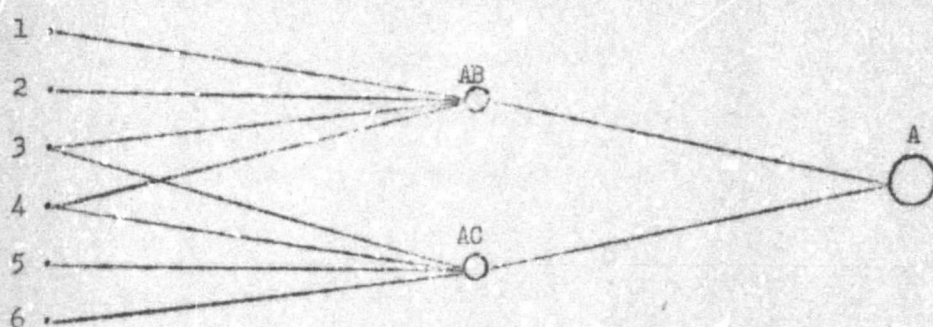


In the previous chapter it was stated that it was reasonable to assume that the systematisation achieved by a first-order analysis of the responses to the many and varied questions concerning behaviour, which constitute the majority of current personality schedules, will in most instances not be determined in terms of the degree to which individuals possess single temperamental characteristics, but rather as the resultant of a combination of these variables in different strengths. It was further postulated that if factors did emerge in a second-order analysis of these data, their interpretation would hinge around the common elements in the first-order factors and there was the possibility that this interpretation would reveal a more basic behaviour pattern. A diagrammatic representation of the argument is given below.

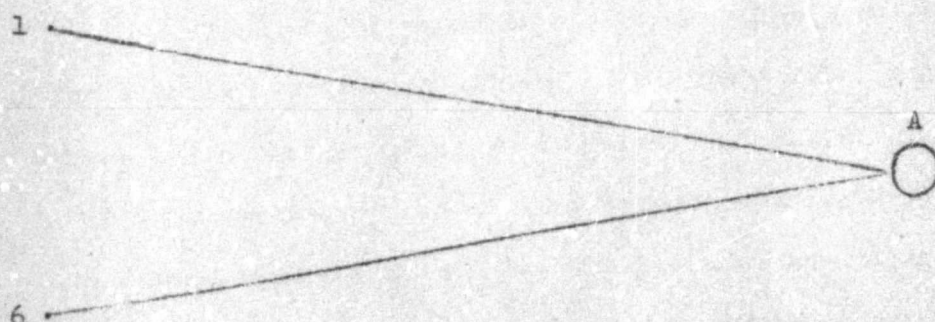
Figure I



Unselected personality items. First-order factors. Second-order factor.

By a careful selection of the personality items it may be possible to circumvent the first systematisation and to obtain Factor A directly. This process could be sketched as follows:

Figure II.



Selected personality items.

First-order factor.

If this should be the outcome of this investigation it would indeed be a compelling argument in favour of the stated hypothesis.

Unfortunately such a result cannot be relied upon because of the extreme difficulty of accurately defining and describing the facets of so complex an entity as human behaviour.

If we can demonstrate the validity of, and unambiguously define even one clear-cut and basic behaviour pattern we will have made a contribution to the field of personality assessment.

For the sake of clarity the five concepts to be covered in this investigation are recapitulated briefly.

Factor A. Emotionally Stable (Preliminary Investigation)

Outgoing adjustive responses facilitated by emotionality such as agreeableness, co-operativeness, confidence, and general sociable behaviour.

Factor B. Primary Function (Preliminary Investigation)

Quick and vigorous extravertive responses including impulsiveness, restlessness, and gregariousness.

Factor C. Activity (Preliminary Investigation)

Purposive behaviour responses characterised by determination, vigour, and enthusiasm.

Factor D. Emotionally Unstable (Preliminary Investigation)

Non-adjustive responses including social and thinking introversion, moodiness, and seclusiveness.

Factor E. General Emotionality (Heymans-Wiersma scheme)

A warm feeling tone and readily expressed emotions including sympathy and demonstrativeness.

