PART II

Critique of ESRI Paper No. 97

T. J. Baker, D. F. Hannan, D. B. Rottman and B. M. Walsh

CONTENTS

TABLES

1 INTRODUCTION

In Part I of this publication, Davis and Sinnott answer criticisms of the methodology and results of their Paper, *Attitudes in the Republic of Ireland Relevant to the Northern Ireland Problem* (ESRI Paper No. 97). Our principal aim in Part II is to offer a reasoned critique of Paper No. 97 as it was published, although we also take account of the additional information and explanation provided by the authors in Part I. We thus hope that our critique provides in itself an independent and balanced assessment of Paper No. 97.

Our appraisal runs parallel for some way with the authors' exposition in Part I. To a large extent we share a common view of the aims of Paper 97 and of the appropriate methodology for realising these aims. Like the authors, we believe that a careful examination of this methodology and the manner in which it was applied can go a long way towards answering the key question of whether there is sufficient evidence to justify the more contentious findings of Paper No. 97. It is in our assessment of the extent to which they correctly followed the prescribed rules of attitudinal research that we part company with the authors, and are led to dispute the findings of Section IV of Paper No. 97.

2 GENERAL OVERVIEW OF PAPER NO. 97

The foundation of any social research which seeks to establish the opinions or attitudes of the population as a whole is the sample which is chosen to represent that population. If the sample is inadequate or biased then the results obtained cannot be accurately grossed up to obtain a valid indication of the views of the population. Suggestions have been made that the sample used in Paper No. 97 was deficient. As mentioned in the Preface, the issues concerning sampling are dealt with in Appendix I by the Head of the ESRI Survey Unit. We are confident that the sample was as good as could have been obtained, and that information derived from it can be taken, within the normal statistical confidence limits, as reflecting national responses.

Similarly, we have no criticisms to make of the actual administration of the survey questionnaire which was carried out by a professional and experienced field force. There is no reason to suppose that there is any significant bias in the responses due to interviewer effects.

Given that the sample and interviewing techniques were satisfactory, it follows that the data collected in the survey reflect with reasonable accuracy the answers of the national population in the summer of 1978. How valuable these data are therefore depends solely on the nature of the questions asked. The questionnaire was long, with over 200 questions, although the answers to many of these are due to be analysed in a subsequent study. The questions relevant to Paper No. 97 fall into three distinct groups.

The first group of questions need not detain us. Its purpose was simply to establish the basic demographic characteristics of each respondent, such as sex, age group, marital status and educational level. Such information is obviously necessary for

analytical purposes as well as to monitor the validity of the sample as a whole.

The second group of questions concerned opinions on alternative possible solutions to the Northern Ireland problem and preferences in relation to policy issues facing the Irish or British governments. The answers to these questions form the basis to Section III of Paper No. 97, in which the authors analyse the responses in considerable detail. Most of the questions concerning solutions were of a "forced choice" variety, in which the respondents had to make first, second and last choices between clearly defined alternatives presented to them. With regard to policy proposals most of the questions involved agreeing or disagreeing with well defined statements concerning possible policy initiatives.

On the whole, this group of questions appears to be well thought out and clearly presented. The analysis, in terms of relating answers to different questions to each other and to demographic factors, is comprehensive and illuminating. We may have drawn one or two slightly different inferences from the data, but this could be said of practically any research.

The third group of questions was designed to elicit underlying attitudes towards aspects of the Northern Ireland problem. These questions, and the analysis of the answers, occupy Section IV of Paper No. 97. Most of the criticism of Paper No. 97 has been focused on this Section, and particularly on its first eight pages (94-101). These pages form the core of the authors' attempt to measure attitudes concerning Northern Ireland, and we have grave misgivings about the validity of these measurements. The remainder of this critique is accordingly devoted to a detailed examination of the methodology applied to this short but vital portion of Paper No. 97.

3 THE AIMS AND PROCEDURES OF ATTITUDE RESEARCH

In Section IV the aim is to identify and analyse attitudes relevant to the Northern Ireland problem. Because some critics appear to have misunderstood what is meant by the word "attitude" in this context, it is necessary to be very clear as to the nature of attitude research before assessing how well the authors succeeded in their aims.

Attitude research is a well established branch of social psychology, sharing many techniques with the better-known field of personality measurement and possessing an extensive literature of its own. By no means all other social scientists are convinced of its practical utility in extending our understanding of society, but few would deny that it is a legitimate field of study.

In a broad psychological sense, attitudes are commonly described as containing three elements: the "cognitive", or rational awareness and understanding; the "affective", or emotional disposition or feeling; and the "behavioural", or manifest response in terms of observable actions. However the authorities in the field[1] are in general agreement that what attitude research attempts to measure are sets of basically emotional dispositions towards given target objects, issues, or groups of people. Many of the definitions quoted imply that attitude will influence behaviour, but none the less they make it clear that the term "attitude" refers to the underlying feeling towards the object and not to any pattern of behaviour in respect of it. Similarly, attitudes are clearly distinguished from beliefs, although again one can influence the other.[2] Finally attitudes are evaluative, in the sense that

1 See Davis and Sinnotts' discussion in Part I (pp. 3-5) and in particular their quotations from Allport (1935) Guilford (1954) Oppenheim (1966) Katz (1966), Kerlinger (1973) and Nunnally (1970).

2 Edwards (1957, pp. 10-12).

feelings towards the object can be positive or negative, but the evaluation is essentially on an emotional, rather than a rational, plane. In technical terms, attitude research looks for "affective" rather than "cognitive" responses.

This fundamental characteristic of attitude research is not set out clearly in the course of Paper No. 97. It is left to the reader to be sufficiently acquainted with the field to recognise that attitudes are related to feelings rather than to thought out positions. Nevertheless, Part I of this document shows that the authors do in fact accept the standard interpretation of the term. As they say, (p. 5), "In summary, when analysing attitudes in this sense we are seeking to identify psychological states, dispositions, evaluative orientations or feelings toward the object in question".

The main purpose of attitude research is to examine the association between particular attitudes and a number of other variables. These may be other attitudes, opinions on specific issues, or the socio-demographic characteristics of the respondent, such as age, sex, education, occupation and location. In order to make such comparisons, it is necessary to construct some sort of measure for the attitude concerned.

Because attitudes are on an emotional rather than a rational level, they cannot usually be ascertained through the posing of single, clear-cut questions. Rather they may be "tapped" by an array of questions, generally referred to as "attitudinal items". These may be presented in various ways, but one method frequently used, and that chosen for use in Section IV, is the "Likert" item. This consists of presenting a statement relevant to the attitude being studied, with which a respondent may express differing degrees of agreement or disagreement. Because the answers may be graded in intensity, the responses to an individual Likert item can be converted into a crude numerical scale. By adding a respondent's score on a number of these items together, a "summated scale" or index is obtained which, it is hoped, will represent a complex attitude more accurately than could the answer to any of the individual items on its own.

The success of any piece of attitude research depends almost entirely on the "validity" of the summated scales constructed. Validity simply means that the researcher has

indeed measured what he set out to measure. To be valid, a scale must be reliable, in the sense that it measures consistently, and it must also correspond closely to the concept that the researcher wishes to measure. By its nature, absolute validity can never be fully established, because confidence in the concepts used can never be complete, and perfect correspondence between an abstract concept and a concrete measure cannot be demonstrated. Nevertheless, the researcher can hope to produce scales possessing a high degree of validity. Procedures have been evolved, rules established and tests constructed which, if applied correctly, can greatly increase confidence that the scales are reasonably accurate measures of the attitude defined by the researcher.

Concepts can be developed and refined through a structured process of reading, pilot interviews, pretest and analysis before the main study is undertaken.[3] By this stage the major concepts should be clear and unambiguous. No formal rules or tests can be applied to the conceptual framework of a research paper, but commonsense and logic can usually detect whether it avoids confusion and known fallacy.

The first step towards producing scales is the selection of attitudinal items for inclusion in the survey questionnaire. The stages of pilot interviews and pretest are valuable in suggesting and checking possible sets of questions in relation to each expected attitude. Fairly clear-cut rules have been laid down concerning the characteristics which should be possessed by attitudinal items. While these rules should be followed, the ultimate test of how well the items were selected lies in how good the scales derived from them are.

Although items are generally selected in the hope that each one will help tap a particular attitude, it is necessary to confirm that they do group together as expected. This is usually done through the application, at both pretest stage and in the main study, of factor analysis. This is a statistical technique for examining the correlations among a group of variables, in order to abstract a number of clusters of those variables which are closely related, and which can therefore

3 We follow, here and throughout, the authors' terminology for these stages, although it is more usual to refer to the initial unstructured interviews as pretesting, and the application of a draft questionnaire to a small sample as a pilot survey.

be interpreted as possessing some common meaning. As with any statistical technique there are established rules governing the use of factor analysis, and there are commonly accepted criteria for determining the number of clusters to extract and which items to include in each cluster.

When the responses to individual items are added together to provide summated scales for each cluster as a whole, it is possible to test the reliability of these scales. While a scale of low reliability is of little value for any purpose, the degree of reliability required depends very largely on the use to be made of the scales. If an attempt is to be made, as in Paper No. 97, to present the scales not ordinally, but as absolute measures of the percentage of the population holding particular attitudes, then a very high level of reliability is needed, and other firm rules must also be obeyed.

