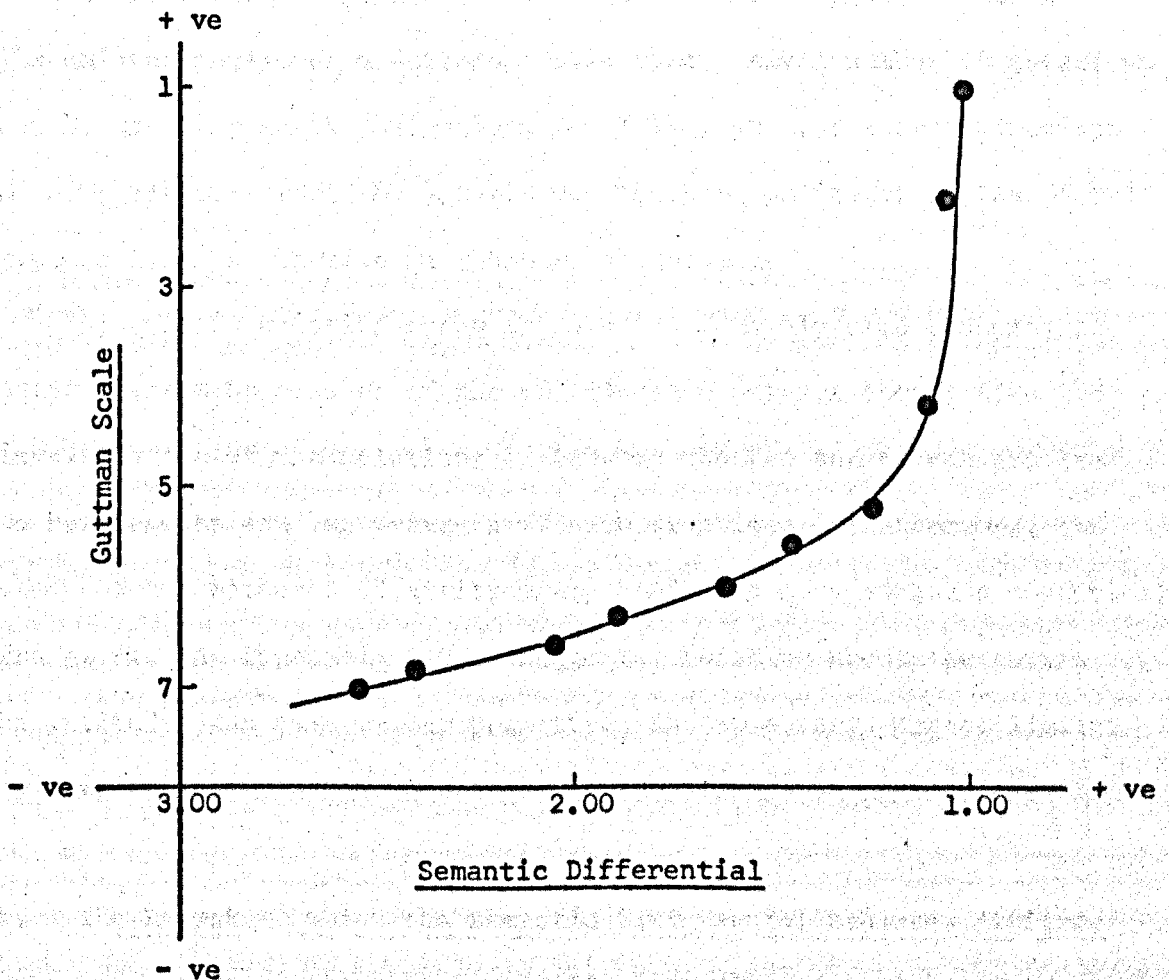


TABLE 36

Graph V : Relationship between Guttman Scale and Semantic Differential responses of 1008 main study subjects. The Guttman Scale intervals are mapped onto the Semantic Differential scale intervals by plotting graphically derived deciles.



To investigate the possible curvilinear relationship between the semantic differential and a Guttman scale it would be necessary to use a very large number of questions, and wise to employ a supplementary equal interval or equal-appearing interval procedure. Also, it would be necessary to select semantic differential concepts such that responses to them are normally distributed. The data would then be in a form to permit the determination of regression lines. The main advantage of a Guttman scale, that a small number of questions can be used for an investigation, would be lost, and the construction of both measures would be laborious. However, an investigation of this type would be worthwhile on theoretical grounds.

It is unsatisfactory to have two commonly used measures of attitude, and be unaware of the relationship between them. Also the investigation of the relationship between the two techniques can lead to refinements and improvements of each technique. Furthermore, the study of two measures of attitude may lead to a more adequate conceptual framework. As Oppenheim (1966) suggests, thinking about the nature of attitudes has been rather primitive, and to a large degree limited to simple linear models when these may not always be appropriate.

##### 5. The Extent to which the Semantic Differential measures Attitude

Attitudes have both cognitive and affective elements. A cognitive judgement, such as the statement "4 is greater than 1" is not an attitude statement. Similarly an affective expression may not contain the type of evaluation which normally characterizes an

attitude. For example, Lazarus and M<sup>C</sup>Cleary (1951) show that people may have an emotional response to nonsense syllables which have previously been associated with shock even when these syllables are not consciously recognized. As the stimuli produce behavioural reactions they have meaning, in terms of Osgood's definition of meaning as an  $r_m$ . However, Katz and Stotland (1959) suggest that

"The concept of attitude does not include such affective response without cognitive evaluation (p.429)."

The cognitive component may be minimal:

"Attitudes or evaluations thus have both an affective and cognitive component. The amount of cognition may be minimal; it need merely specify the object sufficiently for its recognition and relate the object to some evaluative standard. In addition, some attitudes may have a more elaborate cognitive component, including beliefs about the object, its characteristics and its relation to other objects, including the relation to the self (p.429)."

The semantic differential does measure attitude insofar as it specifies an affective object, and asks the subject to evaluate it against some standard, but the cognitive component may be minimal. In another type of attitude scale, there is an attempt to minimize the affective component. Terms such as FASCISM, COMMUNISM, DEMOCRATIC will be studiously avoided as a person agreeing with some of the related policies and ideas will react simply to the negative or positive associations of the terms. For example, the F scale asks questions such as "Obedience and respect for authority are the most important

virtues children should learn" rather than "Fascism is a good ideology" to study Fascist tendencies. Many people who agreed with Fascist ideas would not admit to being pro-Fascist as this is socially unacceptable in this culture. The relationship of the predominantly cognitive aspect of the attitude, to the affective value of the label may be expected to vary as some people, and some attitudes are more rational than others. The relationship between a graphic semantic differential profile of the concept FASCISM, to the F scale would possibly be very slight. As one of the main features of the F scale personality is thought to be a stress on social acceptability high F scale scorers could be expected to rate FASCISM as being negative evaluatively. That is, the name of a value system and the ideas that constitute it are two different things, and the individual's evaluation of the name may differ from his evaluation of the ideas. The semantic differential, as a rule, measures the evaluation of a name or symbol, whereas other attitude scales tend to evaluate the ideas and attitudes which are constituents of the value systems which have this name or label. If the label is not a highly emotive one, the constituent ideas and the label itself may be consistently structured by the individual, but if the label is a highly emotive one the individual's reaction to the label may not be indicative of his attitude to the ideas logically subsumed by the label. The relationship between a semantic differential and another attitude scale can thus be expected to be related to the content of both of these.

The semantic differential could be used during the construction of attitude scales to eliminate emotive terms, to ensure that the scale measures content rather than reactions to symbols. For example, for the people studied in this investigation "respect" seemed to be a very favourable word. It is possible that the high rate of agreement to the statement "Obedience and respect for authority are the most important virtues children should learn" was indicative of the favourable reaction to the word "respect", rather than of attitudes to obedience or respect for authority. Subjects could be asked to differentiate the concepts OBEDIENCE, RESPECT FOR AUTHORITY, RESPECT, A CHILD WHO RESPECTS OTHERS and so on, to investigate the affective components of the words.