The following personality items were selected to cover these concepts in this investigation. The concepts embodied in Factors A (Emotionally Stable) and C (Activity) are slightly different in the alternate rotation but all these concepts are covered so that these factors could appear in either form in the new investigation. It was attempted to make the selected personality items as unambiguous as this field of investigation will allow.

Word List of Descriptive Adjectives.

<u>Positive Pole</u>	<u>Negative Pole</u>
1. Cheerful	Depressed
2. Talkative	Uncommunicative
3. Self-confident	Lacking in Confidence
4. Even-tempered	Moody
5. Sympathetic	Detached
6. Seeks Company	Avoids Company
7. Happy-go-lucky	Serious
8. Energetic	Lethargic
9. Unselfconscious	Selfconscious
10. Enthusiastic	Indifferent
11. Agreeable	Disagreeable
12. Lively	Sedate
13. Steady Worker	Spasmodic Worker
14. Socially at Ease	Socially Uneasy
15. High-strung	Placid
16. Co-operative	Unco-operative
17. Impulsive	Cautious
18. Persevering	Quitter
19. Emotionally Stable	Emotionally Unstable
20. Demonstrative	Undemonstrative
21. Tranquil	Agitated
22. Quick Worker	Slow Worker
23. Prompt Starter	Procrastinator
24. Decisive	Indecisive

Some of the refinements introduced into this word list may be mentioned here. Reserved was first chosen as the opposite of Talkative. On second thoughts, since Reserved implies not only a restraint in speech but also in action, Uncommunicative was substituted. Similarly, Procrastinator was substituted for Dawdler as the opposite of Prompt Starter, as Dawdler implies not only a reluctance to start a given task but also spasmodic work once the job has been started. Seeks Company and Avoids Company were substituted for Sociable and Unsociable respectively as they appeared to have a less ambiguous interpretation.

As a check of the individual behaviour items chosen and of the experimental design as a whole, steps were taken to ensure that any classification which emerged as a result of the factor analysis would not represent only some arbitrary categories imposed by the investigator. It seems logical and entirely possible that if we were to construct a temperament rating schedule whose items were composed of (say) three distinctly different behaviour characteristics and any number of synonyms of these three items that a factorial investigation of the responses would yield only three factors. Such studies can hardly be regarded as contributing anything to our knowledge of human behaviour.

To avoid this pitfall the 24 descriptive adjectives for the positive pole were arranged in all possible combinations of two, yielding 276 pairs. This was also done for the adjectives for the negative pole and the total of 552 pairs were submitted to 6 independent judges. These judges were all graduate students in psychology or practising psychologists. They were asked to write down the numbers of those pairs of words in which they considered the two words to be synonymous or in which they considered that the words described behaviour characteristics which could not be distinguished.

The results of this "Synonym Investigation" are given in Table XIX below.

Table XIX.

	Word-pairs.	Judges Regarding the Word-pairs as Synonymous in Positive Pole.						Judges Regarding the Word-pairs as Synonymous in Negative Pole.						Posi- tive Pole.	Nega- tive Pole.
		1	2	3	4	5	6	1	2	3	4	5	6	Total	Total
1	Even-tempered Tranquil	1				1	1							3	0
2	Even-tempered Emotionally Stable	1				1								2	0
3	Energetic Quick Worker						1							1	0
4	Self-confident Socially at Ease		1											1	0
5	Talkative Lively		1											1	0
6	Socially at Ease Tranquil						1							1	0
7	Unselfconscious Socially at Ease	1	1		1			1		1				3	2
8	Steady Worker Persevering		1											1	0
9	Energetic Lively		1											1	0
10	Cheerful Even-tempered												1	0	1
11	Enthusiastic Lively	1						1						1	1
12	Emotionally Stable Tranquil		1						1					1	1
	Total Number of Word-pairs Regarded as Synonymous by Each Judge	4	6	0	1	2	3	1	2	0	1	0	1		21

It will be seen from the table that only 12 of the total number of the 552 word-pairs obtained from the positive and negative poles were regarded by any of the judges as being comprised of two synonymous words. Of these 12 pairs, 8 were only regarded as containing synonymous words for the positive pole, 1 as containing synonymous words for the negative pole and 3 were regarded as containing synonymous words for both poles. One would have expected that the word-pairs regarded as containing synonymous words for the positive pole would also have been regarded as containing synonymous words for the negative pole. That this is not always the case is due to the extreme difficulty in obtaining exact antonyms to represent the positive and negative poles respectively, especially when there is only a fine shade of distinction between the two words comprising a pair.