Even if highly reliable scales have been constructed, the interpretation of the meaning of the scale is, in the last resort, a matter of personal judgement on the part of the researcher. His judgement, of course, is likely to be informed by his knowledge of previous research in the field, by his own experience in similar work and by his awareness of evidence from other, non-survey, sources. In certain cases it is possible to seek some confirmation that scales have been validly interpreted by obtaining logically concurrent results from other scales or series or by observing behaviour which has been accurately predicted by the scales. Nevertheless, full validity can never be assured, and the conclusions drawn from attitude research remain judgemental in the sense that statistical evidence can be cited as supporting, but not as demonstrating, the interpretations adopted by the writer.

It follows from this that results should be presented in a tone appropriate to the methodology. Where the results of reliable scales are being reported, the role of personal judgement in interpreting their meaning should be duly acknowledged and decisive statements implying full validity should be eschewed. Where the results of scales of doubtful reliability are concerned, the tone of reporting should be frankly speculative, especially if the scales relate to a new and unfamiliar field of study.

4 SECTION IV IN PRACTICE

Having briefly summarised the aims of the authors as we see them, and the basic elements of the methodology they have used, it is now possible to turn our attention to the actual content of Section IV of Paper No. 97. The principal task in this section will be to assess how far Section IV in practice conforms with the established guidelines of attitude research. This will be done under six headings: concepts, attitudinal items, application of factor analysis, construction and use of scales, interpretation and labelling, and presentation.

(a) Concepts

The authors went through the prescribed process of reading, pilot interviews pretest, analysis and main survey, with, presumably, intervals for thought between each. Despite this the concepts still appear to be confused and unclear. This lack of clarity concerning the concepts pervades the Section and probably underlies many of the other problems encountered.

The most serious example of lack of clarity, because it affects the identification of the stimulus presented to respondents as well as the interpretation placed on their replies concerns the use of the term "IRA". The term is not defined explicitly in Paper No. 97, but the tone of certain passages implies that the results are taken as applying to the Provisional IRA. This is clearest from the following (p. 100): "opposition to IRA activities is not overwhelming and certainly does not match the strong opposition so often articulated by public figures". Public figures in the Republic generally articulate opposition specifically to the Provisional IRA, and certainly not to the "old" IRA in its historical context.[4]

4. See below p. 65.

If it was intended to study attitudes to the Provisional IRA, then the terms "Provos", "Provisionals" or "Provisional IRA" would have been in keeping with Irish vernacular usage. Irish newspapers and RTE use these terms consistently in referring to the activities of the group concerned, and it is only in sections of the British popular press that the initials IRA are used without qualification to denote any and every form of militant Irish republicanism.

It should have become clear from the study phase preceding the first pilot interviews that the label "IRA" is inherently ambiguous because of the long and complicated history of Irish republicanism's militant wing, and that it is therefore unsuitable as a stimulus in attitudinal research in Ireland. The "IRA" could have been taken by many respondents as referring exclusively to the Provisionals, but it could have been taken by others as referring to the "old" IRA, which is now honoured at governmental level through the provision of state pensions and attendance at funerals and commemorative ceremonies. Some may even have associated the term with the Official IRA, which formed the illegal wing of the republican movement until the foundation of the Provisionals in 1969, and which remains in existence although militarily inactive.

The important point is that the emotional responses to these various potential interpretations are likely to be very different. The divergence in response to the "old" IRA and to the Provisionals can be seen in the speeches of almost any politican or other public figure. An extremely clear illustration of this divergence was provided recently by Dr Cathal Daly, Bishop of Ardagh and Clonmacnoise.[5]

> The term "Irish Republican Army", with the noble name and record which it earned 60 years ago, can still evoke powerful emotional responses. It cannot be too emphatically asserted that those who usurp the name now have no right or title, historical or moral, to use it. Their present methods, their aims, their ideology, place them in a totally different category ... The new IRA is a radically new phenomenon in Irish history – and it is a sinister one.

5 Quoted in *The Irish Times*, Jan. 2nd, 1980.

Failure to qualify the initials "IRA" in the questionnaire has led, in our view, to a situation where the emotional content of the stimulus presented to the respondents was not clear-cut. Consequently, the meaning of the responses is inherently ambiguous and clarity in interpretation is precluded.

A second major example of conceptual confusion concerns the attitude towards Northern Ireland Protestants. Here the target group is clearly enough defined, but the nature of the attitude is uncertain. The authors state that they are seeking a political attitude and not a measure of social or religious prejudice. However the meaning of the political attitude is not explained, and the extent to which it reflects a perception of difference rather than opposition cannot be ascertained.

Thirdly, the decision to drop the hypothesised attitude towards British involvement, and to regard this omission as a gain in understanding, indicates a confusion between the relative importance of logical models and mere statistical measurement. In any political analysis of the situation, Partition and the British presence must be two separate aspects of the problem, because the possibility exists of a British withdrawal resulting in a continuation of Partition between the Republic and an independent Northern Ireland. Simply to subsume British involvement under Partition, because of the statistical tests on the results obtained, involves ignoring the logical basis of the original hypothesis, and excessive pragmatism in rationalising results which do not support that hypothesis.

Finally, and pervasively, there is a lack of clarity over the basic concept of the research, the nature of attitudes as usually defined. This shows in the assertion that attitudes remain stable over time (p. 19). Given that attitudes are essentially affective, one would wish to see convincing evidence before accepting the assumption that they are unlikely to change significantly in response to events with a high emotional impact. The commonsense presumption would be that attitudes are reasonably stable in normal times, but that they are by no means immune from the shocks of current history. The implicit assumption that attitudes to Northern Ireland were impervious to the several dramatic events between the date of the survey and the date of going

to press seems incompatible with the concept of attitude adopted.

Similarly in their selection of attitudinal items and in their presentation of results the authors seem frequently to lose sight of the essentially affective or emotional character of an attitude and to veer uneasily towards ascertaining their respondents' thought-out positions or beliefs about issues. This will become clear from consideration of the items selected for the main survey questionnaire.

(b) Attitudinal Items

Drawing on the various authorities already cited, it is possible to summarise as follows the characteristics desirable in the Likert items selected to elicit each attitude.

1 They should be adequate in number to tap the expected attitude and to fulfil the remaining criteria.

2 They should be balanced between statements favourable and unfavourable to the subject.

3 They should incorporate strong, but not too extreme, expression of view, using vernacular language where possible.

4 They should be worded appropriately for the purpose of eliciting feelings. Factual statements capable of being interpreted factually, should be avoided.[6]

5 They should clearly identify the object concerning which attitudes are being tapped. Statements that may be interpreted in more than one way, or that are not relevant to the psychological object being measured, should be avoided.

In the light of these criteria, it is instructive to examine carefully the 17 attitudinal items on which the analysis was based, which for convenience are set out in Table 1.

6 As Edwards (1957 pp. 11, 12) says, "if a given statement is equally likely to be endorsed by those with favourable and unfavourable attitudes, then this statement will not be useful in differentiating between those with favourable and those with unfavourable attitudes. . .As a first step in developing an attitude scale, therefore, we eliminate from consideration all statements. . .that are factual or that might be interpreted as factual".

Table 1: *Attitudinal items employed in Section IV*

*Item No.**		
	I	*Items loading on to Partition Factor*
1		Reunification is essential for any solution to the problem in Northern Ireland.
4		This is an island and it cannot be permanently partitioned.
6		The presence of British Troops in Northern Ireland amount to foreign occupation of part of Ireland.
7		There will never be peace in Northern Ireland until partition is ended.
10		The sooner we get the idea that the North belongs to us out of our heads the better.
13		The major cause of the problem in Northern Ireland is British interference in Irish affairs.
	IIA	*Items loading on to IRA Activities Factor*
3		Were it not for the IRA, the Northern problem would be even further from a solution.
8		The methods of the IRA are totally unacceptable.
12		The IRA are basically a bunch of criminals and murderers.
	IIB	*Items loading on to IRA Motives Factor*
14		Leaving aside the question of the methods, I basically support the aims of the IRA.
16		The IRA are basically patriots and idealists.
	III	*Items loading on to Northern Ireland Protestant Factor*
2		The vast majority of Protestants in Northern Ireland are willing to reach an agreement acceptable to the Catholic community.
5		The basic problem in Northern Ireland is that Protestants are prepared to defend their privileges at all costs.
9		Since they are the majority, it is only right that Protestants should have the last say in how Northern Ireland is to be governed.
11		Northern Ireland Protestants have an outlook and an approach to life that is not Irish.
		Items omitted from the Scales
15		Were it not for the British, the situation in Northern Ireland would be worse than it is.
17		The Northern Ireland problem will not be solved by ending partition.

*The numbering of items is that used in Paper No. 97, and corresponds with the order of items in the correlation matrix in Appendix A3 to Part I.

(i) With regard purely to number of items, 17 could perhaps be regarded as an adequate total for representing the four attitudes originally postulated. The items were selected following a pretest, and were chosen so that at least four items should be included in the scale for each expected factor. In the event, however, the items did not group together as expected, and consequently some of the attitudes identified by the authors are represented by an inadequate number of items. The attitudes regarding Partition and Northern Ireland Protestants may be adequately served by six and four items respectively, especially as these are more or less in line with the groupings established by pretest. The attitude relating to the IRA would also have been tapped by sufficient items had it not been split into two separate aspects. As it is, these are based on two and three items respectively, which is too few for confidence, especially as the existence of these two aspects had not been established at the pre-test stage.