A high correlation between a content scale and a semantic differential may not always be desirable, as it could indicate that the affective value of the symbols, rather than the content of the attitudes was being gauged by the content scale. For example Murphy, Murphy and Newcomb (1937) criticise Thurstone scales for the repetitive use of terms such as Communism and Church, and suggest that these scales may gauge a reaction

"not to the actual content of the attitude in question, but to the symbols which indirectly suggest it (p.903)."

This could be tested by comparing the results of subjects on the scales in question with their semantic differential profiles of the symbols in question. In fact, their criticism is supported by the fact that Osgood et al. (1957) report consistently high correlations between the differentiation of concepts such as CHURCH and NEGRO and appropriate Thurstone scales.

This is not to suggest that the semantic differential is a less valid measure of attitude than other attitude scales. The attitude to the symbol may be just as validly studied as may the attitudes to related ideas and issues. But it is a more fruitful approach to not assume they are identical until the relationship has been empirically investigated.

Soper and Menzies (1960) suggest that

"All studies of attitudes are inferential because of the mediation of symbols between the observer and the phenomena in which he is interested (p.42)."

It is useful to have a method which can study the properties of symbols in their own right, so that the effects of the symbols may be taken into account when the phenomena are investigated.

Klapper (1959) also suggests that the semantic differential taps emotional and non-conscious responses.

"It helps get around people's tendency to give well reasoned, logical, socially acceptable replies. It encourages intuitive, impulsive, emotional expression of reactions. Essentially, it may be regarded as a projective measure of somewhat the same order as sentence completions or free associations (p.437)."

While disagreeing that semantic differential responses are free from a social desirability factor, for as Ford and Meisels (1965) point out the semantic differential may measure this, it is possible to see the basis of Klapper's other comments. But if the semantic differential does encourage intuitive impulsive and emotional expression

of reactions, it should not be assumed that these will be correlated with well reasoned and logical statements of attitude that other attitude scales attempt to gauge. That is, even if the semantic differential and another attitude scale are constructed to measure the same attitude, avoiding as far as possible the use of highly emotive terms, the results may not be equivalent as the former attempts to tap the emotional while the latter attempts to measure the cognitive component of the attitude. If the cognitive component is very slight, the results may coincide and the dual approach would give information about the emotional nature of the attitude.

Thus, although the semantic differential does seem to measure attitude it measures a specific type of attitude, indicated by an avoidance-approach type of response, whereas the general definition of attitude is broader than this. The semantic differential does not necessarily measure the same type of attitude as another attitude simply because the name of the attitude object is common to both.

## 6. Correlation with Other Techniques

In attempting to correlate the semantic differential with other measures of attitude there is the problem of selecting concepts, such that the attitude object and not the symbol of the attitude object is studied. There is also the problem of selecting scales, as a standard set of scales which can be shown to offer a meaningful set of dimensions of judgement for any concept, without any scale-concept

interaction has not been established.

The attempt to correlate the semantic differential with other scales seems to be based on a rationale which is common in psychology. This is the view that if two sets of items are highly correlated they can be used as identical measuring instruments. There is some justification for this approach when the two sets of items are standardized, and the correlation between them is established over a wide sample of people. Even though they do not necessarily measure the same thing, if the variables they measure are always correlated, one can be used as an index of the other. However the semantic differential is not a standardized test. Rather it is a technique which may alter in composition (as may the scales it is compared with) and the plausibility of generalizing from one or two instances of an observed high correlation can be questioned.

Even if a semantic differential could be constructed such that the investigator could be confident that the two techniques were measuring the same component of the same attitude the results may not be highly comparable due to the particular distribution of responses. To take an extreme example the other scale may differentiate only between respondents who marked interval 1 on the semantic differential. Or, if the other scale has unequal scale intervals the relationship between the two techniques may be blurred by the arbitrary classification of responses into a small number of scale intervals on the semantic differential. That is, the continua generated by the two methods, each measuring the same variable may have different scale intervals, and

may not locate a common zero. Only a small section of the basic continuum may be delineated by each set of data.

In considering the correlation between the semantic differential and the Guttman scale a further complexity arises. Guttman scales, drawn from the same universe of content may not have a high biserial correlation with one another. Theoretically they should give the same ordering of individuals, but they may differentiate at different parts of the continuum, and the grouping of individuals may be so different as to obscure the relationship. They could, of course, be put together to form a joint scale but, as is the case with individual items, they are both functions of the scale variable rather than of each other. The tetrachoric coefficient, which expresses the correlation between two quantitative variables of which the items or sets of items are functions would be unity if a bivariate normal distribution could be assumed, but one of the aims of scale analysis is to avoid making untested hypotheses about normal distributions. It is conceivable, then, that a semantic differential could correlate with one scale or set of items, and not with the other, even though both scales are equally valid.

In using the semantic differential as a measure of attitude, then, its equivalence with other attitude scales should not be assumed in advance. Osgood et al. (1957) report high correlations with other techniques, but few cases or attitude objects are sampled. Tittle and Hill (1967) in a more thorough study of the relationship between measures of attitude express surprise that a lower correlation between

the semantic differential and other techniques is observed. A consideration of the rationale of the semantic differential suggests that an observed correlation will be a function of the particular instruments employed, in relation to their content and metric properties, and cannot be easily generalized. In this investigation the instruments were constructed with the aim of making them as directly comparable as possible, in an attempt to repeat Osgood's comparison and identification of the two techniques, with a more general attitude object and larger sample of respondents, but the two techniques by no means gave equivalent results. As Osgood suggests himself the semantic differential does not tap much of the content of an attitude, that is the

"specific reactions which people having various attitudes might make, the specific statements they might accept (Osgood et al. (1957) p.195)."

One wonders, then, why he tries to show that the semantic differential is correlated with techniques which do attempt to gauge specific reactions, and responses to specific statements, to establish the validity of the semantic differential.

## CHAPTER VIII

### THE STRUCTURE OF SEMANTIC DIFFERENTIAL DATA

#### 1. The Evaluative Factor : Some Qualifications

The use of both the semantic differential and a Guttman type scale to study attitudes to obedience suggested that the techniques are not unrelated, but are also not, for practical purposes, identical. The techniques seemed rather to supplement one another.

Scale analysis gives a very precise definition of linearity, and enables the relationships between items to be studied with clarity. For example, the scale pattern of Questionnaire B was of particular interest. The question "Parents should always be obeyed" is a generalization with which many respondents agreed. However, when a possible exception was pointed out by the questionnaire in the form "Even if young people know they are in the wrong they should do as they are told", many respondents made this exception to the generalization. The question "There are no circumstances under which a parent should be disobeyed" was included in an attempt to repeat the generalization. Had the original question been repeated, independence of responses could not be guaranteed. Unfortunately, the altered form involves the use of a double negative which is undesirable as respondents may have misunderstood it. Any conclusions cannot be drawn from it for this reason. However, fewer respondents agreed with this item, than with the original item and it is tempting to postulate that an attitude

questionnaire may not only measure attitudes, but also help to create them. It may be that in pointing out exceptions to generalizations the questionnaire tends to make people less liable to hold them. The scale analysis of these questions indicated that a design which allows for the rotation of items may give valuable results.

Similarly scale analysis indicated that the idea of obedience is not related to that of age in a linear fashion. For example, although a large number of respondents agreed that "Children under 12 years of age should always obey their parents" and far fewer believed that teenage children should do so, indicating the lessening need for obedience with age, the majority of respondents felt that children under 5 years of age should not be expected to be obedient. Respondents, then, seemed to think that obedience is important mainly for the 5 to 12 year old child.