It will be seen from the table that the total number of pairs regarded as containing synonymous words ranged from 0 for judge 3 to 8 for judge 2. On the basis of these results it cannot be said that a structure is being imposed upon the experiment. However, 2 of these 12 pairs (No. 1, Even-tempered and Tranquil) and (No. 7, Unselfconscious and Socially at Ease) were regarded by as many as 3 out of the 6 judges as containing synonymous words. One word from each of these pairs was therefore removed from the word list, Tranquil and Unselfconscious and conversely Agitated and Selfconscious were removed. The final word list of adjectives therefore contains 22 words for both the positive and negative poles.

The Method of Assessment.

Each of the many methods which have been devised for assessing human behaviour or preferences have specific advantages and disadvantages. There is no absolute criterion whereby an investigator can select the "best" method of assessment; he must choose the method which seems best suited to his particular investigation.

Rating scales are very commonly employed in which the rater has to classify the trait to be assessed (say) talkative, as present, average, or absent in the individual, or a more extended numerical or descriptive scale may be employed in which each step in the scale is

designated by a number or phrase which describes particular behaviour patterns, or finally a combination of a graphic and descriptive scale¹ illustrated below may be used.

Directions:

Adjustment to Situations:	Sometimes at a loss in fa- miliar situations	Slow to adapt to new situations	Learns new arrange- ments fairly soon	Quick to adjust to new routine	Very quick to respond to emer- gencies
---------------------------------	--	--	---	---	---

These scales enjoy a number of advantages in that they are easily grasped, quickly filled in and require little space so that a large number of traits can be included in one schedule. The number of scale intervals can also be varied at will though in practice they are restricted to the number of steps that can be clearly distinguished by the judges in a reasonable time. Fortunately there is some empirical evidence as to the number of scale intervals that can profitably be employed, the consensus of opinion² seems to be that 5, 7, or 9 intervals can be used depending upon the preciseness with which the trait is defined.

In some rating scales the intervals used are completely arbitrary in which case the investigator would have some reservations concerning the equality of the scale intervals and should devise some method of equating the results before comparing the assessments of different individuals. On the other hand devices such as those employed in the methods of rank order, equal appearing intervals, or paired comparison may be used to determine the scale intervals. In an investigation of this sort, where this would be only a preliminary step, the labour involved would be prohibitive.

There are some sources of error that are common to all rating scales such as end effects due to a large number of diversified judgments being lumped together in the categories at the two extremes of the scale, when raters are forced to fit their assessments to a stipulated number of scale intervals.

¹ Greene, Edward B. (8) p. 700

² *Ibid.*, p. 704

Guilford, J.P. (9) p. 268

There are a number of constant errors such as the error of leniency where raters tend to rate the individuals with whom they are well acquainted above average in the desirable traits, or the error of central tendency where judges avoid using the extremes of the scale. Some judges may constantly over rate or under rate individuals in traits as compared with the average ratings of all the judges. These errors can sometimes be identified by an examination of the frequency distributions and corresponding adjustments made. Errors which are more difficult to identify are those caused by the halo effect where the judge rates the individual according to some general attitude to the personality as a whole and not with due regard to the specific traits under consideration, or logical errors where judges give similar ratings on traits which seem to them to be logically related.¹ It seems likely that these last two sources of error will never be completely eradicated.

The method of assessment finally decided upon for this investigation is a modified form of the paired comparison technique. The 22 adjectives of the positive pole were combined in all possible groups of 2 yielding 231 pairs. Similarly 231 pairs were obtained for the descriptive adjectives for the negative pole giving a total of 462 pairs. These were presented randomly in a single schedule, and the order of the appearance of the word in each pair (i.e. whether it appeared first or second) was also randomised. For each pair the rater will underline that adjective which, in general, is more descriptive of the behaviour of the person being assessed. The rater will be urged to make a choice of one word in each pair whenever this is possible, and only to mark both words of a pair when he considers that they both apply equally to the person being assessed. On the other hand, both words of a pair can be left unmarked if the rater is convinced that neither word is in any way descriptive of the behaviour of the person being assessed. A copy of this Temperament Rating Schedule, which shows the exact instructions given to the raters, is included in

¹ Guilford, J.P. (9) pp. 272 to 275

the appendix.