(ii) Inspection of Table 1 shows that the balance between favourable and unfavourable expression of items is far from adequate. Of the six "Partition" items, five are anti-partition and only one pro-partition. Both of the "IRA Motives" items can be regarded as favourable. The "Northern Ireland Protestant" items are evenly divided, while the "IRA Activities" items are divided as evenly as is possible with only three items.

It could be, as the authors argue in Part I of this document that "acceptance response set" is not a major issue in a survey of this nature, and that consequently any bias imparted by unbalanced items is likely to be relatively minor. All the same the lack of balance must reduce confidence in the scales constructed from these items, and the failure of the authors to allude to the problem in Paper No. 97 exacerbates this concern. The fact that the lack of balance resulted from *post hoc* groupings of items replacing the expected grouping does nothing to restore confidence.

(iii) How far the terminology employed is too extreme is largely a matter of opinion, as there is no definite rule as to what constitutes an extreme statement. However, most

people would probably perceive "The methods of the IRA are totally unacceptable" and "The IRA are basically a bunch of criminals and murderers" as extreme expressions, as also they would "The basic problem in Northern Ireland is that Protestants are prepared to defend their privileges at all costs". The effect of extreme statements on the pattern of responses is not really known, but it seems unlikely that they do not impart some distortion.

(iv) It is with regard to the suitability of the wording of the items that we most take issue with the authors' selection. Remembering that the purpose of attitudinal items is to elicit felt responses rather than rationalised answers or beliefs, then far too high a proportion of the questions invite a thought-out reply. Opinions may legitimately differ on how far individual questions are likely to have evoked cognitive responses. It does, however, seem self-evident that such statements as "were it not for the IRA, the Northern problem would be even further from a solution", and "the vast majority of Protestants in Northern Ireland are willing to reach an agreement acceptable to the Catholic community" will have been interpreted by a significant proportion of respondents as calling for a rational assessment of the facts, regardless of their individual feelings towards the IRA or Northern Ireland Protestants. Although the two questions just quoted are the clearest examples of inappropriately "factual" items, many of the others could also have been treated as factual by some respondents. Certainly it is quite feasible to be implacably opposed to both the aims and methods of the IRA, however defined, and yet to concede that its members are basically patriots and idealists. Throughout history many extremely unpleasant movements have been motivated by idealism and patriotism, and even those most hostile to such movements have been willing to acknowledge the idealism from which they sprung.

The criticism that some items are too factual cannot be adequately answered by pointing out that many respondents' perception of facts may be coloured by their attitudes. Provided that any significant proportion of the sample answered questions in a factual manner which cut across their feelings

towards the subject, the resulting scales cannot be interpreted as measuring attitude. If one takes Edwards (1957) dictum that attitude items should directly and unambiguously measure people's feelings, very few of the items fully meet this criterion. At least half of the items appear to tap beliefs about "what is" rather than reflect preferences or feelings. It requires an unacceptable degree of inference to interpret responses to these questions as expressing affect.

(v) Finally, the items relating to the IRA are deficient in respect of defining clearly the object to which the attitude refers. Not only does the IRA itself remain undefined, as discussed earlier, but the concepts of "activities" and "motives" are never made clear to the respondents. This may well be because these are merely labels, attributed *post hoc* by the authors, rather than concepts which the survey was designed to investigate. Whatever its cause, the effect of this lack of clarity is obvious: neither the authors nor the reader can interpret with any confidence just what it is that is being measured. This is particularly damaging in the case of "activities", where a mildly positive attitude might mean a willingness to condone some degree of violence or might alternatively mean that the respondent is discriminating between violence and such non-violent activities as H-Block protests, rent-strikes or the operation of "black-taxis". There is simply no way of knowing.

(c) Application of Factor Analysis

So far we have attempted to express ourselves in relatively straightforward language, and have kept our use of technical terms to a minimum. This and the succeeding section, however, are inescapably technical in nature, and, although we shall continue to strive for general readability, the more widespread use of technical language cannot be avoided.

This is an important section, because the use of factor analysis is central to the portion of Paper No. 97 which we are examining. The results of their factor analysis are cited by the authors (p. 23) as "further confirming and, if necessary modifying, our initial hypotheses and interpretations." Similarly the grouping of items on to the scales used to measure

the hypothesised attitudes is dependent on the factor analysis carried out. Thus, if the factor analysis itself cannot be shown to have been correctly carried out, then the authors' interpretations must be suspect, and their measures will lack meaning.

Factor analysis is a technique for identifying separate clusters within a group of variables. More specifically, it is a method for studying and grouping the correlations or covariances between the variables. Certain common tendencies underlie the pattern of correlations and account for a proportion of the common variance among the items. These tendencies are referred to as components or factors. There are as many components as there are items in the group, but most of them are very weak and account for only a small proportion of the variance. However, there are usually a few strong factors present which account for a high proportion of the total variance, and it is these strong factors which are sought in factor analysis.

Some of the variance in each individual item is accounted for by one or more of the underlying factors. The greater the proportion of an item's variance which can be attributed to a particular factor, the more strongly that item is said to "load" on to that factor. However, the number of useful factors and the way the items load on to them is dependent on the structure of the basic correlation matrix. As Nunnally puts it, "factor analysis is nothing more than a set of mathematical aids to the examination of patterns of correlations" (*op. cit.* p. 371). Basically, it seeks clusters of variables such that the average level of inter-correlation between items within the cluster is high, while the correlations between any items contained in the cluster and items outside it are relatively slight. Nunnally sums up, "a test should 'hang together' in the sense that the items should correlate with one another. Otherwise, it makes little sense to add scores over them and speak of the total scores as measuring any attribute" (*op. cit.* p. 215). In other words, for a factor to form a useful basis for measurement, the items within it should be reasonably homogeneous, and clearly distinct from items outside.

There are no absolute rules for establishing how many factors should be extracted from a group of items, or what

the constituents of each factor should be.[7] There are, however, basic guidelines in the use of the technique.

The most commonly used guideline is that originally suggested by Guttman and restated by Kaiser: all factors with eigenvalues greater than 1.0 are extracted.[8] When the correlation matrix is used as the basis for factor analysis, the eigenvalue of a factor may be interpreted as the proportion of the variance among all the items that can be attributed to that factor: If, for example, one has five items in a factor analysis and the first factor has an eigenvalue of 2.4, then we know that the first factor is responsible for 48 per cent of the variance in the five items. Thus, the eigenvalue rule is a practical and sensible guideline, for any factor with an eigenvalue lower than one is of less use than would be a single item from the set. The size of the eigenvalue also has a clear effect on the interpretability of a factor: "The higher this figure is, the more substantial can be the claim that the items with significant loadings have some property in common" (Child, *op. cit.*, page 42). Other tests such as Bartlett's[9] or Cattell's "scree-test",[10] have been used as guidelines in the determination of the number of factors to be extracted. While different guidelines may suggest the extraction of different numbers of factors, normal research procedure dictates that whichever guideline or test is selected it should be used consistently throughout the exercise. At the very least any change of criterion should be explained and justified by the researcher.

Having decided how many factors to extract, a solution for that number of factors is specified. This gives the loading of items on to each of the factors included. It should be noted that these loadings will vary if a solution for a different number of factors is undertaken.

7 The difficulty of applying classical statistical tests of significance to factor analysis imposes some limitation on the inferences which can properly be drawn from it. This provides a further argument in favour of adopting a cautious manner in presenting such inferences.

8 Kaiser (1960 pp. 141-151). The degree of acceptance of that criterion can be seen in its advocacy in, among other, the following basic texts: Child (1970 pp. 43-44), Rummel (1970 pp. 362-364), Van de Geer (1971 p. 147), Tatsuoka (1971 p. 147), and Taylor (1977 p. 116).

9 See Part I, p. 23.

10 See below p. 75.

Once the loadings on to the appropriate number of factors have been obtained, the next stage is to distribute the individual items between the factors. One method is to include each item in every factor, weighting the item's score by the item's loading on to each factor. In this way factor scores can be calculated which make use of all the information contained in the factor loadings. A further advantage is that this method recognises that the distribution of answers to an item may be influenced by two or more underlying attitudes. For these reasons this is now the generally preferred method of utilising factor analysis in the construction of scales.

An alternative, much cruder, method is adopted in Paper No. 97. This is to assign each item exclusively to the single factor on to which it loads most heavily, providing that this loading is above some arbitrary cut-off point and that the loadings on to other factors are below this point. The common choice for the cut-off point is 0.30, but a loading of 0.40, which is generally followed in Paper No. 97, has occasionally been used.

The major drawback of this method is that it wastes much of the information available. In the first place, items with split loadings of 0.40 or above on two factors, or just failing to reach 0.40 on any factor, have generally to be discarded, even though their relationship with each of the factors is highly significant. In the second place, the assignment to single factors involves treating any loading of 0.40 or higher as if it were a loading of 1.0, and any loading of 0.39 or below as if it were a loading of 0. Those latter include many statistically significant loadings, which are lost through the use of this method.[11]

Despite these drawbacks, the method has the benefit of simplicity and it remains quite widely used. Standard practice, of course, is for the selected cut-off point, whether it be 0.40,

11 Approximate formulae for calculating sampling errors of factor loadings were developed by Holzinger and Harman (1941) and are widely used today (e.g., Oster 1979). These suggest that 0.40 is considerably higher than the level required for a loading to attain significance with a sample size of 2,000. Although still arbitrary, a lower cut-off point would thus have been preferable, but would have intensified the problem of split loadings which was serious enough even when the 0.40 level was used. If factor scores are used, split loadings cease to be a problem.