Similarly scale analysis studies the relationship between semantic differential scales, when applied to these. It was postulated, when the semantic differential for the main study was constructed, that evil is more negatively evaluative than bad, and that virtuous is more positively evaluative than good. That is, that the scale evil/virtuous is an evaluative scale of greater length than the scale good/bad which conventionally defines the evaluative factor. In functional terms it was postulated that the term good would be used more liberally than the term virtuous, and that the term evil would be used more sparingly than the term bad. The relationship of bad to evil could not be examined as the concept OBEDIENCE was rated positively by the majority

of subjects. The scale analysis of semantic differential data, shown in Table 31, supported this hypothesis in relation to the positive defining terms. Some respondents rated OBEDIENCE as both extremely good and extremely virtuous. However many rating the concept extremely good thought it only slightly virtuous (and not vice versa). Similarly many respondents rating OBEDIENCE as slightly good, rated the concept as not at all virtuous. Scale analysis suggests that the two scales are linear, though there is some error, but do not have equivalent cutting points or scale intervals. Factor analysis, as it assumes equal intervals would need to express the differences in response in terms of a further dimension.

The scale interesting/uninteresting also presents some difficulty. The scale analysis indicates that it has less error than other scales, and that it is of the same dimension as the other scales. Inspection of the content suggests that it is an evaluative scale, but that it is more extreme than "good" or "necessary". As it is logically quite possible that one should think obedience necessary but uninteresting the factor analytic interpretation, in terms of an extra dimension may be more useful.

The other evaluative scales, foolish/wise, valuable/worthless, and important/unimportant also seem to differ from the good/bad scale. They seem more relevant to the concept. It does not seem particularly meaningful to judge obedience as a good or bad think in an absolute way as obedience seems for most people to be a means rather than an end in itself. A child is taught obedience for his own safety, and for the

benefit of those around him, not for the sake of obedience itself.

It could be expected then that scales expressing a pragmatic evaluation of the concept, would be more relevant and hence more evaluative than good/bad. In a general way it might be said that although all concepts can be judged against an evaluative dimension, this might be defined by different terms. For example, it might be a more positive evaluation to describe a person as "interesting" than "good" or "nice", depending on the outlook of the person doing the describing. On the other hand if the concept is a moral one such as ABORTION, the term "good" may be more positively evaluative than "necessary" or "interesting".

Osgood et al. (1957) admit there is some concept-scale interaction, but feel this is the exception rather than the rule for in postulating a semantic space they make some claim to invariant relationships between scales.

The factor analysis of the semantic differential scales, Factor Analysis III, Table 37, is extremely interesting. Like the scale analysis, factor analysis indicates the presence of one major factor and a number of specific factors. The influence of the content is evident. The scale evil/virtuous and the scale interesting/uninteresting which were thought to have different scale intervals, or in the case of the latter even to be to some extent independent, have the most complex factor structure. That is, the scales with more equal intervals, as judged on an a priori basis, tend to be selected by the analysis on the basis of having the purest evaluative factor loadings. However, these scales, good/bad, foolish/wise, valuable/worthless and important/

TABLE 37

Factor Analysis III : Coordinates of scales, based on the responses of 1008 main sample subjects to the concept OBEDIENCE.

As the D-method of factoring begins with raw scores, the loadings or coordinates can be greater than 1.

Dimensions	Semantic Differential Scales					
	1	2	3	4	5	6
I	51.90	55.82	61.03	54.63	52.54	73.65
II	0	15.41	9.80	8.45	5.50	4.05
III	0	0	9.53	19.21	6.19	12.66
IV	0	0	17.97	0	21.91	6.44
V	0	0	22.76	0	0	0.17
VI	0	0	0	0	0	341.1

unimportant do not have the largest coordinates on the evaluative dimension.

The factor analysis of this data indicates the importance of the content of scales and their relevance to the concept to be differentiated. The choice of scales which seem relevant to the concept, and to have poles of comparable intensity may lead to a clearer factorial structure of data, than the use of scales such as good/bad which are widely applicable to many concepts, but with varying degrees of relevance.

To locate an object on an evaluative dimension is not simply to say whether it is good or bad. For example, when the attitude "Jews are weak" is expressed certain negative overtones are evident. As weakness and strength are potency rather than evaluative terms Osgood could hardly admit that this is a statement of attitude at all, except on the assumption that all people who agree with the statement would tend to rate the concept JEWS as bad, dirty and so on. In this case he would say that the potency and evaluative factors align. Not surprisingly he suggests that a full profile of meaning, including activity and potency scores, may be a better predictor of behaviour than a simple evaluative score. But the question remains as to whether activity and potency scales may also measure attitudes. For example, many people are against pacifism, not because it is bad, but because it is weak. Similarly many people consider Christianity to be good, but consider it ineffectual or irrelevant to modern life. In differentiating the concept PACIFISM scale terms such as weak and unrealistic

may better define the dimension of judgement than good/bad, clean/dirty, sweet/sour which are the more conventional evaluative scales.

Similarly, respondents with a negative attitude to CHRISTIANITY may think it is good, clean and sweet, but inadequate, inappropriate, ineffectual and unrealistic. That is, the negative evaluation may be largely related to its lack of potency.

It may be simpler to think of individual concepts generating their own evaluative dimensions, than to try and establish a general evaluative dimension, which is common to all concepts but sometimes aligns with other factors and acts in an exceptional manner. Many sets of concepts may be found to have similar judgemental frames of reference, giving comparability across related concepts.

## CHAPTER IX

### DISCUSSION AND CONCLUSIONS

This investigation related, both theoretically and empirically, scalogram analysis and the semantic differential test. Comparable forms of each technique were constructed in an attempt to replicate the study reported in Osgood et al (1957) which indicated a correlation of 0.78 between a semantic differential dealing with farmers' attitudes to CROP ROTATION, and a 14 item Guttman scale on the same topic, when responded to by 28 subjects. In this research a more substantial sample of 1008 subjects was used, and a more general attitude object, OBEDIENCE, was considered. The coefficient of contingency relating the two scales was found to be 0.41. This lower level of association accords with that of 0.52 reported by Tittle and Hill (1967) in their investigation of student political activity.

Emphasis was given to theoretical considerations, as the relationship between the two techniques does not seem to be invariant. It was suggested that further studies relating the two scales will discover differing degrees of association between them, and that the relationship has not yet been shown to be linear. In any given application of the two techniques the degree of association will be contingent on a number of related factors: content, symbolization, the nature of the attitude and the attitude object, and the metric properties of the scales.

The selection of concepts for a semantic differential form is subjective. Similarly, although there is some standardization of scales, scale-concept interaction is an unknown quantity. It has not been shown that the conventional evaluative scales are equally applicable, or exhaustive, for any given concept. In the same way, the sampling of Guttman scale items involves some subjectivity. The scaling procedure guarantees that only one continuum is being studied, but it cannot indicate that it is the same continuum that is being measured by the semantic differential.

The affectivity of symbols is also a probable determinant. The use of highly affective symbols in both the Guttman scale and the semantic differential, may lead to the observation of a high degree of association between the two techniques insofar as both measure the reaction to the connotative associations of the symbols. Neither would measure the attitude towards the designated object. More information could be obtained by using the techniques appropriately: the semantic differential to study the affective reaction to the symbols, and the Guttman scale to study the content of the attitudes.

The nature of the attitude and the attitude object studied may also influence the observed relationship between scalogram analysis and the semantic differential. If the attitude object is such that attitudes towards it can be adequately represented by a simple approach-avoidance model, the semantic differential may be used, and the additional use of a Guttman scale would yield little further information. On the other hand, the attitude object may be a value system or ideology,

and a Guttman scale will be more appropriate. The joint use of scales may indicate how far the value systems or ideologies are merely rationalizations of underlying approach-avoidance tendencies, or how far they are rationally based.

The metric properties of a Guttman scale and a semantic differential may also affect the obtained correlation between them. There is a particular problem here as the rationale of scale analysis indicates that correlation techniques cannot always be suitably applied to Guttman scales. A rank order correlation, such as is used by Osgood et al. (1957) can only be applied when the number of subjects is very small.