The score for each item will be the number of times that it is underlined; and for the positive and negative poles respectively, the scores will be comparable from item to item and from person to person since each word is judged in relation to the same set of standard words. L.L. Thurstone used comparative judgments to determine the relative interest in ten vocational fields. Each vocational field was represented by ten different occupations, and each occupation was compared with every other, the subjects underlining the preferred occupation. The relative interest in a vocational field was represented by the number of times that the occupations pertaining to that field had been underlined. With regard to the scores so obtained he states: "When an ordinary raw score in a psychological test is to be interpreted, one turns naturally to a table of norms in order to ascertain whether the given score is high, average, or low. In a complete table of norms, one can ascertain the percentile rank, or standard score equivalent, for the given raw score in comparison with any specified standardisation group, such as an age group, an occupational group, or the students in some particular school. Such tables for different occupational groups are useful for some types of problems but, since the scores in the present schedule are comparable, the profile can be interpreted directly to find a man's relative interests in different fields without comparing the profiles for groups of people."¹

In this investigation each item is scored separately and does not form a part of any larger group. For the positive and negative poles respectively we will obtain, therefore, an estimate of the relative degree to which each item is descriptive of the individual's behaviour. Intercorrelations calculated between these items for the group of subjects assessed will give the relationships between these items in the experimental population. In other words they will give an estimate of the degree to which each item is associated with every other item in the individuals comprising the experimental

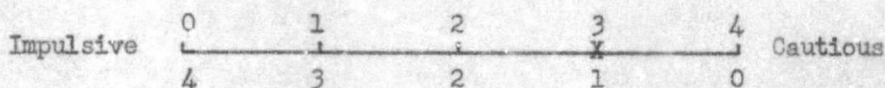
¹ Thurstone, L.L. (35) p. 9

population. These data will then be subjected to a factor analysis in an attempt to discover the underlying order in this domain.

Throughout this investigation the adjectives from the positive and negative poles will be treated separately. In no instance will the pairs presented in the schedule involve comparisons between words from these different word lists, and the results from the two word lists will be factored separately. This procedure was followed as it seemed to afford some distinct advantages.

In a number of rating scales the opposite end of the continuum is defined and a single assessment determines the score on two behaviour characteristics. In Figure III the assessment could represent a score of 3 for cautiousness or 1 for impulsiveness. The one score is necessarily the complement of the other.

Figure III.



In this investigation the opposite end of the scale is not assumed or imposed and it can be determined experimentally whether the scores obtained for a word and its postulated opposite are indeed complementary. Factorially, the question will be phrased in these terms: Will the separate factorings of these two different word lists yield comparable factors?

If there are factors which appear in complementary pairs in the two investigations, their interpretation will be direct and unassailable. If the two factorings do not yield comparable factors, a question of practical importance in the compilation of rating scales can be set:

What behaviour items do represent the opposite poles of a continuum? This can be demonstrated experimentally by a third factor study of the correlations between the words in both word lists, or by a study of the correlations between the scores on the actual factors obtained.

Another advantage of this method should be the greater

reliability of the scores, since the score for each behaviour item is not dependent on a single scale rating but will be the result of 21 separate judgments. A check for the consistency of the ratings will be described later.

It was considered that by treating the two word lists separately the effects of some of the constant errors inherent in rating scale assessments would be minimised. The lists were so designed that, in so far as it was compatible with the concepts to be covered, the desirable and socially acceptable behaviour characteristics appeared in the word list for the positive pole and vice versa. In general, therefore, the judgments called for in the schedule are between two socially desirable traits or between two socially undesirable traits. The rater is forced to make assessments some of which will be flattering and others which will be unflattering to the individual. This should reduce the halo effect and the effects of the error of leniency since the rater is not in a position to consistently check the desirable characteristic as opposed to the undesirable one or vice versa. If a particular judge has a relatively large number of pairs in which both items have been underlined, and/or a large number where both items have been left unmarked, these schedules will be discarded.