0.30 or any other level, to be specified by the researcher and to be used consistently throughout the exercise.

The pattern of correlations between the 17 attitude items was not shown in Paper No. 97. The authors have, however, made it available as Appendix Table A3 to Part I of this document. It provides the starting point for the factor analysis.

Following the pretest, the hypothesis on which the authors based their selection of items was that four separate factors should be identifiable from this matrix of correlations. These four factors should represent attitudes towards Partition, the IRA, Northern Ireland Protestants, and British Involvement. As Table 2 shows, there are indeed four factors with eigenvalues greater than 1.0, although the first of these factors is clearly much stronger than the other three.[12]

Table 2: *Eigenvalues and their corresponding percentages of variance: 17 attitudinal items*

Component number	*Eigenvalue*	*Percentage of variance*	*Cumulative percentage of variance*
1	4.61	27.1	27.1
2	1.65	9.7	36.8
3	1.33	7.8	44.6
4	1.08	6.3	50.9
5	.94	5.5	56.4
6	.83	4.9	61.3
7	.78	4.6	65.9
8	.74	4.4	70.3
9	.68	4.0	74.3
10	.66	3.9	78.2
11	.61	3.6	81.8
12	.59	3.5	85.3
13	.54	3.2	88.5
14	.52	3.1	91.6
15	.50	2.9	94.5
16	.49	2.9	97.4
17	.44	2.6	100
	17	100	

12 Cattell's "scree test" gives somewhat indeterminate results. The slope resulting from plotting the proportion of variance explained by each factor declines to an approximately straight line after five factors, indicating that this is the solution to be sought. However, the test could just possibly be interpreted as allowing a 1, 3 or 4 factor solution instead. Bartlett's test indicates that five factors could be extracted.

Given this indication of the presence of four viable factors, and their hypothesis that there should be four specified factors present, the authors must have sought a four-factor solution in the next stage of their analysis. No results of this exercise are quoted in Paper No. 97, so we have ourselves undertaken it, using the correlation matrix and supporting information provided by the authors. The outcome is shown in Table 3. This indicates that at first sight the four-factor solution is moderately satisfactory, as would be expected from the eigenvalues shown in Table 2. All 17 items load on to factors at above the 0.40 cut-off point, although two of the items load at above this level on to two separate factors. "The vast majority of Protestants in Northern Ireland are willing to reach an agreement acceptable to the Catholic community" loads on to both Factor 3 and Factor 4, while "The sooner we get the idea that the North belongs to us out of our heads the better" loads on to both Factor 1 and Factor 4. Leaving these two items aside, there are clear-cut loadings of six items on to Factor 1, five items on to Factor 2, 2 items on to Factor 3, and two items on to Factor 4.

Table 3: *Factor analysis of 17 Likert items from Paper No. 97: Loadings on four factors (varimax rotation), principal factoring without iteration*

	Factors				
	1	*2*	*3*	*4*	*Communality**
1 Reunification is an essential condition for any solution of the problem in Northern Ireland	.74	.17	.14	.01	.59
2 The vast majority of Protestants in Northern Ireland are willing to reach an agreement acceptable to the Catholic community	.11	-.04	-.45	.55	.52
3 Were it not for the IRA, the Northern problem would be even further from a solution	.25	.66	.12	.08	.52
4 This is an island and it cannot be permanently partitioned	.75	.02	.09	-.07	.58
5 The basic problem in Northern Ireland is that Protestants are prepared to defend their privileges at all costs	.23	-.02	.69	-.04	.53

Table 3: (Continued)

	Factors				
	1	*2*	*3*	*4*	*Community**
6 The presence of British Troops in Northern Ireland amounts to foreign occupation of part of Ireland	.55	.18	.38	-.23	.53
7 There will never be peace in Northern Ireland until partition is ended	.74	.16	.20	.01	.61
8 The methods of the IRA are totally unacceptable	.07	-.73	.05	.10	.55
9 Since they are the majority, it is only right that the Protestants should have the last say in how Northern Ireland is to be governed	-.07	.03	-.07	.73	.54
10 The sooner we get the idea that the North belongs to us out of our heads the better	-.51	-.12	.01	.47	.50
11 Northern Ireland Protestants have an outlook and an approach to life that is not Irish	.08	.09	.71	-.06	.52
12 The IRA are basically a bunch of criminals and murderers	-.08	-.73	.05	.25	.61
13 The major cause of the problem in Northern Ireland is British interference in Irish affairs	.43	.27	.38	-.16	.43
14 Leaving aside the question of their methods, I basically support the aims of the IRA	.29	.57	.16	-.27	.50
15 Were it not for the British, the situation in Northern Ireland would be worse than it is	-.18	-.26	-.02	.51	.36
16 The IRA are basically patriots and idealists	.24	.57	.11	.07	.40
17 The Northern Ireland problem will not be solved by ending partition	-.47	-.23	.24	.24	.39

*Communality is a measure of the common variance among the items, representing the correlation between the item concerned and the factors extracted from the original matrix.

From the point of view of interpretation or measurement the four-factor solution is less helpful. Factor 1 is fairly satisfactory, including four expected Partition items, and two postulated British items. One expected Partition item is subject to a split loading although its heaviest loading is on this factor. Factor 2 is even clearer, with the five expected IRA items loading unambiguously on to it. Factors 3 and 4 are awkward. With only two items clearly on each, they are not suitable as measurement scales, while the content of the items makes interpretation very difficult. Indeed, one possible interpretation, especially if the split loaded Item 2 is attributed to Factor 4, is that Factor 3 groups together items antipathetic to Northern Ireland Protestants, and Factor 4 items sympathetic to them and their British connection, which would indicate that "response set" may have been a problem after all.

What is clear is that these factors do not represent the hypothesised attitudes towards Northern Ireland Protestants and the British presence respectively. The authors at this stage were obviously in some difficulty, with which we can fully sympathise. In spite of the proper sequence of pilot and pretest, the items selected had not clustered in the expected manner. This is by no means a rare occurrence in social scientific research, and the problems posed by such a situation have been faced by many research workers. The purest solution, which is in keeping with a strict hypothesis-testing approach to research, is simply to report the results and concede that parts of the hypothesis have failed to be borne out and must be abandoned at this stage. Attention can then be focused exclusively on the parts which have received confirmation, in this case Factors 1 and 2. Alternatively it is quite legitimate to argue that the results may indicate flaws in the original hypothesis and proceed, in an exploratory manner, to seek alternative hypotheses which would be compatible with the results.

What is not legitimate is to move on to an alternative three-factor solution, without reporting the four factor, and to announce to the reader that the factor analysis "produced not four but three factors" (p. 94). Nor is it acceptable to claim that this procedure, which involves abandoning the

hypothesis on which the items were selected, can "be seen as a substantive gain in terms both of our understanding of the attitudes in question and in terms of our confidence in the validity of our composite measures" (p. 94).

As we have seen, the eigenvalues provide no basis for adopting a three-factor solution. Even if a level of 1.0 is regarded as arbitrary, there are no obvious grounds for adopting, without explanation, a cut-off point somewhere between 1.08 and 1.33. The three-factor solution possesses no advantage in terms of the loading of the items. There are still two items, although a different two, with split loadings, and in this case they fail to reach 0.40 on any factor rather than reach it on two at once. Moreover, one of the items "lost" in the three-factor solution, "The Northern Ireland problem will not be solved by ending partition" is one of the most direct of the Partition items and one which loaded satisfactorily and unambiguously in the four-factor solution.

Nevertheless, the three-factor solution, like the four-factor, provides two clear factors, with six and five items respectively, and reasonably in line with the expected groupings of items. The third factor, however, is very weak, although it does contain the four expected Northern Ireland Protestant items. As Table 4 shows, the average inter-correlation among its items is only 0.19 and one correlation within it is as low as 0.09. This means that the two items concerned share less than one per cent of their variance. Moreover, the four items in the factor have correlations with several items from other factors which exceed those they have with one another. Thus, as a potential scale, the factor possesses neither homogeneity nor distinctness. In Nunnally's term, it does not "hang together". Any scale derived from it will lack clear meaning and as will be seen later, lack statistical reliability.

If Table 4 demonstrates that the third factor is too weak to use, it also shows that the first two factors do "hang together" reasonably well. In each case the pattern of correlations is relatively homogeneous and shows little overlap with items loading on to other factors. The authors, however, seek to establish that the IRA factor, which emerges quite clearly and unchanged from both the four and three-factor solutions, is in fact "two clearly different factors relating to

Table 4: *Correlation between items within factors*

3 factor solution, coefficients of correlation, signs omitted

I. Items loading on to partition factor

*Item**	1	4	6	7	10	13	*Number of correlations with items outside factor higher than lowest correlation within factor*
1	—	.47	.41	.55	.32	.33	0
4		—	.41	.49	.34	.31	0
6			—	.43	.34	.46	1
7				—	.32	.37	0
10					—	.28	1
13						—	2

Average coefficient among partition items = .389

II. Items loading on to IRA Factor+

*Item**		3	8	12	14	16	
3	IIA	—	.33	.37	.35	.29	0
8	(activities)		—	.42	.28	.20	1
12				—	.43	.34	0
14	IIB				—	.39	6
16	(motives)					—	4

Average coefficient among "activities" items (3, 8, 12) =.373
Average coefficient among "motives" items (14, 16) = .39
Average coefficient between "activities" and "motives" items =.315
Average coefficient among all IRA items (3, 8, 12, 14, 16) =.340

III. Items loading on to Northern Ireland protestant factor

*Item**	2	5	9	11	
2	—	.22	.24	.21	0
5		—	.09	.26	10
9			—	.12	9
11				—	7

Average coefficient among N.I. Protestant items = .190

*Items numbered as in Appendix Table A2 Part I.
+Correlations within each IRA sub-factor are indicated by the dotted lines.

the IRA" (p. 94). The reason given in Paper No. 97 for dividing the factor is that "the different levels of support which these items relating to the IRA showed in the nationwide sample suggested that, although these five items clustered together in a global factor analysis, the attitude towards the

IRA which they measured might itself be multi-dimensional" (p. 94). This is amplified (p. 98) as: "If our identification and interpretation of a two dimensional attitude to the IRA is valid, then the two dimensions should have contrasting distributions in the population. This is in fact the case".