The relationship between the semantic differential, and other measures of attitude, is more complex than Osgood et al. (1957) assume. In this investigation, the evaluative factor of the semantic differential was considered as a measure of attitude and compared with a Guttman scale. But it is sometimes difficult to judge when the semantic differential does measure attitude, and when it does not do so. For example, scales which are in one context evaluative, and indicative of an attitude, may, in another context, be denotative. The rating of a concept, such as MY MOTHER, on the scale sweet/sour, may be indicative of an attitude. The rating of another concept, SUGAR, on the same scale, seems not to be indicative of an attitude. The connection between the scale and the concept, here, is inherent in the definition of the words; if a subject suggests that sugar is sour, he is mistaken about the meaning of the words. A statement of fact, rather than a statement of attitude, is involved.

Similarly, scales which are usually associated with factors other than the evaluative factor, may produce responses which seem to be indicative of attitude. For example, the rating of PACIFISM on the scale weak/strong, and the rating of the concept JOHN SMITH on the scale active/passive, would seem to be directly comparable with attitude statements such as "Pacifism is weak" or "John Smith is passive", which could be elements of a Guttman scale.

There is no evidence to suggest that the relationship between a Guttman scale and a comparable semantic differential form is a simple linear one which can be expressed neatly in terms of an invariant index of association. Although it is possible that under some conditions they may, to a large extent, measure the same thing; the techniques should not be used, in a practical situation, as if they were interchangeable.

If the aim of an investigation is to describe the properties of an attitude object, rather than to differentiate between subjects on the basis of the attitudes they hold, the semantic differential is probably the most useful tool available. However, the results of this investigation indicate that research workers should keep in mind the limitations of the semantic differential, and not use it as if it were a more conventional attitude scale such as one constructed by means of scalogram analysis.

In conclusion, then, the theoretical and empirical comparison of scalogram analysis and the semantic differential test indicates

that they have different properties, qualities and functions.

Although Osgood suggests a Guttman scale and the semantic differential can be used to measure the same thing to a considerable degree, they should not in practice be used as if they are identical.

BIBLIOGRAPHY

- Adorno, J.W., Frenkel-Brunswick, E., Levenson, D.J. and Sanford, R.N.  
The Authoritarian Personality, New York : Harper, 1950.
- Allport, G.W. Attitudes. In Murchison, C. A Handbook of Social Psychology, Worcester, Mass. : Clarke University Press, 1935
- Blalock, H.M. and Blalock, A.B. Methodology in Social Research, New York : M<sup>C</sup>Graw Hill, 1968.
- Bogardus, E.S. Measuring social distance. Journal of Applied Sociology, 1925, 9, 299-308.
- Bogardus, E.S. A social distance scale. Sociological and Social Research, 1933, 17, 265-271.
- Bray, D.W. The prediction of behaviour from two attitude scales. Journal of Abnormal and Social Psychology, 1950, 45, 64-84.
- Burt, C. Scale analysis and factor analysis. British Journal of Statistical Psychology, 1953, 6, 5-23.
- Clark, K.E. and Kriendt, P.H. An application of Guttman's new scaling techniques to an attitude questionnaire. Educational and Psychological Measurement, 1948, 8, 215-223.
- Coombs, C.H. Theory and methods of social measurement. In Festinger, L. and Katz, D. (Eds.) Research Methods in the Behavioural Sciences, Chicago : Holt, Rinehart & Winston, 1953.
- Corey, S.M. Professed attitudes and actual behaviour. Journal of Educational Psychology, 1937, 38, 271-280.
- Defleur, M.L. and Westie, F.R. Verbal attitudes and overt acts : an experiment on the salience of attitudes. American Sociological Review, 1958, 23, 669-673.
- Edwards, A.L. Techniques of attitude scale construction, New York : Appleton-Century-Crofts, 1957.

- Edwards, A.L. and Kilpatrick, F.P. Scale analysis and the measurement of social attitudes. Psychometrika, 1948, 13, 99-114.
- Eysenk, H.J. Review by H.J. Eysenk. Measurement and Prediction. A discussion of Volume IV of Studies in Social Psychology in World War II, Part I. International Journal of Opinion and Attitude Research, 1951, 5, 95-102.
- Eysenk, H.J. The psychology of politics, London : Routledge and Kegan Paul, 1954, reprinted 1963.
- Festinger, L. The treatment of qualitative data by scale analysis. Psychological Bulletin, 1947, 44, 149-161.
- Ford, L.H. and Meisels, M. Social desirability and the semantic differential. Educational and Psychological Measurement, 1965, 25, 465-475.
- Ford, R.N. A rapid scoring procedure for scaling attitude questions. Public Opinion Quarterly, 1950, 14, 507-532.
- Goodenough, W.H. A technique for scale analysis. Educational and Psychological Measurement, 1944, 4, 179-190.
- Green, B.F. A method of scalogram analysis using summary statistics. Psychometrika, 1956, 21, 79-89.
- Green, R. and Goldfried, M.R. On the bipolarity of semantic space. Psychological Monographs. General and applied, 1965, 79 (6, Whole number 599), 1-31.
- Guilford, J.P. Fundamental statistics in psychology and education, New York : M<sup>C</sup>Graw-Hill, 1950.
- Guttman, L. The quantification of a class of attributes. In Horst, P. et al., The prediction of personal adjustment, New York : Social Science Research Council, 1941.
- Guttman, L. A basis for scaling qualitative data. American Sociological Review, 1944, 9, 139-150.

- Guttman, L. On Festinger's evaluation of scale analysis. Psychological Bulletin, 1947, 44, 451-465.
- Guttman, L. The Cornell technique for scale and intensity analysis. Educational and Psychological Measurement, 1947, 7, 247-280.
- Guttman, L. Chapters 2,3,6,8,& 9 in Stouffer, S.A. et al. Measurement and prediction. Studies in social psychology in World War II, Vol. IV Princeton, New Jersey : Princeton University Press, 1950.
- Guttman, L. Scale analysis, factor analysis and Dr. Eysenk. Measurement and prediction. A discussion of Volume IV of studies in social psychology in World War II, Part II. International Journal of Opinion and Attitude Research, 1950, 5, 102-120.
- Guttman, L. The principal components of scalable attitudes. In Lazarsfeld, P.F. (Ed.) Mathematical thinking in the social sciences, Glencoe, Illinois : The Free Press, 1954.
- Guttman, L. An outline of some new methodology for social research. Public Opinion Quarterly, 1954, 1, 31-55.
- Hall, J. and Jones, C.O. Social grading of occupations. British Journal of Sociology, 1950, 1, 31-55.
- Harper, R. The measurement of meaning in personality research. Unpublished doctoral dissertation. University of Edinburgh, 1963.
- Hickman, C.A. and Kuhn, M.H. Individuals, groups and economic behaviour, New York : Dryden Press, 1956.
- Kahn, L.A. and Bodine, A.J. Guttman scale analysis by means of IBM equipment. Educational and psychological measurement, 1951, 11, 298-314.
- Katz, D. and Stotland, E. A preliminary statement to a theory of attitude structure and change. In Koch, S. (Ed.) Psychology : a study of a science, Vol. III. Formulations of the person and the social context, New York : M<sup>C</sup>Graw-Hill, 1959.