In making comparative judgments of this sort the rater is not constantly aware of a scale as such, and it seems logical that errors of central tendency or a constant bias in using either the upper or lower scale intervals will be largely eradicated.

It is difficult to conceive of any method of subjective assessment in which the effects of the logical error are completely avoided. Any presuppositions concerning the logical coherence of the traits to be assessed, which some of the judges may hold, are unlikely to be affected by the method of presentation of the traits. There does not seem to be any reason why the effects of this error should be more marked in this method of assessment than in any other.

A Check of Consistency.

Fourteen pairs appear twice in the schedule in order to afford a check of the consistency of the assessments of the individual judges. Each of these pairs will, of course, only contribute once in the scoring of the schedule, in each instance the first appearance of the pair will be scored.

Seven pairs were chosen from the positive pole and seven from the negative pole in such a way that each of the twenty-two behaviour items in this schedule is represented either on the positive or negative pole or in some instances on both. The fourteen pairs of words are given below in the order of their first appearance in the schedule.

Page 2	1. Selfconfident Prompt Starter	2. Detached Spasmodic Worker
	3. High-strung Persevering	4. Uncommunicative Emotionally Unstable
Page 3	5. Energetic Quick Worker	6. Lacking in Confidence Moody
Page 4	7. Seeks Company Co-operative	8. Sedate Cautious
Page 5	9. Steady Worker Demonstrative	10. Serious Socially Uneasy
Page 6	11. Cheerful Sympathetic	12. Indifferent Slow Worker
Page 7	13. Talkative Even-tempered	14. Disagreeable Indecisive

The first appearance of the first two pairs on every page is in the tenth and eleventh positions in the first column, and the check for the pairs appear in the first and twentieth positions in the fourth column of each respectively. An example of this arrangement is given for page three.

Figure IV.

[illegible]

The word pairs and their checks were arranged in this orderly way for ready reference and to facilitate the comparisons. For seven of these pairs (numbers 1, 2, 3, 7, 8, 11, and 12) the order of presentation of the two words in the pair is the same as that in the check. In other words, pair number one is Selfconfident and Prompt Starter and the check is in the same order, Selfconfident and Prompt Starter. For the remaining seven pairs (numbers 4, 5, 6, 9, 10, 13, and 14) the words in the checks are in reverse order. Thus pair number four is Uncommunicative and Emotionally Unstable and the check is Emotionally Unstable and Uncommunicative.

For each judge the number of discrepancies in the scoring of the fourteen pairs and their respective checks will be calculated and a criterion of consistency decided upon. For example, this may be set as not more than three discrepancies per schedule. Any assessment which does not conform to this standard of consistency will be discarded.

The Raters.

The material for this investigation was collected at the University of Chicago in the United States of America. The idea was at first entertained of securing an experimental population by asking graduate students in psychology to furnish self ratings on the schedule. On second thoughts it was decided that this population would be too highly selected for an investigation of this sort in which it is essential to cover a wide range of behaviour on both the socially acceptable and socially unacceptable traits.

It was finally decided to ask those people who during the course of their studies or their work dealt with the problems of human behaviour, such as graduate students in psychology, practising psychologists, vocational guidance counsellors, and psychiatrists, to rate their friends and patients. In this way the experimental population would be expanded to include not only students but people, chiefly returned soldiers, seeking vocational guidance, members of the general population and some psychoneurotics. It was hoped that this would ensure a wider range of behaviour and some cases falling at the extremes of the scales.

CHAPTER VI.

THE EXPERIMENTAL RESULTS (POSITIVE POLE).

- (a) Description of the selection of the two hundred completed Temperament Rating Schedules used in the experimental investigation.
- (b) Factor analysis of the twentytwo behaviour traits comprising the positive pole of the variables.
- (c) Interpretation of the factors.

A sample of 200 completed Temperament Rating Schedules was obtained following, in general, the procedures laid down in the previous chapter entitled: "The Experimental Design".