This claim is based on a statistical fallacy. The marginal distributions of two variables are irrelevant to the relationship between them. Correlation, on which factor analysis is based, measures whether responses to items tend to move together, not whether they are of similar magnitude. If most individuals who score two on one item score five on the other, while those who score three on the first score six on the second and so on, then the two items will be highly correlated, although their means and marginal distributions are quite different. The claim is also extraordinary because it is quite simply not the case that there is any clear-cut discontinuity in the marginal distribution of replies to the five IRA items. Appendix Table A1 to Part I shows that while there is a considerable spread in the percentage responses to the five individual items, the distributions cannot be sensibly divided into two contrasting groups on this basis.

The method chosen to subdivide the IRA factor was to carry out a second stage of factor analysis, seeking a two-factor solution for the five IRA items in isolation. This is a most unusual procedure. One should not, of course, condemn an approach just because it is novel, for innovation is necessary if methodology is not to stagnate. However, it is incumbent upon authors in such circumstances to justify their techniques. This the authors of Paper No. 97 do not attempt, and indeed it would be difficult for them to do so in this case.

In the first place five items are too few for the proper identification of two factors, especially in the absence of a relevant pretest. Taylor, (1977), for example, recommends that "in practice the number of variables be at least five times the number of factors". While minor deviations from this criterion may be disregarded, halving the recommended guidelines is not reasonable. It ensures that, at best, the new factors will contain three and two items respectively, which in turn means that the resulting scales are likely to be unbalanced and unreliable.

In the second place, the procedure adopted violates the logic of the method. The initial factor analysis provided a factor that is orthogonal to the other factors; that is, a cluster of items that were answered with a high degree of consistency and with maximum independence from all other clusters. To apply a further stage of factor analysis implies that these five consistent items can then be further broken into two groups which are independent of each other. For the assumption of independence to hold, it must be plausible for each of the four combinations between the two factors to be held by some people. In this instance, independence presupposes that some may be favourable to both the activities and motives of the IRA, some unfavourable to both, some in favour of motives but unfavourable to activities and some favourable to activities but unfavourable to motives. It is reasonable to expect a considerable number of respondents to hold each of the first three of these combinations of attitudes, but it is exceedingly implausible that the fourth cell would be populated. Yet unless it can be postulated that some people do accept IRA activities while rejecting IRA motives, then independence between the "attitudes" cannot be hypothesised.

Such a hypothesis was advanced in Paper No. 97. In Part I of the present document, the break among the IRA items is justified using an oblique rotation, which allows the two factors to be correlated. The result is what Nunnally advises us to expect: the loadings on the oblique factors are clearer than those from orthogonal rotation and the results of the two rotations are rather similar (see Nunnally, *op. cit.*, pp. 325-327). This hardly overcomes the strength of the evidence that in fact only one factor is present for the IRA items, nor can it legitimise the novel "two stage" approach to factor analysis developed by Davis and Sinnott.

Leaving aside our objections to the very practice of performing a second-factor analysis on a five-item factor, it is clear that in this case the procedure does not result in the identification of two separate factors. The first factor emerges with an eigenvalue of 2.34, while the second factor has an eigenvalue of only 0.84. This is clear evidence that only one

viable factor is present in the five items. All five items load satisfactorily on to this single factor, as follows:

Item 3	0.683
Item 8	-0.642
Item 12	-0.761
Item 14	0.719
Item 16	0.634

There seems to be no basis here for distinguishing between the first three items and the remaining two.

In view of this evidence that only one factor exists, it is not surprising that in order to allocate their items between their two sub-factors the authors are obliged to abandon the general rule of thumb, which they have used in the first stage of their analysis, that items with a 0.40 loading or better are included in a factor. Item 12 ("criminals and murderers") loads 0.66 on to their first sub-factor and 0.41 on to their second. Rather than discard either the item or the approach, they arbitrarily, and without explanation in the text of Paper No. 97, raise the threshold for inclusion of an item from 0.40 to 0.60.

In any case, these factor loadings must be approached with considerable caution. There is some disagreement among the authorities concerning the correct treatment of the diagonal in a correlation matrix when factor analysis is to be applied. It is now widely accepted that communalities should be inserted on the diagonal (Tatsuoka *op. cit.* p. 145). Since these measure the common variance among the items, and exclude the variance which is unique to each item, it is argued that their use minimises error, both systematic and random, and permits greater confidence to be placed in the results of the factor analysis. On the other hand, Nunnally (*op. cit.* p. 355) prefers in general the use of unities, which represent each item's perfect correlation with itself.

However, on two points there is no disagreement. The researcher should make clear which method is being adopted. This is not done in Paper No. 97. Secondly, where very small numbers of items are involved, the use of unities artifically inflates the factor loadings (Nunnally *op. cit.* p. 369). Thus,

if an attempt is to be made to obtain two factors from a set of five items, then communalities, rather than the unities employed in Paper No. 97, should be used. The use of communalities in a two-factor solution for the five items concerned yields the following loadings.

Item	*Factor 1*	*Factor 2*
3 Were it not for the IRA, the Northern problem would be even further from a solution	.41	-.38
8 The methods of the IRA are totally unacceptable	-.18	.69
12 The IRA are basically a bunch of criminals and murderers	-.48	.48
14 Leaving aside the question of their methods, I basically support the aims of the IRA	.64	-.24
16 The IRA are basically patriots and idealists	.55	-.16

With two of the five items loading almost equally on to each factor it seems clear that, if communalities had been used, the two-factor solution could not have been adopted as the basis for sub-dividing the IRA items.

Against the manifold evidence that the IRA items form just one cohesive factor, the authors could argue that some statistical tests do indicate the presence of more than one factor, although as they rejected the findings of such tests when adopting a three-factor solution in their main analysis, they cannot legitimately rely on them in their secondary analysis. The clearest of such tests is probably that put forward by Bartlett.[13] This indicates that there should be three, not two, sub-factors among the five items.[14] However, Taylor (*op. cit.*) shows that such a solution is mathematically invalid when only five items are used. It might be noted that the danger of looking only at the factor loadings is well illustrated

13 Quoted in Kendall and Stuart, (1968) Volume 3, page 292.
14 This is also the result of the "scree test".

by the trivial three-factor solution, since it appears to yield results which are at least as interpretable as those of the two-factor solution adopted in Paper No. 97.

In assessing the use of factor analysis in Paper No. 97, it is as well to keep constantly in mind Nunnally's dictum: "One way to fool yourself with factor analysis is to ignore the correlations that are used to define a factor" (*op. cit.* p. 368). By paying too little attention to the underlying correlations, the authors would appear to have made the following major errors in their application of factor analysis.

1 To adopt a three-factor solution, where both the eigenvalues obtained and their own hypothesis suggest a four-factor solution should have been used.

2 To have accepted as meaningful a Northern Ireland Protestant factor lacking in homogeneity and which is so unstable that it disintegrates when a four-factor solution is applied.

3 To have forced through a sub-division of a clear IRA factor which the eigenvalues, the matrix of correlations and the single-factor loadings all show to be a single factor.

4 To have used methods which breach normal practice and infringe the rules of logic in effecting this sub-division.

Thus three of the four scales derived by the authors are based on misapplications of factor analysis. Even the fourth scale, that relating to attitudes to Partition, is imperfectly specified, in that the four-factor solution provides it with a slightly different content of items. If the four scales are to be justified, it must be on the grounds of evidence other than the factor analysis. No such external evidence is provided in Paper No. 97.

(d) Construction and Use of Scales

The principal purpose of the factor analysis is to group the 17 Likert items into a number of clusters. The responses to

each item within a cluster or factor are then combined together into a scale which is taken to measure the attitude hypothesised by the authors.

(i) Reliability

For such a summated scale to be taken seriously it must possess a reasonable level of reliability, meaning that the measurements obtained from the survey should be repeatable. Any scale will contain a certain amount of error, both random and systematic, but provided that the items in a scale are measuring a common phenomenon, and that the items are answered consistently, with errors from the various sources tending to cancel each other out, the scale will be sufficiently reliable to use in research.

The reliability of a scale can be tested, and Nunnally (*op. cit.* p. 194) recommends the use of the "coefficient alpha" test since " . . . in many ways this is the most meaningful measure of reliability". This test sets an upper limit on a scale's reliability, based on (a) the consistency among the items as expressed by their intercorrelations, and (b) the test length, or number of items in the scale. The alpha coefficient can range from zero to one, and is the estimated correlation between a scale as measured and the "true scores" that would have been obtained had it been possible to measure the attitude without error.

The alpha coefficients for the four scales used in Section IV are –

1	Attitude to Partition: Anti versus Pro	.793
2	IRA Activities: Support versus Opposition	.641
3	IRA Motives: Sympathy versus Rejection	.561
4	Northern Ireland Protestants: Anti versus Pro	.484

The level of reliability that is required in a scale is governed by the use to be made of the scale. Nunnally (p. 226) offers the following guideline:

> In the early stages of research on predictor tests of hypothesized measures of a construct, one saves times and energy by working with instruments that have only modest reliabilities, for which purpose reliabilities of .60 or .50 will suffice . . . in many applied settings a relia-

bility of .80 is not nearly high enough. In basic research, the concern is with size of correlations and with the differences in means for different experimental treatments, for which purpose a reliability of .80 for the different measures involved is adequate.

As the use made of the scales in Paper No. 97 could be defined as either applied or basic, it is clear that the reliability coefficients of three of the scales (those relating to IRA activities, IRA motives, and Northern Ireland Protestants) must be considered inadequate. In view of the shortcomings in the application of factor analysis, there can be no surprise that the scales do not achieve adequate reliability. However, because an attempt is made to use the scales to measure precisely the average level of support and opposition to particular objects, the absence of adequate reliability in the scales takes on an added importance.

(ii) Neutral Point

Even the individual Likert items are essentially ordinal measures, in that the participant is asked to agree or disagree with a statement more or less strongly. The placing of numbers on responses, such as a score of one for strong disagreement or six for moderate agreement with an item, although standard practice in attitudinal research, is an arbitrary procedure. The score on an individual item can in no way be compared with precise measures, such as those of length, weight or temperature. Rather it is a somewhat crude numerical representation of an order of ranking and equal intervals between the points of the score cannot, strictly, be assumed.

A problem in the application and scoring of individual Likert items is the treatment of the neutral point, where answers change from being in some degree favourable to some degree unfavourable. The purist approach is to have six potential scores, with no point intervening between slightly favourable and slightly unfavourable. This appears to have been the authors' intention, as the Likert items presented to respondents contained no box for a "don't know" score. Nevertheless, in spite of encouragement from interviewers to select a definite response, a small proportion of respondents insisted on returning a "don't know" answer. These were

accommodated by inserting a "don't know" category in the reported scores, and according this a score of four on a scale ranging from one to seven.

This practice allows a small area of indetermination in the score of each individual Likert item. It can be argued that it also leads to some distortion of the scale, as the pattern of answers received is probably somewhat different to that which would have been obtained had the respondents been offered a "don't know" category in the first place. A technical consequence of the constricted "don't know" area in the centre of the scale is that it tends to lead to a bi-modal distribution of answers, with a dip at the middle of the scale. This in turn implies that the items will not possess a normal distribution of responses, yet the assumption of normal distribution is a standard requirement for drawing inferences from correlation analysis, including factor analysis.

Although there is no discussion of the matter in the course of Paper No. 97, it could well be, as the authors argue in Part I of this document, that the issue of the neutral point on individual Likert items is not of great importance, and does not significantly affect the results obtained.

It is when we turn from individual Likert items to the summated scales that the difficulties caused by the treatment of the neutral point become severe. On an individual item, the neutral point between favourable and unfavourable responses possesses psychological meaning, even if it does pose problems in measurement. When individual items are added together no such psychological neutral point can be postulated. On, say, a four-item summated scale an average score of four can be obtained by dozens of different combinations of scores on the component items, ranging from two "strongly agrees" combining with two "strongly disagrees", through many mixtures of agreement and disagreement, to four genuine "don't knows". To interpret all these possible combinations of agreement and disagreement as neutrality towards the stimulus is not justifiable. Rather, the scores towards the centre of a summated scale can only be interpreted as indeterminate.

Secondly, even if the answers to questions are consistent in the sense that they correlate highly, it cannot be assumed

that all items tap the attitude in question with equal intensity. Consequently a score of, say, two on one item does not necessarily imply the same degree of favourability towards the attitude as the same score on a different item. It follows, crucially, that the changeover points from positive to negative responses to different specific items may bear differing relationships to the change from a favourable to an unfavourable position on the attitude as a whole.

Oppenheim (*op. cit.* pp. 140-141) sums up this argument as follows:

> With regard to the neutral point, we must agree that this is not necessarily the mid-point between the two extreme scale scores; moreover, scores in the middle region could be due to lukewarm response, lack of knowledge, or lack of attitude in the respondent (leading to many "uncertain" responses), or to the presence of both strongly positive and strongly negative responses, which would more or less balance each other.

Of course, scales containing items with a poor level of intercorrelation exhibit this problem in an acute form. They will suffer from a high proportion of inconsistent answers, while the relationship of each of a set of weakly linked items with the underlying attitude must be regarded as tenuous. Not only does the mathematical mid-point on such a scale lack any meaning, but the other points towards the centre of the scale are also likely to contain a large number of indeterminate sets of answers, so that they too cannot be interpreted with any degree of confidence. Only average scores, which are extremely high or low, approaching one or seven, can be interpreted without ambiguity, as such averages can only be obtained if the scores on all component items are similarly extreme. Obviously, unless a scale possesses adequate reliability it cannot sensibly be used as a measure at all.

Among scales which are reliable, the seriousness of the problem concerning the neutral point depends largely on the use to which the scale is to be put. Traditionally summated scales have been most widely used in personality research, where the nature of research requires ordinal measures. In attitude research, reliable summated scales are frequently used to compare the co-variance of two attitudes, or to

examine the relationship between certain socio-demographic variables and the strength with which an attitude is held. So long as applications of summated scales are limited to such ordinal uses the issue of the neutral point is not important, but in any attempt to interpret the scales as cardinal measures it becomes crucial.

A useful summary of the matter is given by Edwards (1957, p. 157)

> . . . . nor is there any evidence that the "neutral" point on a summated-rating scale necessarily corresponds to the mid-point of the possible scores . . . If in terms of research, our interest is in comparing the mean change in attitude scores as a result of introducing some experimental variable, such as a motion picture film, then the lack of knowledge of a zero point should cause no concern. Similarly, if our interest is in comparing the mean attitude scores of two or more groups, this can be done with summated rating scales . . . . Or if we wish to correlate scores on an attitude scale with scores on other scales or with other measures of interest, this can also be done without any reference to the zero point on the favourable-unfavourable continuum . . .

But he also warns:

> The absence of knowledge of such a point is a handicap only if our major interest is in being able to assign, on the basis of an attitude score, a single subject to the class of those favourable or unfavourable in attitude toward the psychological object under consideration.

If percentages pro- or anti- a particular group or phenomenon are to be quoted, then it is, of course, necessary to designate individuals as favourable or unfavourable in their attitude. Some method of defining the psychological neutral point therefore becomes essential. Techniques have been proposed whereby the information contained in the responses to individual items can be analysed to obtain a neutral point possessing a reasonable degree of psychological meaning.[15] However, such techniques can only be properly applied if the scale is reliable in the first place.

15 Suchman (1950).

Alternatively, an attempt could be made to establish the neutral point through the use of criteria from outside the scale. This generally needs to be planned for when the Survey questionnaire is being designed. On a crude level some indication of the neutral point could be obtained through asking a direct question, not included in the scale itself, concerning the general attitude towards the stimulus. A more acceptable practice is to ascertain from each respondent the strength of his feeling about an item as well as his degree of agreement or disagreement. This provides sufficient information to allow the use of intensity techniques[16] which reveal the area of neutrality concerning the attitude. Whatever the method adopted, the aim remains the same: to replace the mid-point of the scale with a psychological neutral point which has been established by means of a deliberate and clearly explicable procedure.

(iii) Assessment of Scales

It should be clear from the foregoing discussion that in order to present summated scales as measuring the percentage pro-, anti- and neutral in attitude towards an object such as Partition, two fundamental conditions must be met.

(a) The scales must themselves be reliable.

(b) The "neutral" point must be independently determined.

The four scales presented in Paper No. 97 in the form of percentage distributions meet neither of these conditions.

As we have already seen, three of the scales are insufficiently reliable for use in anything except, perhaps, preliminary exploratory hypothesis formation. The scale relating to Partition probably has sufficient reliability for use in an ordinal form, as in the latter part of Section IV, but not for the applied use of measuring the distribution of an attitude.

For none of the scales was any independent source sought for establishing a viable neutral point. There is a simple reliance, against the explicit advice of the authorities in the field, on the mathematical mid-point of the scales as an expression of neutrality of attitudes. Neither is there any warning to the reader that this procedure departs from accepted practice.

16 See Guttman and Suchman (1947).

The presentation of scales purporting to measure the precise distribution of attitudes on sensitive topics, when these scales meet neither of the fundamental conditions necessary for such measurement, is in our view a serious misuse of summated attitudinal scales.

(e) Interpretation and Labelling

In the two preceding sections we have expressed our reasons for believing the identification of the four factors presented in Section IV and the presentation of these in terms of absolute percentages of the population holding particular attitudes to have been mistaken. If, in fact, these attitudes have not been properly identified and measured, then their labelling and interpretation become matters of secondary importance.

Nevertheless, it is worth considering the issue in isolation, because we believe that the labels adopted in Paper No. 97 would have been unsuited to the findings of the Survey, even if the factor analysis and scale construction had been correctly undertaken. For the purposes of this section, therefore, we temporarily suspend our criticisms concerning analysis and accept the clusters of items selected by the authors as forming their four factors. Thus we can concentrate for the time being on how these factors are interpreted and labelled.

The interpretation of analysed survey data and the labels attached to particular factors are closely intertwined. Indeed, it would hardly be an exaggeration to regard the labels as a summary of the interpretation adopted. Labelling is thus a matter of judgement and not of technique.

To some extent labelling is bound to be arbitrary, but it ought generally to stay within the limits of common sense and the ordinary rules of language. Leaving to one side the major issue of whether attitudes should be labelled in a dichotomous manner, such as "Pro versus Anti" or "Support versus Opposition", when no neutral point in the scale has been established, we can consider the key words used in each of the labels adopted in Section IV.

There is no call to quarrel with the key word "anti" in relation to the attitude to partition. This appears to be an adequate and accurate reflection of the content of the items contributing to the factor. With regard to the IRA Motives

attitude, the principal query is whether this sub-factor exists as a separate entity. Assuming for the moment that it does, then "sympathy" is probably a reasonable label in relation to the content of the constituent items. It is the labelling of the other two attitudes which causes us concern.

From the items on which it is based, the attitude towards Northern Ireland Protestants would appear to be complex and somewhat diffuse. It involves a recognition of differences of outlook, approach to life and political aims together with a reluctance to accept that these Northern Ireland Protestant aims should determine the resolution of the Northern Ireland problem. Not surprisingly some of these items are correlated as closely with some of the Partition items as with each other. To label this attitude "anti" Northern Ireland Protestant is to simplify it unduly, to imply a coherence which the factor lacks, and to show a lack of awareness concerning the effect that labelling can have on the reader of Paper No. 97.

The labelling of the remaining IRA attitude is open to even stronger objections. The factor is based on three items, one of which concerns the effect of the existence of the IRA, one of which is ostensibly concerned with their motives, and only one of which mentions methods. To relate all three to IRA *activities* is thus an unjustified interpretation. To go further and describe the attitude as one of *support* is to adopt an interpretation which simply is not in accord with the normal usage of the English language. Support, in this context, means an actively positive state of mind, if not also implying concrete manifestations of agreement.[17] There is nothing in the three items to warrant the use of the word.

If the criterion of support for a political group is a failure to find its "methods totally unacceptable" or to regard its members as "a bunch of criminals and murderers", then most of the adult population of Ireland simultaneousy "support" Fianna Fail, Fine Gael, the Labour Party, and, for that matter, the Australian Country Party. On the same basis, many pressure groups, armed factions and illegal organisations around the world could claim widespread "support" in Ireland.

17 See the Oxford English Dictionary definition of "support" quoted in Part I (p. 30).

Obviously, the mere absence of outright condemnation should not be described as "support" and would be more accurately reflected in the use of more passive terminology such as "acquiescence in", "toleration of" or even "acceptance of". This is especially so when the authors themselves are at pains to point out (p. 98) that they "have no evidence that an attitude of support for IRA activities, as we have measured it, leads to any concrete actions, by way of monetary contributions or whatever, in support of the campaign of the IRA."

In passing, it should be noted that this denial of any link between attitude and action runs counter to most of the definitions of attitude quoted in Part I of this document. If attitudes are assumed not to influence behaviour, then much of the *raison d'être* of attitude research is removed.

Apart from the labelling of the factors, there are several less important points on which the authors' interpretation of their results appears unconvincing. The most serious of these is probably the discussion of the attitude towards IRA Motives. Here it is claimed (p. 100) that this cannot "be explained away as an alternative expression of the aspiration to reunification" on the grounds that one of the items referred to "aims (plural)" and that the other referred "to patriotic and idealistic characteristics." This evidence is inadequate for such a claim. Given that the majority of respondents themselves see reunification as highly desirable, it would not be unreasonable to assume that they would tend to perceive anybody else holding the same aspiration as being both patriotic and idealistic. Similarly the mere use of the plural in "aims" tells nothing about how these aims are actually perceived by respondents. Either the plural could be overlooked, or could be taken to encompass complementary goals such as British withdrawal. The only clear evidence concerning how the actual programme of the Provisional IRA is accepted comes in Section III, where only eight per cent of respondents selected the Provisional Sinn Fein policy of a four-province Federal United Ireland as their first choice solution. The real point at issue here is that if a deliberate choice was made to omit any reference to the actual aims of the IRA from the questionnaire, it is unwarranted to then make firm assertions concerning the meaning of responses in this area. A legitimate

case can be made in attitude research for leaving such questions vague, but the inevitable price is that uncertainty about the exact meaning of the results must be accepted.

(f) Presentation

Our criticisms of Section IV so far have concerned several aspects of conceptualisation, methodology and interpretation. Taken together, these would be enough to render the findings of this Section of Paper No. 97 invalid. The manner in which the Section is presented greatly exacerbates its shortcomings.

In the first place, it is common practice when presenting an analysis to allude to, if not to discuss fully, the methodological problems encountered, and to seek to justify the particular solutions adopted. As we have shown in this critique, there are several instances in Section IV where important issues have not been addressed. This makes it very difficult for even the informed reader to assess the strength of the evidence adduced.

More centrally, in discussing the desiderata of attitude research, we noted that validation of such work could never be absolute, and that the degree of validity achieved depends on the value of the concepts as much as on the consistency of the statistical results. Furthermore we concluded, in line with authorities in the field, that the presentation of this type of research should be in keeping with its interpretative and speculative nature.

In the discussion of methodology in Section II of Paper No. 97, it is stated, reasonably, that the process of pre-testing factors "greatly increases the confidence one may place in the validity of the measures one uses, since the judgement of validity is based not simply on *a priori* reasoning, but on empirical corroboration" (p. 23). Even at the beginning of Section IV itself there is a reference to a "gain in terms of our confidence in the validity of our composite measures" (p. 94). This too is an acceptable form for a statement of the general role of factor analysis in relation to validity.

However, as the section progresses this concern with the place of statistical evidence in corroborating *a priori* reasoning, and thus improving confidence in validity, gives way to much stronger expression. We find such statements as (p. 96)

". . . the homogeneity of the subsets of items has been empirically established by means of the factor analysis", (when in fact the correlation between some of the "homogeneous" items is as low as .09). Similarly the view that the attitude towards the IRA might possess more than one dimension (p. 97) "proved correct in that a separate factor analysis of the five items produced two distinct factors" (when, as we have shown, the evidence for the existence of these two separate factors is totally lacking).

These claims, which we regard as excessive, still relate to matters of technique, and in themselves are perhaps not very serious. It is when the authors turn from describing their methods to reporting their results that the tone of the section becomes even less cautious, and implies that the findings are fully validated. Examples of over-confident presentation can be found in the discussion of each of the attitudes. However, it is clearest and most serious in the pages devoted to Attitudes to the IRA (pp. 97-100). The three full paragraphs quoted below illustrate in context the dearth of either qualifying clauses or verbs expressing uncertainty, the consistent use of definite verbs including the simple "is", and the frequent occurrence of strong adverbs.

> The interpretation of the grouping of the items and thus of two dimensions identified is that the first three items all refer to aspects of IRA activities: without their activities the problem would be worse (Item 1), the methods underlying their activities are totally unacceptable (Item 2) and their activities make them a bunch of criminals and murderers (Item 3). Taken together the three items clearly represent *Support for versus Opposition to IRA Activities.*
>
> The second set of IRA items relates to the motives of the IRA: support for their aims (Item 1) and attribution of patriotic and idealistic characteristics to them (Item 2). Taken together these two items represent an attitude to IRA motives – an attitude of *Sympathy with, versus Rejection of IRA Motives.* If our identification and interpretation of a two-dimensional attitude to the IRA is valid then the two dimensions should have contrasting distributions in the population. This is in fact the case.

The contrast is first apparent when we look at the average scores on each dimension – these are 3.24 in the case of support for activities and 3.86 in the case of sympathy for motives. Both means fall below the mid-point of four but the mean value of support for IRA *motives* is substantially closer to the mid or neutral point. The contrast between the distributions of the two dimensions of attitude to the IRA is also apparent in Table 43 where the percentage distributions are given on the basis of rounding the scores to integer values. There are contrasts at almost every level of the scale in Table 43. If we focus on the summary percentages, the picture emerges quite clearly. In the case of attitude to IRA *activities* (support *versus* opposition), 61 per cent are on the opposed side of the neutral point compared with 34 per cent on the rejection side of the neutral point in the case of attitude to IRA *motives.* Correspondingly, 21 per cent are on the support side of the neutral point in regard to attitude to activities compared with 42 per cent on the sympathy side of the netural point in the case of attitude to motives.

The identification and measurement of these two distinct dimensions is of crucial importance in assessing attitudes to the IRA. Attitude to IRA activities is a clear and unambiguous measure. Given the nature of the attitude in question it is necessary to be particularly careful and precise in discussing its distribution. The majority of people (61 per cent) are opposed to IRA activities as we have measured this attitude. Overall opposition is also evident in the average score of 3.24 already noted. A further 19 per cent are neutral. In regard to the remaining 21 per cent support for IRA activities, it should first of all be noted that this includes 13 per cent who are slightly supportive as against eight per cent moderately to strongly supportive. This having been said, the stark fact remains that 21 per cent of the population emerge as in some degree supportive in their attitude to IRA activities. It should also be emphasised that we have no evidence that an attitude of support for IRA activities, as we have measured it, leads to any concrete actions, by way of

monetary contributions or whatever, in support of the campaign of the IRA. The context in which these figures for attitude to IRA activity (61 per cent opposition, 19 per cent neutrality and 21 support) should be interpreted is that these attitudes are part of the overall approach of people in the Republic to the Northern Ireland issue. As such it must be acknowledged that, on this evidence, opposition to IRA activities is not overwhelming and certainly does not match the strong opposition so often articulated by public figures.

If these paragraphs, like the rest of the section, do not imply an assumption of full validity for the scales, then it is difficult to conceive of a style of presentation that would. In other words, the findings in this section are presented as scientifically measured facts. Even if the statistical evidence were more reliable, the most that should be claimed is that the authors' interpretations, while inevitably speculative, are consistent with the statistical results. This is particularly so given the evolution of the scales being presented. Because the final composition of the scales was determined only after the responses were received, the hypotheses embodied in the scales must be regarded as being formed from the main survey data. It is therefore logically untenable to claim that these data corroborate or confirm the hypotheses.

While our main criticism of the presentation concerns the unduly dogmatic tone of the reporting, there is another point illustrated in the quotations above which should not go unremarked. This is the use throughout of the present tense, implying that a survey carried out in the Summer of 1978 represents views still held, in the same proportions, in the Autumn of 1979. Nowhere in the relevant sections is there any discussion as to whether events in the intervening period might have led to a significant change in attitudes.

Because of the innate characteristics of the methodology used, because the application of that methodology diverges at several points from accepted practice, and because of the time-lag between survey and report, the reader ought to be warned that the results set out in Section IV should be treated with caution. The complete failure to state, or even imply,

that such caution is necessary is not merely a question of using an inappropriate style. When such sensitive subject matter is involved, it could encourage the misuse of the findings by those who are unaware of the limitations inherent in the calculations.

5 SUMMARY

(a) The sample on which Paper No. 97 was based appears to have been well drawn, and the questionnaire properly administered. It follows that the answers to individual questions collected in the Survey can be taken as representative of the responses of the population as a whole.

(b) The eliciting of opinions concerning alternative solutions to the Northern Ireland problem and the policy options open to the Irish and British Governments was well conceived and competently executed. Section III of the Paper, in which these results are reported, is, in our opinion, a solid and useful contribution to knowledge.

(c) The methodology of attitude research is well established and an extensive literature lays down clear guidelines for its application. Section IV seeks to apply this methodology and it can therefore be judged largely on the extent to which it adheres to these guidelines and on how well the authors justify any major departure from them.

(d) In any social research, it is important that the concepts used be clear. In Section IV there appears to be considerable conceptural confusion. Among several instances where clarity is lacking, two stand out. Unnecessary ambiguity is caused by using the simple initials "IRA", with their complex historical associations, when the object of the research would appear to be attitudes to the Provisional IRA. More fundamentally there is uncertainty about the basic concept of the entire section, the nature of an attitude. This uncertainty is shown in various ways, but it is reflected most obviously in the wording of the questions chosen to tap the attitudes being studied.

(e) Attitudes are commonly defined in social psychology as emotional dispositions towards particular objects or stimuli. Questions designed to identify or measure attitudes should

therefore be concerned with the respondent's feelings towards the stimulus, not with his considered opinions, beliefs or behaviour. Of the 17 questions in Section IV, far too many are worded so as to invite beliefs or thought-out positions, and there is no evidence that the answers to these are correlated with basic emotions.

(f) In attitude research, several questions, or items, are usually taken together as representing and measuring an attitude. To select items which group well together for this purpose, factor analysis is normally applied at the "pretest" stage of a study. It is generally used again on the results of the main survey, to confirm that the items fall into the expected clusters. Like other statistical techniques, factor analysis should be applied according to well established rules. We have shown that these rules are breached in Section IV, with the result that the factor analysis of the main survey data contains at least four serious errors. The final grouping of items into three factors, one of which is then divided into two sub-factors, is quite unacceptable.

(g) Following factor analysis, the results of each cluster of items are combined to produce summated scales. These can be tested for reliability. Only the scale measuring Attitude to Partition shows a level of reliability which would justify its use in some form of further analysis. The other three scales presented are simply not reliable enough to be regarded as measuring any attitude consistently.

(h) Even where reliable scales have been obtained, the mid-point of the scale cannot legitimately be taken as the division between favourable and unfavourable attitudes towards the stimulus in question. This does not matter much if the scale is to be used ordinally for purposes of comparison. However, if the scale is to be presented as a cardinal measure, then it is essential to establish the psychological neutral point independently. This neutral point must be substituted for the mid-point of the scale before the percentages pro- or anti- a particular stimulus can be calculated. In Section IV, there is total reliance on the mid-points, and no attempt is made to establish independent neutral points for the scales. Thus the percentages quoted in relation to each attitude would be misleading, even if the scales themselves were reliable. In the

event, of course, the presentation as cardinal measures of scales with such low levels of reliability cannot be accepted.

(i) The interpretation of the scales, as embodied in the labelling adopted, involves a debatable use of the English language. The content of the items on which they are based does not, in our view, justify the use of the word "support" in relation to the activities of the IRA, nor of the word "anti-" in relation to Northern Ireland Protestants.

(j) The style of presentation in Section IV implies that the scales produced are fully validated. In view of the demonstrably low reliability of the scales, and of grave doubts whether the concepts used were sufficiently clear to be measurable, this assumption of validity is unjustifiable. The definite and over-confident tone used is therefore highly inappropriate, and serves to compound the methodological faults already summarised.

6 CONCLUSION

On the basis of a careful comparison of the methods used in Section IV with the standard practices of the methodology employed, we have indicated serious shortcomings in conceptualisation, selection of questions, application of factor analysis, construction and use of scales, interpretation and presentation. Each, on its own, would cast doubt on the findings of Section IV. Taken together, they completely invalidate the findings. In our view, the evidence adduced in Section IV does not justify the conclusions reached.

The issue is not that the figures presented are wrong, in the sense that a more rigorous analysis of the data would produce alternative, more accurate, percentages. It is rather that the scales presented in Section IV are inadequate measuring devices. Standard tests show them to be statistically unreliable, while their meaning in relation to the attitudes they are designed to represent is problematic. Quite simply, these scales cannot legitimately be used to determine the percentages of the population holding favourable or unfavourable attitudes towards Partition, IRA Activities, IRA Motives, or Northern Ireland Protestants.

Because the sample was satisfactory and the questions appropriately selected, we regard Section III of Paper No. 97 as a useful and instructive exercise. Even parts of Section IV, where the Partition scale is used ordinally to compare the strength of the attitude with various socio-demographic features or with the holding of certain opinions, could have been a potentially valuable contribution to an understanding of Irish problems. It is unfortunate that the virtues of a large part of Paper No. 97 tend to be obscured by the deficiencies contained in the first eight pages of Section IV.

Note: The agreed format for this publication precludes our replying to the points raised in Part III. We should like to make clear that, having studied Part III, we stand by our major conclusions as set out in Part II.

REFERENCES

ALLPORT, G. W., 1935. "Attitudes", in C. Murchinson (ed.) *Handbook of Social Psychology*, Worcester, Mass: Clark University Press.

CHILD, D., 1970. *The Essentials of Factor Analysis*, London: Holt, Rinehart and Winston.

EDWARDS, A. L., 1957. *Techniques of Attitude Scale Construction*, New York: Appleton-Century-Crofts.

GUILFORD, J. P., 1954. *Psychometric Methods* (2nd edn.), New York: McGraw-Hill.

GUTTMAN, L. and E. A. SUCHMAN, 1947. "Intensity and a Zero Point for Attitude Analysis", *American Sociological Review*, 12, pp. 57-67.

HOLZINGER, K. and H. HARMAN, 1941. *Factor Analysis*, Chicago: University of Chicago Press.

KAISER, H. F., 1960. "The Application of Electronic Computers to Factor Analysis", *Educational and Psychological Measurement*, 20, pp. 141-151.

KATZ, D., 1966. "Attitude Formation and Public Opinion", *Annals* 367, pp. 150-162.

KENDALL, M. and A. STUART, 1968. *The Advanced Theory of Statistics, Volume III Second Edition*, London: Charles Griffin.

KERLINGER, F. N., 1973. *Foundations of Behavioural Research* (2nd ed.), London: Holt, Rinehart and Winston.

NUNNALLY, J. C., 1970. *Introduction to Psychological Measurement*, New York: McGraw-Hill.

OPPENHEIM, A.N., 1966. *Questionnaire Design and Attitude Measurement*, London: Heinemann Educational Books Ltd.

OSTER, G., 1979. "A Factor Analytic Test of the Theory of the Dual Economy", *Review of Economics and Statistics* 61, pp. 33-39.

RUMMEL, R. J., 1970. *Applied Factor Analysis*, Evanston: Northwestern University Press.

SUCHMAN, E. A., 1950. "The Intensity Component in Attitude and Opinion Research". Chapter 7, in S. Stouffer, L. Guttman, E. Suchman, P. Lazerfeld, S. Star, and J. Clausen, *Studies in Social Psychology in World War II: Measurement and Prediction*, Princeton: Princeton University Press.

TATSUOKA, M. M., 1971. *Multivariate Analysis: Techniques for Educational and Psychological Research*. New York: Wiley.

TAYLOR, C. C., 1977. "Principal Component and Factor Analysis" Chapter 4 in C. A. Ó Muircheartaigh and C. Payne, (eds.), Exploring *Data Structures*, Chichester: Wiley.

VAN DE GEER, J. P., 1971. *Introduction to Multivariate Analysis for the Social Sciences*, San Francisco: Freeman.