- Klapper, J.T. The semantic differential : its use and abuse. Public Opinion Quarterly, 1959, 23, 435-438.
- Kofksy, E. A scalogram study of classifactory development. In Sigel, J. and Hooper, F. (Eds.) Logical thinking in children, New York : Holt, Rinehart and Winston, 1968.
- Kutner, B., Wilkins, C. and Yarrow, P.R. Verbal attitudes and overt behaviour involving racial prejudice. Journal of Abnormal and Social Psychology, 1952, 47, 647-652.
- La Piere, R. Attitudes versus actions. Social Forces, 1934, 14, 230-237.
- La Piere, R. Type-rationalizations of group antipathy. Social Forces, 1936, 15, 232-237.
- Lazarus, R.S. and M<sup>C</sup>Cleary, R.A. Autonomic discrimination without awareness. Psychological Review, 1951, 58, 113-122.
- Lazarsfeld, P.F. Latent structure analysis. In Koch, S. (Ed.) Psychology: a study of a science. Vol. III. Formulations of the person and the social context, New York : M<sup>C</sup>Graw-Hill, 1959.
- Likert, R. A technique for the measurement of attitudes. Archives of Psychology, 1932, 22, 1-55.
- Likert, R., Roslow, S. and Murphy, G. A simple and reliable method of scoring the Thurstone attitude scale. Journal of Social Psychology, 1935, 5, 228-238.
- Linn, L.S. Verbal attitudes and overt behaviour: a study of racial discrimination. Social Forces, 1965, 43, 353-364.
- M<sup>C</sup>Nemar, Quinn. Opinion attitude methodology. Psychological Bulletin, 1946, 43, 289-374.
- M<sup>C</sup>Quitty, L.L. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. Educational and Psychological Measurement, 1957, 17, 207-229.
- Marder, E. Linear segments: a technique for scalogram analysis. Public Opinion Quarterly, 1952, 16, 417-431.

- Marshall, T.H. Sociology at the crossroads, and other essays, London : Heineman books on sociology, 1963.
- Messick, S.J. Metric properties of the semantic differential. Educational and Psychological Measurement, 1957, 17, 586-593.
- Murphy, G., Murphy, L.B. and Newcomb, T.M. Experimental Social Psychology, New York : Harper and Brothers, 1937.
- Nettler, G. and Golding, E.H. The measurement of attitudes towards the Japanese in America. American Journal of Sociology, 1946, 52, 31-39.
- Neuhaus, J.O. and Wrigley, C.F. The quartimax method, an analytic approach to orthogonal simple structure. British Journal of Statistical Psychology, 1954, 7, 81-91.
- Noland, E.A. Factors associated with absenteeism in a south central New York State industry. Unpublished doctoral dissertation, Cornell University, 1944. Reported in Kahn, L.A. and Bodine, A.J. Guttman scale analysis by means of IBM equipment. Educational and Psychological Measurement, 1951, 11, 298-314.
- Noland, E.A. Worker attitude and industrial absenteeism: a statistical appraisal. American Sociological Review, 1945, 10, 503-510.
- Nunnally, J.C. Psychometric theory, New York : M<sup>C</sup>Graw-Hill, 1967.
- Oppenheim, A.N. Questionnaire design and attitude measurement, London : Heinemann Educational Books Ltd., 1966.
- Osgood, C.E. The nature and measurement of meaning. Psychological Bulletin, 1952, 49, 197-237.
- Osgood, C.E. and Luria, Z. A blind analysis of a case of multiple personality using the semantic differential. Journal of Abnormal Psychology, 1954, 49, 579-591.
- Osgood, C.E., Suci, G.J. and Tannenbaum, R.H. The measurement of meaning, Urbana, Illinois : University of Illinois Press, 1957.

- Osgood, C.E. Studies on the generality of affective meaning systems. American Psychologist, 1962, 17, 10-28.
- Pace, C.R. Opinion and action: a study in validity of attitude measurement. American Psychologist, 1949, 4, 242.
- Poppleton, P.K. and Pilkington, C.W. A comparison of four methods of scoring an attitude scale in relation to its reliability and validity. British Journal of Social and Clinical Psychology, 1964, 3, 36-39.
- Riley, M.W., Riley, J.W. and Toby, J. Sociological studies in scale analysis. New Brunswick, New Jersey : Rutgers University Press, 1954.
- Rogers, H.W. Some attitudes of students in the R.O.T.C. Journal of Educational Psychology, 1935, 26, 291-306.
- Sample Census 1966, England and Wales. County Report, Staffordshire. London : Her Majesty's Stationary Office, 1966.
- Shapiro, G. Myrdal's definitions of the south : a methodological note. American Sociological Review, 1948, 13, 619-621.
- Shuessler, K.F. A note on the statistical significance of the scalogram. Sociometry, 1961, 24, 312-318.
- Slater, P. Weighting responses to items in attitude scales. British Journal of Statistical Psychology, 1956, 9, 41-48.
- Sofer, C. and Menzies, I. Problems and opportunities in inferring attitudes. In Attitude Scaling, London : Market Research Society, 1960.
- Solomon, L.N. A factorial study of complex auditory stimuli (passive sonor sounds). Unpublished doctoral dissertation, University of Illinois, 1954. Reported in Osgood, C.E., Sici, G.J. and Tannenbaum, R.H. The measurement of meaning, Urbana, Illinois : University of Illinois Press, 1957.

- Stagner, R. and Osgood, C.E. Impact of war on a nationalistic frame of reference: I Changes in general approval and qualitative patterning of certain stereotypes. Journal of Social Psychology, 1946, 24, 187-215.
- Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarsfeld, P.F., Star, S.A. and Clausen, J.A. Measurement and prediction. Studies in social psychology in World War II, Vol. IV. New Jersey : Princeton University Press, 1950.
- Suchman, E.A. The scalogram board technique for scale analysis. In Stouffer, S.A. et al. Measurement and prediction. Studies in social psychology in World War II, Vol. IV. New Jersey : Princeton University Press, 1950.
- Thurstone, L.L. Scales for the measurement of social attitudes, Chicago : University of Chicago Press, 1931.
- Thurstone, L.L. Multiple factor analysis, Chicago : University of Chicago Press, 1947.
- Thurstone, L.L. and Chave, E.J. The measurement of attitude, Chicago : University of Chicago Press, 1929.
- Tittle, C.R. and Hill, R.J. Attitude measurement and prediction of behaviour : an evaluation of conditions and measurement techniques. Sociometry, 1967, 30, 199-213.
- Torgerson, W.S. Theory and methods of scaling, New York : John Wiley and Sons, 1958.
- Trenaman, J. Guttman's scalogram analysis of attitudes. In Attitude Scaling, London : The Market Research Society, 1960.
- Tucker, W.T. Experiments in aesthetic communications. Unpublished doctoral dissertation. University of Illinois, 1955. Reported in Osgood, C.E., Suci, G.J. and Tannenbaum, R.H. The measurement of meaning, Urbana, Illinois : University of Illinois Press, 1957.

- Walker, D. Answer pattern and score scatter in tests and examinations. British Journal of Psychology, 1931, 22, 71-86.
- Wallin, P. A Guttman scale for measuring women's neighbourliness. American Journal of Sociology, 1953, 59, 243-246.
- Wohlwill, J.F. A study of the development of number concept by scalogram analysis. In Sigel, I. and Hooper, F. (Eds.). Logical thinking in children, New York : Holt, Rinehart and Winston, 1968.
- Wrigley, C. and McQuitty, L.L. The square root method of factor analysis. Department of Psychology, University of Illinois (Mimeographed). Reported in Osgood, C.E., Suci, G.J. and Tannenbaum, R.H. The measurement of meaning, Urbana, Illinois : University of Illinois Press, 1957.
- Zunich, M. A study of the relationship between child rearing attitudes and maternal behaviour. Journal of Experimental Education, 1961, 30, 231-241.

# APPENDIX I

## Least Squares Method of Scale Analysis

The least squares method of scale analysis was developed by Guttman (1941) to define a "best" answer, which other methods of scale analysis approximate visually.