Because of the difficulty of obtaining Raters, however, it was necessary to include amongst these not only practising psychologists, psychological counsellors, psychiatrists and graduate students in the department of Psychology, as at first stipulated, but also graduate students from other departments in the Division of the Social Sciences such as the department of Education, and from the department of Human Development administered by the Division of the Biological Sciences and the Division of the Social Sciences at the University of Chicago. It was also necessary to include 35 undergraduate students.

Some students had transferred from the department of Psychology to that of Human Development or vice versa since these are closely related. All the Raters had a background of courses in general psychology, the vast majority were graduate students in psychology with varying degrees of experience in vocational guidance, psychological counselling and psychotherapy. Many graduate students were part-time members of the university lecturing staff, some were on the staff of Clinical Psychologists in the Neuropsychiatric Wards of Hines Veterans Hospital, others were part-time members of the staff of the Veterans Administration Vocational Guidance Bureau. The remainder of the Raters were full-time practising psychologists. Because of the considerable overlap in the categories "graduate student", "staff" and "practising psychologists", it was considered that a numerical analysis of the composition of the group of Raters would be both

unsatisfactory and misleading.

The criterion of consistency mentioned in the previous chapter was adhered to. Each Rater was allowed a maximum of three discrepancies in the scoring of the fourteen word-pairs and their respective checks. In other words, the consistency score is at least 11 out of 14 for each of the 200 schedules comprising the sample. A distribution of these consistency scores is given below:

Table XX.

Consistency Scores for Sample Group of 200 Schedules.

Score.	Number.
11	50
12	49
13	57
14	<u>44</u>
Total Number	200

A group of 68 completed Temperament Rating Schedules was discarded on the grounds of the consistency criterion. A distribution of these consistency scores is given below:

Table XXI.

Consistency Scores for Discarded Group of 68 Schedules.

Score.	Number.
6	1
7	2
8	10
9	27
10	<u>32</u>
Total Number	68

The Temperament Rating Schedule was so designed that, in so far as it was compatible with the concepts to be covered, the Raters were required to make judgements between two socially desirable traits or between two socially undesirable traits. It was thought that this would reduce the halo effect and the effects of the error of leniency since the Rater was not in a position to consistently check the desirable characteristic as opposed to the undesirable one or vice versa.

The only way in which this could be accomplished by the Rater would be to consistently omit all the word-pairs describing socially undesirable characteristics and to mark both words in the word-pairs describing socially desirable characteristics or vice versa. In order to avoid

this pitfall all completed Temperament Rating Schedules which had more than 24 word-pairs omitted or more than 24 word-pairs double marked were discarded.

Since there are 462 judgments to be made in the schedule, the number of "omissions" or "double marks" allowed each comprise only 5.4 per cent of the total number of responses. On the other hand, this procedure has the salient psychological advantage of allowing the Rater some latitude in the difficult discriminations which he has to make, by not continually forcing his decisions when he genuinely believes that either both words in the pair are descriptive of the subject's behaviour, or that neither are descriptive of the subject's behaviour. Twentyseven completed Temperament Rating Schedules were discarded because of excessive "omissions", 1 because of excessive "double marks", and 1 because of excessive "omissions" and excessive "double marks".

Distributions are given below of the number of "omissions" and of the number of "double marks" in the sample group of 200 completed Temperament Rating Schedules:

Table XXII.

Number of Omissions.	Number of Schedules.	Number of Double Marks.	Number of Schedules.
0	35	0	70
1	9	1	17
2	7	2	11
3	10	3	10
4	9	4	12
5	11	5	11
6	7	6	5
7	7	7	9
8	2	8	2
9	8	9	2
10	7	10	5
11	11	11	7
12	1	12	8
13	5	13	2
14	4	14	3
15	5	15	7
16	5	16	1
17	5	17	3
18	5	18	2
19	5	19	3
20	6	20	2
21	4	21	2
22	13	22	0
23	3	23	2
24	16	24	4
Total	200		200

Author Baehr Melany Erna

Name of thesis An Investigation Into The Determinants Of Temperament. 1987

PUBLISHER:

University of the Witwatersrand, Johannesburg

©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.