### 1. Solution employing Correlation Ratio

Suppose P subjects respond to a set of m items which have common content. These responses can be set out in a table in the following manner:

Question	Response	Subjects					
		1	2	3	4	...	P
A	A <sub>1</sub>	✓		✓		...	
	A <sub>2</sub>		✓			...	✓
	A <sub>3</sub>				✓	...	
B	B <sub>1</sub>			✓		...	✓
	B <sub>2</sub>	✓	✓		✓		
.	.	.	.	.	.	...	.
Z	Z <sub>1</sub>		✓			...	✓
	Z <sub>2</sub>	✓		✓		...	
	Z <sub>3</sub>				✓	...	
	Z <sub>4</sub>					...	

This table may be conceived of as being indefinitely large, but in practice only a few responses can be studied. If the question checking behaviour is sufficiently consistent inferences can be made from the sample of behaviour about the totality of behaviour. Scale analysis tests the hypothesis that the responses form a unitary class of behaviour. Numerical values are assigned to individuals, and to question subcategories. The former values are termed "scores", and the latter are termed "weights". The aim, then, is to derive a set of weights for the  $n$  subcategories that enables the maximum possible reconstruction of behaviour. This set of weights is a row vector,

$$X = (x_1 \ x_2 \ x_3 \ \dots \ x_n)$$

The behaviour in the table can then be represented by the matrix  $M_x$ .

$$M_x = \begin{vmatrix} x_1 & 0 & x_1 & 0 & \dots & 0 \\ 0 & x_2 & 0 & 0 & \dots & x_2 \\ 0 & 0 & 0 & x_3 & \dots & 0 \\ 0 & 0 & x_4 & 0 & \dots & x_4 \\ x_5 & x_5 & 0 & x_5 & \dots & 0 \\ . & . & . & . & \dots & . \\ 0 & x_{n-3} & 0 & 0 & \dots & x_{n-3} \\ x_{n-2} & 0 & x_{n-2} & 0 & \dots & 0 \\ 0 & 0 & 0 & x_{n-1} & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{vmatrix}$$

$M_x$  can be regarded as the product of a diagonal matrix of weights and the matrix  $M$ .

$$M = \begin{vmatrix} 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 1 \\ 1 & 1 & 0 & 1 & \dots & 0 \\ . & . & . & . & \dots & . \\ 0 & 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{vmatrix}$$

The matrix  $M_x$  can be thought of as representing what subjects do: what subjects do not do can be thought of in terms of a similar matrix in which the zeros are replaced by weights, and the weights are replaced by zeros. To differentiate between what a subject does and does not do, the difference between these two distributions is maximized. This difference will tend to be maximized for all subjects if the relative variability in the columns of  $M_x$  is minimized. This is done by maximizing the correlation ratio defined by

$$\eta_x^2 = \frac{m \sum_{i=1}^P a_i^2 - mP\alpha_x^2}{\sum_{j=1}^n N_j x_j^2 - mP\alpha_x^2}$$

where  $a_i$  = mean of  $m$  weights in column  $i$  of  $M_x$

$N_j$  = number of individuals who check subcategory  $j$

$\alpha_x$  = mean of all non zero entries in  $M_x$ .

The origin is arbitrary and so may be chosen such that  $\alpha_x = 0$ . Then  $m\alpha_x^2 = 0$ , and the correlation ratio can be written

$$\eta_x^2 = \frac{m \sum_{i=1}^P a_i^2}{\sum_{j=1}^n N_j x_j^2}$$

Now, since

$$\frac{\sum_{i=1}^P x_i}{m} = a_i$$

it is known that

$$\frac{1}{m} \times M = (a_1 \ a_2 \ \dots \ a_p)$$

Therefore,

$$m \sum_{i=1}^P a_i^2 = \frac{1}{m} \times M M' X'$$

If a diagonal matrix is formed

$$D = \begin{vmatrix} N_1 & & & \\ & N_2 & & \\ & & \ddots & \\ & & & N_n \end{vmatrix}$$

then  $\sum_{j=1}^n N_j x_j^2$  can be written  $x D x'$ , and  $\eta_x^2$  can be written

$$\eta_x^2 = \frac{x M M' x'}{m x D x'}$$

Maximizing  $\eta_x^2$  is equivalent to maximizing  $x M M' x$  under the restriction that  $m x D x'$  be some finite constant. Hence the expression  $x M M' x - m \phi x D x'$  is maximized where  $\phi$  is a Lagrange multiplier. That is, the condition is imposed that

$$x(M M' - m \phi D) = 0$$

Suppose this condition is satisfied for  $x_0$  and  $\phi_0$ , then by postmultiplying both members by  $x_0'$  and solving for  $\phi_0$

$$\phi_0 = \frac{x_0 M M' x_0'}{m x_0 D x_0'} = \eta_{x_0}^2$$

The number of possible solutions equals the rank of  $M$ , that is  $n - m + 1$ . One solution is extraneous and gives

$$\phi_0 = \eta_{x_0}^2 = 1$$

This extraneous solution, which gives a row vector of  $n$  units, each of which is unity, is subtracted out. Therefore  $n - m$  solutions are possible.

## 2. Solution employing correlation coefficient

If the behaviour of the P individuals is conceived as a distribution of weights, and subcategories are thought of as distributions of scores, the entire configuration will be more consistent if people with similar scores tend to check subcategories with similar weights.

The aim then is to determine a set of weights

$$w = (w_1 \ w_2 \ . \ . \ . \ w_n)$$

and a set of scores

$$z = (z_1 \ z_2 \ . \ . \ . \ z_p)$$

which will maximize the correlation between the mP pairs of values in  $M_{(w,z)}$ .

$$M_{(w,z)} = \begin{vmatrix} w_1 z_1 & 0 & w_1 z_3 & 0 & \dots & 0 \\ 0 & w_2 z_2 & 0 & 0 & \dots & w_2 z_p \\ 0 & 0 & 0 & w_3 z_4 & \dots & 0 \\ 0 & 0 & w_4 z_3 & 0 & \dots & w_4 z_p \\ w_5 z_1 & w_5 z_2 & 0 & w_5 z_4 & \dots & 0 \\ . & . & . & . & \dots & . \\ 0 & w_{n-3} z_2 & 0 & 0 & \dots & w_{n-3} z_p \\ w_{n-2} z_1 & 0 & w_{n-2} z_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & w_{n-1} z_4 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{vmatrix}$$

Since the correlation coefficient is invariant with respect to a change of origin for  $w$  and  $z$ , the origins may be chosen such that the sum of squares is zero. That is, so that

$$\sum_{i=1}^P z_i = 0$$

and

$$\sum_{j=1}^n N_j w_j = 0$$

The correlation coefficient can then be written

$$\rho = \frac{\sum wz}{\sqrt{\sum w^2 \sum z^2}}$$

This can be rewritten in matrix notation

$$\sum wz = wMz'$$

$$\sum w^2 = wDw'$$

$$\sum z^2 = mzz'$$

So

$$\rho = \frac{wMz'}{\sqrt{wDw'zz'}}$$

To maximize this correlation coefficient the bilinear form  $wMw'$  is maximized under the restriction that  $wDw'$  and  $mzz'$  be finite constants. The solution can be obtained by maximizing the expression

$$wMz' - 1/2\lambda wDw' - 1/2\mu mzz'$$

where  $\frac{1}{2}\lambda$  and  $\frac{1}{2}\mu$  are Lagrange multipliers. Differentiating this with respect to  $w$  and  $z'$  yields the conditions

$$Mz' - \lambda Dw' = 0$$

and

$$wM - \mu m z = 0$$

If  $\lambda \neq 0$  and  $\mu \neq 0$  then

$$w = \frac{1}{\lambda} zM'D^{-1}$$

and

$$z = \frac{1}{\mu m} wM$$

With the exception of the extraneous solution, for which  $\rho_0 = 1$ , the condition

$$\sum_{j=1}^n N_j w_j = \sum_{i=1}^p z_i = 0$$

will be satisfied for  $n - m$  solutions.

### 3. The identity of results

Since the correlation ratios in both directions equal the correlation coefficient for any particular solution, the regressions of  $w$  on  $z$  and of  $z$  on  $w$  must both be linear for the bivariate distribution  $M(w, z)$ . The subcategory weights are proportional to the mean of scores of individuals who checked them, and the scores of individuals are proportional to the mean of the weights of the subcategories they

checked. The constants  $\lambda_0$  and  $\mu_0$  can be seen to be regression coefficients and the linearity is directly expressed.

There are, then,  $m - n$  bivariate distributions for  $M_{(w,z)}$  and the one chosen is the one with maximum internal consistency. A bivariate frequency function fitted to this distribution,  $f_{(w,z)}$ , where both regressions are linear can be assumed.

Suppose it is known that an individual has a score  $z_i$ , and it is wished to reconstruct his behaviour with respect to item A, where the subcategories have weights  $w_1$ ,  $w_2$  and  $w_3$ . Comparing the three probability densities

$$f_{(w_1 z_i)} \quad f_{(w_2 z_i)} \quad f_{(w_3 z_i)}$$

it could be estimated that the individual checked the subcategory which corresponded to the highest of the three densities. If this procedure were applied to all individuals on all items the proportion of erroneous estimates would tend to decrease as the internal consistency, measured by the correlation coefficient, increased.

In practice  $f_{(w,z)}$  is usually estimated from a sample of individuals and a sample of items. Then, if an individual is outside the sample and does not have on record his behaviour to, say, item A, his behaviour can be predicted by the process described. The weights  $w_1$ ,  $w_2$  and  $w_3$  are known from the sample. If an individual's score  $z_i$  can be estimated this can be substituted into  $f_{(w,z)}$  with the known weights to reproduce his behaviour.

## APPENDIX II

### The D-method of Factoring

The D-method of factoring is described by Osgood et al. (1957) for use with semantic differential data on page 332.

This technique is essentially equivalent to Thurstone's diagonal method (1947). Although the diagonal begins with correlation coefficients, and this method begins with raw scores, the two techniques are equivalent under certain conditions.

		concepts										
		1	...	f	...	g	...	h	...	i	...	m
	1	$X_{11}$	...	$X_{1f}$	...	$X_{1g}$	...	$X_{1h}$	...	$X_{1i}$	...	$X_{1m}$
	.	.		.		.		.		.		.
	.	.		.		.		.		.		.
	.	.		.		.		.		.		.
scales	j	$X_{j1}$	...	$X_{jf}$	...	$X_{jg}$	...	$X_{jh}$	...	$X_{ji}$	...	$X_{jm}$
	.	.		.		.		.		.		.
	.	.		.		.		.		.		.
	.	.		.		.		.		.		.
	k	$X_{k1}$	...	$X_{kf}$	...	$X_{kg}$	...	$X_{kh}$	...	$X_{ki}$	...	$X_{km}$

The assumption is made that the matrix defines a space of  $k$  dimensions, such that each concept  $i$  has coordinates  $(X_{1i} \dots X_{ji} \dots X_{ki})$  on the  $k$  dimensions. The aim is to find the coordinates of a new set of  $k'$  dimensions where  $k' < k$ .

For k dimensional space the following definitions are made

$$D_{oi}^2 = \sum_{j=1}^k x_{ji}^2; \text{ the squared distance between}$$

concept i and the origin D.

$$D_{hi}^2 = \sum_{j=1}^k (x_{jh} - x_{ji})^2; \text{ the squared distance}$$

between any two concepts h and i.

$\theta_{hi}$  ; the angle between two vectors where one vector extends from 0 to h and another from 0 to i.

In the k-dimensional space:

$$D_{hi}^2 = D_{oh}^2 + D_{oi}^2 - 2D_{oh} D_{oi} \cos \theta_{hi}$$

therefore

$$D_{oi} \cos \theta_{hi} = \frac{D_{hi}^2 - D_{oh}^2 - D_{oi}^2}{-2D_{oh}} = C_{Ii} \quad (1)$$

This is the coordinate of the concept i on a dimension passing through h.

The dimensions are symbolized I, II, III ..., and the coordinate of i on the first dimension is symbolized as  $C_{Ii}$ .

To find the coordinates on a second dimension, II, orthogonal to I, the distances in k space must be reduced to their k - 1 values by subtracting from the  $D^2$  values, their squared components on I.

The reduced distances ( $D'$ ) may be substituted in equation (1).

$$C_{III} = \frac{(D_{gi}')^2 - (D_{og}')^2 - (D_{oi}')^2}{-2D_{og}'} \quad (2)$$

where

$$(D_{gi}')^2 = D_{gi}^2 - (C_{Ig} - C_{Ii})^2$$

$$(D_{oi}')^2 = D_{oi}^2 - C_{Ii}^2$$

$$(D_{og}')^2 = D_{og}^2 - C_{Ig}^2$$

To find a third dimension orthogonal to I and II, a concept  $f$  through which III will pass in  $k - 2$  space is selected. The distances  $D''$  in the  $k - 2$  space are found by subtracting their components on I and II.

$$C_{III} = \frac{(D_{fi}'')^2 - (D_{of}'')^2 - (D_{\alpha}'')^2}{-2D_{of}''} \quad (3)$$

where

$$(D_{fi}'')^2 = D_{fi}^2 - (C_{If} - C_{Ii})^2 - (C_{IIIf} - C_{III})^2$$

$$(D_{oi}'')^2 = D_{oi}^2 - C_{Ii}^2 - C_{III}^2$$

$$(D_{of}'')^2 = D_{of}^2 - C_{If}^2 - C_{III}^2$$

This process is continued until the coordinates are reduced to zero, or a negligible amount.

Now as

$$D_{oi}^2 = \sum_{j=1}^k x_{ji}^2$$

and

$$D_{hi}^2 = \sum_{j=1}^k (x_{jh} - x_{ji})^2$$

it is possible to work with sums of cross products and squares rather than distances. By substituting the equivalences into equation (1) it is found that:

$$c_{Ii} = \frac{\sum_{j=1}^k x_{jh} x_{ji}}{\sqrt{\sum_{j=1}^k x_{jh}^2}}$$

Similarly

$$c_{IIIi} = \frac{\sum_{j=1}^k x_{jg} x_{ji} - c_{Ig} c_{Ii}}{\sqrt{\sum_{j=1}^k x_{jg}^2 - c_{Ig}^2}}$$

and

$$c_{IIIi} = \frac{\sum_{j=1}^k x_{jf} x_{ji} - c_{If} c_{Ii} - c_{IIIf} c_{IIIi}}{\sqrt{\sum_{j=1}^k x_{jf}^2 - c_{If}^2 - c_{IIIf}^2}}$$

This method has been shown to yield results corresponding closely with those obtained with the centroid method when both are applied to correlation matrices. When the technique is applied to raw score matrices the distances (D) between variables can be reproduced and hence the name D-method of factoring.

APPENDIX III

(i) Questionnaire A

Background Information

Name

Address

Place of Birth

Age

Sex

Marital Status

Religion

Education

Occupation

Length of time lived in the Potteries

Questionnaire

Here are a number of statements of opinion about how much obedience children owe to their parents. Each question or statement is followed by a list of items; you are asked to tick one of these to show how far you agree or disagree with the statement.

1. Obedience and respect for authority are the most important virtues children should learn.

1 - strong agreement

2 - moderate agreement

3 - slight agreement

4 - no opinion or unsure

- 5 - slight disagreement
- 6 - moderate disagreement
- 7 - strong disagreement

How strongly do you feel about this?

- 1 - very strongly indeed
- 2 - quite strongly
- 3 - not at all strongly

2. Parents should always be obeyed.

- 1 - agree
- 2 - disagree
- 3 - undecided

How strongly do you feel about this?

- 1 - very strongly indeed
- 2 - quite strongly
- 3 - not at all strongly

3. Even if young people know their parents are in the wrong they should do as they are told.

- 1 - agree
- 2 - disagree
- 3 - undecided

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

4. Do you think young people should always do what their parents tell them to?

- 1 - yes, parents always know what is best
- 2 - yes, even if parents make an occasional mistake they should be obeyed
- 3 - no, sometimes young people know better than their parents
- 4 - undecided

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

5. Young people sometimes get rebellious ideas but as they grow up they ought to get over them and settle down.

- 1 - strong agreement
- 2 - moderate agreement
- 3 - slight agreement
- 4 - no opinion or unsure
- 5 - slight disagreement
- 6 - moderate disagreement
- 7 - strong disagreement

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

6. There are no circumstances under which a parent should be disobeyed.

- 1 - agree
- 2 - disagree
- 3 - undecided

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

7. No-one is better qualified to say what is right for a young person than his parents.

- 1 - agree
- 2 - disagree
- 3 - undecided

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

8. What children need most from their parents is strict discipline.

- 1 - agree
- 2 - disagree
- 3 - undecided

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

9. A good parent is one who teaches his child obedience and respect for authority.

- 1 - agree completely
- 2 - agree moderately
- 3 - undecided
- 4 - disagree moderately
- 5 - disagree completely

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

10. The quality which should be most admired in a child is obedience.

- 1 - strong agreement
- 2 - moderate agreement
- 3 - slight agreement
- 4 - no opinion or unsure
- 5 - slight disagreement
- 6 - moderate disagreement
- 7 - strong disagreement

How strongly do you feel about this?

- 1 - very strongly
- 2 - quite strongly
- 3 - not at all strongly

11. It is wrong for children to question the authority of their parents.

1 - agree

2 - disagree

3 - undecided

How strongly do you feel about this?

1 - very strongly

2 - quite strongly

3 - not at all strongly

(ii) Questionnaire B

"Here are a number of statements of opinion about how much obedience children owe to their parents. Each question or statement is followed by a list of items; please circle the number next to the one which comes closest to your opinion. For instance, if you agree with the first statement you will circle number 3 next to "agree". If you disagree you will circle number 1 next to "disagree".

1. Parents should always be obeyed.

3. agree

2. undecided

1. disagree

2. Even if young people know their parents are in the wrong they should do as they are told.

3. agree

2. undecided

1. disagree

3. Do you think young people should always do what their parents tell them to?

4. yes, parents always know what is best

3. yes, even if parents make an occasional mistake they should be obeyed

2. undecided

1. no, sometimes young people know better than their parents

4. There are no circumstances under which a parent should be disobeyed.
  3. agree
  2. undecided
  1. disagree
5. A good parent is one who teaches his child obedience and respect for authority.
  5. strongly agree
  4. agree
  3. undecided
  2. disagree
  1. strongly disagree
6. Do you think that children under 5 years of age should always be expected to be obedient?
  4. yes, at this age children need discipline
  3. undecided
  2. no, one should not expect a small child to always be obedient
  1. no, obedience is not very important at this age
7. Do you think that children under 12 years of age should always obey their parents?
  3. yes
  2. undecided
  1. no

8. Do you think that children under 12 years of age should always obey their parents?
3. yes, always
  2. undecided
  1. no, it depends on the circumstances
9. People tend to place too much emphasis on obedience and discipline today.
3. agree
  2. undecided
  1. disagree
10. At what age do you think a boy should begin to make up his own mind about things, even if this means he is disobedient and goes against the wishes of his parents?
4. by the time he is 5 years of age
  3. by the time he is 12 years of age
  2. by the time he is 21 years of age, married or earning his own living
  1. never, he should not go against the wishes of his parents at any age.
11. Parents should expect unquestioning obedience from their children at all times.
3. agree
  2. undecided
  1. disagree

12. At what age do you think a girl should make her own decisions even if this means disobeying her parents?

5. by the time she is 5 years of age
4. by the time she is 12 years of age
3. by the time she is 21 years of age, or earning her own living
2. by the time she is married
1. never, she should not disobey her parents at any age.

The following form asks for your feelings about obedience in another way. Underneath the word 'obedience' there are rows with five spaces. At the end of the rows are adjectives that form opposites like "good" and "bad". You are asked to mark one of these spaces to show your feelings about obedience. A mark in the space next to "good" will mean you think obedience is "extremely good". A mark in the next space will mean you think it is "quite good", while the middle space stands for "neither good nor bad".

The fourth space will indicate "quite bad" and the fifth space "extremely bad". Please place one mark in each row to show your feelings in this way.

OBEDIENCE

good	-	-	-	-	-	bad
gentle	-	-	-	-	-	violent
fast	-	-	-	-	-	slow
active	-	-	-	-	-	passive
sour	-	-	-	-	-	sweet
clean	-	-	-	-	-	dirty
weak	-	-	-	-	-	strong
large	-	-	-	-	-	small
masculine	-	-	-	-	-	feminine
useful	-	-	-	-	-	useless
unnecessary	-	-	-	-	-	necessary
valuable	-	-	-	-	-	worthless
unimportant	-	-	-	-	-	important
desirable	-	-	-	-	-	undesirable
wise	-	-	-	-	-	foolish

Thank you very much for telling us about your opinions.

We are most interested in the differences in people's opinions.

For instance, in how the ideas of young and old people differ, and

in whether people of different religions have different ideas about

discipline. Perhaps we will find that people of different occupations,

or people from different areas have different ideas. For this reason we would like you to fill in the following items of information.

Of course this paper is confidential and there is no need to disclose your name.

Address:

Place of Birth:

Age:

Sex:

Marital Status:

Religion:

Education:

Occupation:

Occupation of husband or wife:

Length of time lived in the Potteries:

(iii) Questionnaire C (Items included on general questionnaire)

"Here are a number of statements of opinion about how much obedience children owe to their parents. Each question or statement is followed by a list of items; please circle the number next to the one which comes closest to your opinion. For instance, if you agree with the first statement you will circle number 3 next to "agree". If you disagree you will circle number 1 next to "disagree".

1. Parents should always be obeyed.

3. agree

2. undecided

1. disagree

2. Even if young people know their parents are in the wrong they should do as they are told.

3. agree

2. undecided

1. disagree

3. Do you think that young people should always do what their parents tell them to?

4. yes, parents always know what is best

3. yes, even if parents make an occasional mistake they should be obeyed

2. undecided

1. no, sometimes young people know better than their parents

4. There are no circumstances under which a parent should be disobeyed.
3. agree
  2. undecided
  1. disagree
5. Do you think children under 12 years of age should always obey their parents?
3. yes
  2. undecided
  1. no
6. Do you think that teenage children should always obey their parents?
3. yes, always
  2. undecided
  1. no, it depends on the circumstances

Now we wish to ask your opinions in another way. What do you think about "obedience"? Do you think it is good, or bad; evil, or virtuous; foolish, or wise; valuable, or worthless; important, or unimportant; interesting, or uninteresting? If you think it is extremely good you can mark the space like this

good      X      -    -    -    -    bad

If you think that obedience is good, but only quite good you can mark the next space

good    -      X      -    -    -    bad

A mark in the middle space will mean neither good nor bad, while this

good    -    -    -      X      -    bad

means slightly bad and this

good    -    -    -    -      X      bad

means extremely bad.

#### OBEDIENCE

good    -    -    -    -    -    bad

foolish    -    -    -    -    -    wise

evil    -    -    -    -    -    virtuous

valuable    -    -    -    -    -    worthless

important    -    -    -    -    -    unimportant

interesting    -    -    -    -    -    uninteresting

..."Thank you very much for telling us about your opinions.

We are most interested in differences in people's opinions. For instance, in how the ideas of young and old people differ, and in whether people of different religions have different ideas about discipline.

For this reason we would be most grateful if you would fill in the following items of information. Of course, this paper is confidential and there is no need to disclose your name.

Address:

Place of birth:

Age:

Sex:

Marital status:

Religion:

Education:

Occupation of head of family: