

A PROBABILISTIC THEORY OF PREFERENTIAL CHOICE

by

Joseph Louis Zinnes

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
1959

Doctoral committee:

Professor Clyde H. Coombs, Chairman
Associate Professor J. David Birch
Professor Paul S. Dwyer
Doctor Robert A. Hefner, Jr.
Assistant Professor James V. McConnell

ACKNOWLEDGMENTS

This study was supported by a grant from the Feint Research Institute to the University of Michigan Research Institute project on Psychological Color Space and Color Preference.

I wish to express my overall indebtedness to Professor Coombs. As the inspiring mentor he will recognize the source of most of the ideas expressed in this study.

Acknowledgments are also due to Professor Birch for his thoughtful and thorough analyses; to Doctor Hefner, the principal investigator of the U.M.R.I. project, for his continuous contributions to this study from its earliest inception to its completion in the present form; and to the other members of the committee for their gracious cooperation and suggestions.

To Mrs. Percival Tomlinson I owe special thanks for permitting her husband to spend Saturday nights persevering with an IBM 650; and, of course, thanks go to the IBM 650 for its determined efforts to yield the necessary numerical results before drawing its last breath on the University of Michigan campus.

Finally, I wish to acknowledge my great debt to the Undergraduate Library at the University of Michigan for obvious reasons.

TABLE OF CONTENTS

LIST OF TABLES	iv
LIST OF FIGURES	vii
LIST OF APPENDICES	viii
Chapter	
I. INTRODUCTION.	1
II. PROBABILISTIC THEORIES OF CHOICE AND STRONG STOCHASTIC TRANSITIVITY.	16
III. DEVELOPMENT OF A PROBABILISTIC QUADRANT I MODEL. . . .	42
IV. THE SIMILARITY OF PROBABILISTIC THEORIES OF CHOICE. . .	79
V. REVIEW OF THE EXPERIMENTAL LITERATURE ON COLOR PREFERENCES	92
VI. EXPERIMENTAL METHOD	99
VII. RESULTS	119
VIII. DISCUSSION.	174
IX. SUMMARY	182
APPENDICES	184
BIBLIOGRAPHY	196

LIST OF TABLES

Table		Page
1	Preference Data for a Hypothetical Subject	7
2	Data for Artificial Example.	7
3	Comparison of U and V Scales	40
4	Accuracy of Numerical Integration for $p_{jk} = .5$	61
5	Accuracy of Numerical Integration for Asymptotic Values of p_{jk}	61
6	Values of p_{jk} for Different Values of C_1	62
7	Accuracy of the Unilateral Equations	64
8	p_{jk} Matrix Generated from Hypothetical J Scale.	66
9	Hypothetical Data Generated from V Theory.	82
10	Similarity of U and V Theories	82
11	Similarity of U and Q Theories as Indicated by Reproduction Method.	85
12	Similarity of U and Q Theories as Indicated by Prediction Method.	87
13	Similarity of V and Q Theories as Indicated by Prediction Method.	89
14	The Similarity of the U and V Theories with Respect to Data Generated from the Q Theory	90
15	Designations of the Stimuli	100
16	Obtained p_{jk} Matrix	125
17	Reproduced p_{jk}^* Matrix from the ICJ Scale (U1)	126
18	Predicted p_{jk}^* Matrix from ICJ Scale (U2).	128
19	Variance Components Analysis for Paired Comparisons	129

LIST OF TABLES - Continued

Table	Page
20 Variance-Components Analysis for U1	130
21 Variance Components Analysis for U2	130
22 Reproduced p_{jk}^* Matrix (R Stimuli Only) from LCJ Scale (U3)	133
23 Variance Components Analysis for U3	133
24 Predicted p_{jk}^* Matrix (R Stimuli Only) from LCJ Scale (U4)	135
25 Variance Components Analysis for U4	135
26 Reproduced p_{jk}^* Matrix from V Scale (V1)	137
27 Predicted p_{jk}^* Matrix from V Scale (V2)	138
28 Reproduced p_{jk}^* Matrix (R Stimuli Only) from V Scale (V3).	139
29 Predicted p_{jk}^* Matrix (R Stimuli Only) from V Scale (V4) .	139
30 Residual Sum of Squares for U and V Analyses.	140
31 I Scales for R Stimuli.	143
32 p_{jk} Matrix (R Stimuli Only) from Unilateral Judgments (Qa)	144
33 r_{jk} Matrix (R Stimuli Only) from Unilateral Judgments (Qa)	144
34 Separate Estimates of Each Adjacent Interval on the Qa Scale	146
35 p_{jk}^* Matrix from the Qb Analysis.	146
36 p_{jk} Matrix (R Stimuli Only) from Unilateral Judgments (Qc)	152
37 p_{jk}^* Matrix (R Stimuli Only) from the Qc Analysis	152
38 I Scales for the Combined Stimuli	156
39 The Relationship Between First and Second Choices as Indicated in the I scales	157
40 The Relationship Between the R and G Stimuli as Indicated in the I Scales	159
41 p_{jk} Matrix from Unilateral Judgments (J1).	159

LIST OF TABLES - Continued

Table	Page
42 p_{jk} Matrix from Unilateral Judgments (J2)	163
43 p_{jk} Matrix from Unilateral Judgments (J3)	163
44 Predicted p_{jk} Values from Normal Curves (1) and (2) in the Qd Analysis	167
45 p_{jk}^* Matrix from the Combined Predictions of Curves (1) and (2) in the Qd Analysis	168
46 Comparison of Residual Sum of Squares for U, V, and Q Analyses.	168

LIST OF FIGURES

Figure		Page
1	The relation between T_j and p_{jk} for three values of $Q_j - Q_k$	63
2	Hypothetical J scale	66
3	Non-SST property for non-overlapping distributions	70
4	Example of two types of observers.	95
5	The display board.	104
6	Successive positions of the panel with respect to display board.	105
7	The number of inconsistent stimulus pairs for all subjects	121
8	The number of inconsistent stimulus pairs for all unilateral and bilateral subjects.	122
9	Histogram of subjects on the Qb scale.	148
10	Histogram of subjects on the Qc scale.	154
11	Histogram of subjects on the Qd (J3) scale	165
12	Comparison of J scales with Munsell scale	169
13	Comparison of J scales with Munsell scale (R stimuli). . .	169

LIST OF APPENDICES

Appendix	Page
1 Numerical Integration.	184
2 Evaluation of Equation A	187
3 Solution to the Scale Values in the V Theory using the Prediction Method	191
4 Glossary of Symbols	192

CHAPTER I

INTRODUCTION

Choice, defined in general terms as a differential or selective response to the environment, may be regarded as a distinguishing characteristic of higher organisms' behavior. Speaking somewhat loosely for the present, two components of choice may be distinguished: a discrimination component and a preference component. Choice can be determined largely by either component. Discriminative choices occur, for example, when a subject judges lengths of lines in a psychophysical experiment; while a preferential choice occurs in buying commodity x rather than y.

Preference and discrimination are, of course, intimately related. One cannot prefer x to y unless one can discriminate them. In practice the converse is frequently true although it is not as obvious. An animal psychologist investigates an animal's discriminatory capacity by using a positive or rewarded stimulus and a negative or unrewarded stimulus. The experimenter uses, in effect, the animal's preferences to determine what discriminations the animal can make. Preferences are also utilized in a psychophysical experiment with human subjects. The observer in a psychophysical experiment presumably will not attempt to discriminate unless some reinforcement, either psychological or monetary, or both, is attached to "being right".

This interrelationship of preferences and discriminations makes the problem of measurement difficult. The usual procedure has

been to make some simplifying assumptions which would permit treating these two processes independently. Attitude measurements, for example, are frequently made by a two step process. The attitude items are first scaled by a discrimination procedure requiring the subject to act as an impartial judge and indicate, for each pair of attitude statements, which statement is more "pro" (segregation, for example). The subject is next asked to indicate the item he endorses, that is, prefers. If one subject endorses item A while another subject endorses item B, it is then assumed that, in some sense, the amount of disagreement between the two subjects can be measured by the number of discrimination units between A and B.

The reverse procedure is often used in a modified form in animal experiments. Instead of measuring discrimination capacity first, and preferences second, the experimenter first determines the animal's preference for, say, food to water when hungry and then uses this fact to determine his discriminations. Such experiments do not permit testing whether the discriminations an animal makes to receive one type of reward will agree with his discriminations under different reward conditions. If, for example, an animal exhibits certain color discriminations to obtain food, it need not exhibit these same color discriminations to obtain water.

The assumption that the discriminatory capacity of an animal is independent of the nature of the reinforcements will be termed the independence assumption. In general terms, the independence assumption

refers to the assumption that preferences and discriminations may be treated independently. (In a later section the independence assumption will be identified more concretely with the stability of a discrimination scale under different choices.)

This independence assumption may be tenuous under certain conditions. The choices one might make in selecting a color for a car could be based on color discriminations which differed from the discriminations made in a color discrimination experiment. It is possible for two colors only a few jnd's apart, as indicated by a psychophysical measurement, to be seen as quite dissimilar in terms of their desirability as a color for a car. Guilford (1940), working with the color preferences of human subjects suggests the following:

It may be that affective reactions are more discriminating with regard to the spectral composition of chromatic stimuli than are cognitive reactions, that wave lengths as such have a more direct and unique influence upon our feelings than upon awareness of color as such (1940, p. 457).

If "affective reactions" can be interpreted as preferential choices, then Guilford's remarks may also be viewed as questioning the validity of the independence assumption.

One alternative approach which avoids the independence assumption is to assume something about the nature of the relationship between discriminations and preferences. If some relationship is postulated,

that is, if one assumes a model, it may be possible analytically to separate these two processes while varying both experimentally. One could assume, for example, that if an individual most preferred A, then if he were presented with two other items, he would prefer to endorse the item closer to A. This seemingly innocuous assumption permits, as will be seen, the analytical separation between the discriminative component and the preferential component of an individual's choices. This model is the basis of the unfolding technique developed by Coombs (1950).

Quadrant I model

Although a precise axiomatic treatment of the unfolding technique has been given (Coombs, 1952) the underlying model (called by Coombs a Quadrant I model and denoted here as Q model) will be described in general terms. This model postulates a space consisting of points representing stimuli or items (denoted hereafter by stimulus points or q points) and points representing individuals (denoted by ideal points or c points). Coombs refers to this space as a joint or J space to denote the fact that it contains both classes of points. A one dimensional J space is called a J scale.

It is further stated that an individual with an ideal point at c_i will prefer stimulus $\underline{j}$, (S_j), to stimulus $\underline{k}$, (S_k), if, and only if q_j is nearer to c_i in the J space than q_k . More formally, $j \succ k$ (read stimulus j is preferred to stimulus k) if and only if

$$|c_i - q_j| < |c_i - q_k|$$

which is further abbreviated as

$$s_{ij} < s_{ik} ,$$

If certain assumptions are made concerning the stability and discriminability of the c_i and q_j for all i and j ---these assumptions are discussed more fully below---the unfolding technique provides a calculus for determining both the stimulus points and the ideal points.

Before describing this calculus it would be well to define more clearly the two interrelated concepts: preference and discrimination. This can be done most simply in terms of the parameters of the Q model.

Preference and discrimination

A scale of the q values of a set of stimuli will be called a discrimination scale or a stimulus attribute scale. The difference between q_j and q_k will indicate the discriminability between S_j and S_k in terms of some attribute, such as size or color. A preference scale for an individual (denoted by Coombs as an I scale) will consist of the s values of the set of stimuli. The difference between s_{ij} and s_{ik} will indicate how different S_j and S_k are to individual i in terms of their preferability. It should be noted that this scale consists of positive values only, the larger values denoting the less desirable stimuli.

It remains to distinguish a preferential judgment from a discriminative judgment. **Preferential judgments are judgments that satisfy the Q model.** The unfolding technique then is an analytical procedure for analyzing preferential judgments and separating out the preference scale and the stimulus attribute or discrimination scale.

A discriminative judgment can be defined as one which gives rise only to a discrimination scale. The observer in a psychophysical experiment is asked essentially to compare the q values of the stimuli only and to ignore his preferences; that is, his ideal point is irrelevant. A discriminative judgment, then, theoretically lacks the preferential component. It remains, however, a moot point whether a discrimination scale of a given set of stimuli obtained through preference judgments gives the same q values as one obtained through discrimination judgments. The independence assumption referred to earlier, in essence, asserts that the q values will agree.

The unfolding technique

The general character of the unfolding analysis will be illustrated with an artificial example consisting of five stimuli A, . . . , E, and 100 subjects. The stimuli are presented pairwise to the subjects---there will be 10 pairs---and for each pair, the subjects indicate their preferences. A typical subject's response may be tabulated as shown in Table 1.

The sum of a row indicates the number of times the stimulus was

TABLE 1

PREFERENCE DATA FOR A HYPOTHETICAL SUBJECT

	A	B	C	D	E	Σ
A		0	0	0	1	1
B	1		1	1	1	4
C	1	0		1	1	3
D	1	0	0		1	2
E	0	0	0	0		0

Note: "1" in each cell indicates that the row stimulus was preferred to the column stimulus when the two stimuli were paired.

TABLE 2

DATA FOR ARTIFICIAL EXAMPLE

I Scales	Order of Midpoints
1. A B C D E	AB
2. B A C D E	AC
3. B C A D E	AD
4. B C D A E	BC
5. C B D A E	BD
6. C D B A E	CD
7. D C B A E	AE
8. D C B E A	BE
9. D C E B A	CE
10. D E C B A	DE
11. E D C B A	

preferred. The order of these row sums gives the order of the s_{ij} values for the hypothetical individual, that is, the order of the stimuli on his I scale. For the subject tabulated in Table 1 the I scale (the order of the stimuli from smallest s_{ij} value to largest) is

B C D A E .

By tabulating each subject's responses as indicated in Table 1, it is possible to obtain the I scales for each of the 100 subjects.

Assume now that, of the 100 I scales, only eleven are distinct; that is, have a different order of stimuli. These eleven I scales are presented in Table 2.

The following step-wise procedure is then used to determine whether the conditions for a J scale are met by these data.

- 1) The number of distinct I scales should not exceed $\binom{n}{2} + 1$ where n equals the number of stimuli.
- 2) The least preferred stimulus in each I scale must be one of two stimuli.
- 3) It must be possible to order the I scales, $I_1, I_2, \dots, I_k$, such that the following condition is met for each stimulus: if R_i denotes the rank of a given stimulus in the i-th I scale (higher numbers denoting less preferred stimuli), then for this stimulus

$$R_{i+1} > R_i \Rightarrow R_{h+1} \geq R_h$$

for all $h \geq i$.

An examination of Table 2 in which the I scales have been correctly ordered indicates that the three conditions are satisfied by the data.

Since the present data satisfy these conditions it is possible to proceed with the determination of the stimulus attribute scale. The order of the q values is given by I scale (1) or I scale (11), the two I scales which are mirror images of each other. The direction of the stimulus attribute scale is not important here so that either A B C D E or E D C B A may be used. The metric information is obtained by determining the order of the midpoints between pairs of stimuli on the J scale. The order is given in the right hand column of Table 2 and can be obtained by observing which pair of stimuli have reversed orders in adjacent I scales. The order of the midpoints AD and BC imply, for example, that $|q_A - q_B| > |q_C - q_D|$. From the remaining midpoints the following metric relations may be concluded;

$$|q_D - q_E| > |q_A - q_B|$$

and

$$|q_D - q_E| > |q_A - q_C|$$

Limitations of the unfolding technique

Unfortunately, as may be seen in the preceding example, the unfolding technique has severe limitations. The limitations stem from the highly restrictive set of assumptions which the data must meet---assumptions rarely met in practice---and the relatively low

level of measurement obtained from the analysis. The J scale values are measured at an ordered metric level. This means that the c and q values can be ordered or ranked on the J scale and that certain intervals between the q 's and c 's can also be ordered.

The assumptions which must be made in order for the unfolding technique to provide a solution consist of at least the following:

- 1) The ideal points of each person are fixed---at least for the duration of the experiment.
- 2) The stimuli are completely discriminable; they are never confused with one another. **This assumption is implied in 3).**
- 3) The distances between the stimuli and the ideal points are also completely discriminated. If $s_{ik} = s_{ij} + \epsilon$ where $\epsilon > 0$ and is small, individual i nevertheless consistently prefers S_j to S_k .
- 4) No other random elements which would tend to produce "errors" in judgment are present.

Since these assumptions are rarely met, **and since the unfolding technique is a highly vulnerable method--it is extremely sensitive to error--it** therefore has a high likelihood of not yielding a solution.

In recent work Coombs (1958) has attempted to relax these assumptions. Instead of requiring s_{ij} to be consistently discriminated from s_{ik} he has only required that the subject prefer S_j to S_k more than half the time--assuming still that $s_{ik} = s_{ij} + \epsilon$ or $s_{ij} < s_{ik}$. The unfolding analysis is then carried out in the usual manner using only the ordinal information (that s_{ij} is

less than s_{ik}) and ignoring further the actual proportion of times that S_j was preferred to S_k . One might be lead to suspect, however, that the size of this proportion could give some information of the magnitude of the difference between s_{ij} and s_{ik} . It would seem to follow then that what is needed is a model, a probabilistic model, which relates the preference proportions to the magnitude of the parameters s_{ij} or c_i and q_j . The development of such a model and its testing is the basic problem of this study.

In Chapter III the equations of this probabilistic model are developed, their character described and a solution indicated. Here the general nature of this probabilistic extension of the model will be described.

A probabilistic Q model

The probabilistic extension proposed here assumes, for the single dimensional case, that the point corresponding to a stimulus can have different positions at different moments. However, the distribution of these successive positions will be normal. The discrimination of stimuli is assumed, therefore, to be imperfect, and the degree of overlap of two stimulus distributions is related to their degree of confusion. Similarly, it is also assumed that the ideal point of an individual varies over time, and that the distribution of its successive positions is normal. This assumption implies that a person is not absolutely or consistently certain which item he most prefers, even if the stimulus distributions ~~has~~ zero variance. He vacillates, first wishing to endorse this stimulus, then that.

The probability that a person on any given moment will prefer S_j to S_k will be determined by the probability that c_i (now a random variable) will be nearer to q_j than to q_k (both random variables), i.e.

$$\text{Prob} (j > k) = \text{Prob} (|c_i - q_j| < |c_i - q_k|)$$

or

$$P_{jk} = \text{Prob} (s_{ij} < s_{ik})$$

This extension of the Q model appears natural in the light of the usual psychophysical assumptions. The equations which result, however, are distinctly unnatural, and therein lies one source of difficulty.

Before embarking on the purely mathematical aspects of the model, let us summarize what may be expected from this model. **Firstly, the** probabilistic model places less **restrictive** demands on data. Whereas the unfolding technique required assumptions which were rarely met **and was a** highly vulnerable model, the present model requires, in principle, only a set of preference proportions and, if necessary, a **statistical** procedure for obtaining a "best fit" to the proportions. In general, a "best fit" will always exist so that the model is invulnerable, that is to say, always capable of **yielding** a solution. Secondly, with this model a stimulus discrimination scale can be obtained at the interval scale level, rather than at the ordered metric level. This means that a rigorous comparison can be made between the stimulus discrimination scale obtained under preferential judgments and one obtained under the discriminative judgments used in psycho-

physics. It should be possible then to test the independence assumption. This assumption will be tested in the experimental part of this study. There is another aspect to this experiment, but this requires some preliminary discussion.

Theories of choice and strong stochastic transitivity

For the purpose of illustrating another characteristic of the proposed probabilistic model of choice, let us take the liberty of imagining an argument submitted by an econometrician, one strongly devoted to one of the now familiar theories of choice. He asserts as follows:

"While I can appreciate your attempt at a more parsimonious and picturesque theory of choice---and this seems to be at best what you have accomplished by treating preference and discrimination simultaneously---I am not altogether convinced that I should therefore abandon my own probabilistic theory. After all, I am interested in the preferability (in my language, the utility) of commodities, not their discriminability along certain dimensions of, let us say, size and color. I believe that my particular theory permits me to account for the relationship between the preference proportions---though not necessarily the discriminations---involving a set of commodities. To put the matter differently, my theory accounts for the observed proportions of preferences. If your theory will also describe these data, then the issue between us is merely that

of convenience and certainly you have yet to demonstrate the superior convenience of your theory."

As the reader will no doubt suspect, the conclusion drawn by the, shall we say, straw man, is entirely erroneous. The issue is not simply one of convenience. The two theories are incompatible. A set of **proportions** generated by the present probabilistic theory cannot, in general---there are certain trivial cases which must be excluded---be generated by **the existing theories**.¹ The issue then is: which theory is correct, or at least more correct? It is this issue that will be examined experimentally in this study.

Although the theories of choice are compared in Chapter IV and their basic incompatibilities demonstrated there, the basis for the above conclusions will be indicated here. The fact that the present theory is not an equivalent formulation of one of the classical theories can be made plausible with the introduction of the concept of strong stochastic transitivity (SST). The pairwise preference probabilities between stimuli S_j , S_k and S_l are said to satisfy SST if the following constraint on the three probabilities is satisfied:²

1. This statement is too strong. Any theory can be made to fit a given p_{jk} matrix if a sufficiently large number of parameters are estimated from the data (so called free parameters). The Thurstone Case 3 Law of Comparative Judgment is a case in point. The theories of choice that are referred to here, and discussed more fully in Chapter II use approximately the same number of free parameters and these theories of choice may be said to be incompatible with the probabilistic Q theory.

2. An equivalent formulation, stated by Davidson and Marschak (1958, p. 13), is as follows. If

$$P_{jk} \geq \frac{1}{2} \quad \text{then} \quad P_{jl} \geq P_{kl}.$$

If $p_{jk} \geq \frac{1}{2}$ and $p_{kl} \geq \frac{1}{2}$,

then $p_{jl} \geq p_{jk}$ and $p_{jl} \geq p_{kl}$.

Present day theories of choice require the entire matrix of proportions to satisfy SST. The probabilistic Q model developed here does not require SST and, in fact some non-SST probabilities are usually expected. A matrix of probabilities therefore, which does not satisfy SST (aside from sampling errors) cannot be generated by the classical theories. The present theory and the current theories are therefore incompatible.

Prospectus

The experimental and theoretical issue of SST is considered in Chapter II. Chapter III consists of the mathematical development of a probabilistic Q theory, and in Chapter IV this theory is compared theoretically with two other probabilistic theories of choice. Chapter V introduces the experimental part of this study which focuses on two questions:

- 1) Does the probabilistic Q theory fit preferential data better than other probabilistic theories of choice?
- 2) Is the discrimination scale obtained from preferential judgments equal to the scale obtained from discriminative judgments?

In Chapters VI, VII, and VIII the procedure, results, and conclusions of the experiment are discussed.

CHAPTER II

PROBABILISTIC THEORIES OF CHOICE AND STRONG STOCHASTIC TRANSITIVITY

Estes (1958) has suggested that there are three major approaches to a theory of choice: the axiomatic approach of the economists, the scaling theory approach of the psychometricians or statisticians, and finally the behavior theory approach of experimental psychologists. This chapter reviews and compares these major approaches, showing the similarities between them.

As suggested in the previous chapter, the transitivity properties will be important in comparing these theories to the proposed **extension of the Quadrant I theory**. Therefore the transitivity properties implied by each theory will be indicated and, in the concluding section of this chapter, the experimental evidence for strong stochastic transitivity will be discussed.

The economic approach

Early history of the utility concept . - The concept of utility has been the central term in the theories of choice developed by economists. Edwards (1954) has succinctly reviewed the history of the utility concept up to 1953. In 1953, however, the utility concept was drastically altered---drastic, at least, for economists. It became probabilistic. To provide the historical context for this recent conceptualization of utility, the history of the utility concept is outlined briefly in this section; the post 1953 developments will be discussed in greater detail in the section following.

The utility concept, in the writings of economists (and philosophers) has passed through what may be distinguished as four stages. Stage 4 begins in 1953 with the probabilistic treatment of utility.

The issue in each stage is: can cardinal utilities be obtained? A cardinal utility is a measurement of utility unique up to a positive linear transformation; it is utility measured on an interval scale. If it is asserted that utility can be measured on an interval scale then the next question to be answered is: how?

Stage 1 can be associated with the names of Jevons and Marshall about the turn of this century. In this stage it was assumed that utilities could be attached to objects such that one invariably chose the object with the larger utility. It was also assumed without question that, in principle at least, one could measure utility on an interval scale and that, in fact, it was meaningful to discuss the derivative of utility (called marginal utility).

The definitive writing of Stage 2 was the paper by Hicks and Allen in 1934. In this stage the assumption of cardinal utility was abandoned. It was shown that the utility scale was strikingly non-unique, that any monotonic transformation of the utility scale was permissible. Utility, therefore, was only measured on an ordinal scale; it was not meaningful then to use the concept of cardinal utility.

Stage 3 begins with the Von Neumann and Morgenstern work of 1944, although it had its precursors in the works of Ramsey (1931) and perhaps D. Bernoulli (1738). The concept of cardinal utility was reasserted in this stage by considering risky choices, a risky choice

being one in which there is some uncertainty as to the outcome of one's acts or choices. By assuming that individuals' choices maximize the expected utility of alternative actions, they were able to show, with certain additional assumptions, that one could measure utility up to a linear transformation. Although it was now meaningful to use the term cardinal utility in the context of a risky choice, it was not permissible to apply the concept to exclusively riskless choice, choices involving no uncertainty. It should be noted that probabilities do appear in the Von Neumann-Morgenstern formulations, but these probabilities pertain to the likelihood that certain outcomes will follow from certain acts by an individual. The utility of the outcome, however, does not have a probability attached to it.

Stage 4 appears to have begun about 1953 with the writing of Vail (1953). In this stage the probabilistic notion of utility was introduced and, as a consequence, riskless choices were once again considered. Debreu (1958) has shown that the introduction of the probabilistic version of utility will permit the measurement of cardinal utility for riskless choices. It is clear that what was lacking in the early formulation of utility were sufficient constraints to obtain a relatively unique solution or measurement of utility. The Von Neumann-Morgenstern approach introduced these constraints by the use of chance events in the form of probability mixtures. However, the constraints may be more directly introduced by permitting probabilistic concepts of utility. Indeed, it is interesting to observe that the economists, in their attempt to simplify the problem of utility measurement, made the problem overly difficult. Psychologists

have known since Fechner of the usefulness of inconsistency for establishing a unit of subjective measurement.

Probabilistic formulations of utility. - Before describing the probabilistic development of utility since 1954, the probabilistic discussions which appear in the 1936 writings of the economist Georgescu-Roegen (1936) should be noted. Georgescu-Roegen assumed that the utility of two commodities can not be perfectly distinguished or discriminated if their difference is less than some fixed number, Δ . If the difference is greater than Δ then the probability of choosing the commodity with the larger utility is equal to one. He shows, with some additional assumptions, that while preferences will be transitive the indifference relation will be intransitive.

The initial discussions of Vail (1953) and Davidson (1956) of probabilistic utilities were cautious. The concept of utility was formulated in general terms, and the problem posed by the writer in each case was that of selecting the precise constraints to impose on the probabilities associated with the utilities. Vail assumes merely that the utilities $u_1, u_2, \dots, u_n$ of a set of commodities $S_1, S_2, \dots, S_n$ are distributed according to the density function

$$(1) \quad \phi(u_1, u_2, \dots, u_n)$$

so that the probability that S_i will be preferred to S_j is equal to the probability that u_i will be greater than u_j , that is,

$$(2) \quad \text{Prob}(S_i \succ S_j) = \int_{u_i > u_j} \phi du.$$

With this general definition of utility, Vail asks: what restrictions must be placed on the function ϕ such that the probabilities of choice will satisfy weak stochastic transitivity (WST)? Weak stochastic transitivity, a term evidently first used by Vail, is defined as follows

$$\text{If } p_{AB} \geq \frac{1}{2}, \text{ and } p_{BC} \geq \frac{1}{2} \text{ then } p_{AC} \geq \frac{1}{2}.$$

Vail asserts that WST will be satisfied if the function ϕ is such as to satisfy the inequality

$$(3) \quad \begin{aligned} &\text{Prob}(u_1 > u_2 > u_3) + \text{Prob}(u_1 > u_3 > u_2) + \text{Prob}(u_2 > u_1 > u_3) > \\ &\text{Prob}(u_2 > u_3 > u_1) + \text{Prob}(u_3 > u_1 > u_2) + \text{Prob}(u_3 > u_2 > u_1) . \end{aligned}$$

The second and third term on the left hand side of the inequality refer to the probability of one inversion occurring in the order 1, 2, 3. The first and second terms on the right hand side give the probability of two inversions and the last term the probability of three inversions.

Further, continues Vail, if it is assumed that the utilities are independently distributed, that is,

$$(4) \quad (u_1, u_2, \dots, u_n) = (u_1) (u_2) \dots (u_n),$$

then "intuition demands" that the probabilities satisfy SST. The proof is not given.

Davidson (1956) suggested two probabilistic modifications of of utility that are directly related to the WST and SST conditions

of Vail. One theory, called by Davidson a weak stochastic theory, postulates a utility function having the property that

$$(5) \quad \text{If } p_{ab} \geq \frac{1}{2} \text{ then } u_a \geq u_b.$$

It is clear that this utility function will satisfy the WST condition. The strong stochastic theory, the second one proposed by Davidson, requires the utility function to have the following property:

$$(6) \quad p_{ab} \geq p_{cd} \iff u_a - u_b \geq u_c - u_d.$$

This version of probabilistic utilities, which will hereafter be referred to as theory or model U, will appear frequently in one form or another in the remaining discussion. As one would expect, this strong stochastic utility theory will satisfy the SST condition and, in addition it will satisfy a somewhat stronger condition, called by Davidson and Marschak (1958), the quadruple condition. The quadruple condition states that

$$(7) \quad p_{hi} \geq p_{jk} \implies p_{hj} \geq p_{ik}.$$

This condition, according to Davidson and Marschak (1958), is stronger than the SST condition since it implies the latter, but not conversely.

One of the purposes in introducing the probabilistic concept of utility was, it will be recalled, to reintroduce the notion of cardinal utility. If the utility theory is stated in the form of theory U, the problem is reduced to determining the conditions on the probabilities which lead to a utility function that is unique up to a linear

transformation. Davidson and Marschak (1958) and Debreu (1958) have shown that essentially the quadruple condition on the probabilities is sufficient. More precisely, it is also necessary to assume that the set of alternatives or commodities is stochastically continuous. Debreu defines this condition as follows

$$(8) \quad \begin{array}{l} \text{If } p_{ba} \leq q \leq p_{ca} \text{ then there exists an alter-} \\ \text{native } d \text{ in the set of alternatives such that} \\ p_{da} = q. \end{array}$$

This condition seems fairly mild and need not concern us. Essentially then, it can be said that the utility function in theory U implies the quadruple condition (and therefore SST) and the quadruple condition specifies the utility function in U up to an increasing linear transformation. It will be noted that SST however, is not sufficient to specify a cardinal utility function.

Recently Luce (Luce and Raiffa, 1957) has provided an alternative although more tortuous route for arriving at theory U. He assumes essentially a weak stochastic definition of the utility function, adds eleven axioms and then succeeds in proving that the probability of choice between two alternatives is a function of the difference between the utilities of the two alternatives. Furthermore, he shows that the utility is defined up to a positive linear transformation. This is precisely the property of theory U, but as has been discussed, theory U can be derived from two axioms.

Under certain conditions, Luce asserts that the utility can be determined directly from the probabilities of choice as follows

$$(9) \quad u(a) = \left[\frac{p_{ab^*} - p_{b^*a}}{p_{a^*b^*} - p_{b^*a^*}} \right]^{\frac{1}{\epsilon}} \quad (a \succ b \text{ in the SST sense})$$

where $\epsilon = a$ constant and a^* and b^* are any pair of alternatives such that $p_{a^*b^*} > \frac{1}{2}$.

Theory V. - Another probabilistic model has recently come into prominence as a result of Luce's more recent (1959) axiomatic writings. Luce postulates a single axiom which relates the choice probabilities involving sets of alternatives of different sizes. From this axiom he derives a theory of choice between two alternatives.

The axiom states:

$$(10) \quad \begin{aligned} &\text{If } p_{xy} \neq 0 \text{ or } 1 \text{ for all } x \text{ and } y \text{ in the} \\ &\text{set } T, \text{ then for } R \subset S \subset T \\ &p_T(R) = p_S(R) p_T(S), \end{aligned}$$

$p_T(R)$ is defined as the probability of choosing an element from the set R when the set T is available or is presented. This axiom also considers probabilities which are equal to zero or one but this will not be relevant here. If (10) is satisfied then a function v can be specified which is unique up to multiplication by a positive constant, such that

$$(11) \quad p_{ij} = \frac{v_i}{v_i + v_j}$$

A theory of this form will be denoted a V theory. It will be noted that whereas in the U theory the choice between the two

alternatives depends on the difference between the utilities of the two alternatives, in the V theory it depends on the ratio.

Luce demonstrates further that the V theory will satisfy the SST condition and a somewhat stronger condition on the probabilities, namely,

$$(12) \quad p_{ij} = \frac{p_{kj} p_{ik}}{p_{ki} p_{jk} + p_{kj} p_{ik}}$$

or equivalently

$$(13) \quad \left(\frac{p_{ij}}{p_{ji}} \right) \left(\frac{p_{jk}}{p_{kj}} \right) = \frac{p_{ik}}{p_{ki}}$$

Since this condition holds for any three alternatives S_i , S_j , and S_k , using the terminology of Davidson and Marschak, this condition will be called the triple condition. In fact, the triple condition bears the same relationship to the V theory as the quadruple condition does to the U theory. It can be shown that the triple condition is sufficient to guarantee the existence of the v function and that furthermore the function is uniquely specified up to a positive multiplicative constant. In other words, if choices are confined to pairs of alternatives, it is only necessary to assume the triple condition instead of the axiom postulated by Luce. As has already been noted, both the triple condition and the quadruple condition imply the SST condition.

Psychometric approach

Scaling theorists and statisticians have not been directly concerned with formulating theories of choice. Their problem has been

to interpret or concisely describe actual data. However, in the process of solving this problem, a model is generally assumed (implicitly or explicitly) and this model, it will be seen, may be termed a theory of choice.

Thurstone's Law of Comparative Judgment (LCJ). - According to Thurstone (1927), a random variable, u_j , which is normally distributed over time, is associated with each stimulus (or alternative or commodity). For S_j the distribution of u_j is then

$$(14) \quad N(u_j, \sigma_u).$$

The probability of choosing S_j in preference to S_k , when both are available, is equal to the probability of u_j being greater than u_k . If it is further assumed that the variance of u_j is equal for all S_j and the correlation between u_j and u_k is constant (or zero for simplicity) for all S_j and S_k , ---these assumptions are termed the Case 5 assumptions and are the ones usually employed---then the probability of choice may be expressed as

$$(15) \quad p_{jk} = \int_0^{\infty} N\left[u_j - u_k, \sqrt{\sigma_j^2 + \sigma_k^2}\right] dx.$$

By a substitution of variable this may be written as

$$(16) \quad p_{jk} = \int_{-\infty}^{\frac{(u_j - u_k) \sqrt{\sigma_j^2 + \sigma_k^2}}{\sqrt{\sigma_j^2 + \sigma_k^2}}} N(0,1) dx$$

or more simply as

$$(17) \quad u_j - u_k = x_{jk} \sqrt{\sigma_j^2 + \sigma_k^2}$$

where x_{jk} is the normal deviate corresponding to p_{jk} . It is customary

to set the sum of the variances of u_j and u_k equal to one for all S_j and S_k (this defines the unit of measurement) so that

$$(18) \quad U_j - U_k = X_{jk}.$$

Equation (18) is termed the Case 5 Law of Comparative Judgment.

Equation (18) asserts, in effect, that the probabilities of choice are determined entirely by the difference in scale values between the two alternatives. Furthermore, it is clear that if

$$(19) \quad U_j - U_k \geq U_l - U_m$$

then

$$(20) \quad X_{jk} \geq X_{lm}$$

and this implies that

$$(21) \quad p_{jk} \geq p_{lm}$$

But this is precisely the property of theory U. The Thurstone **theory** therefore **is**, in essence, theory U. It can be immediately concluded, therefore, that the LCJ scale implies probabilities which satisfy SST and, of course, also implies the quadruple condition.¹

¹Case 3, which does not assume that the variance of u_i , is constant for all S_i will probably not imply the SST condition. This is because in Case 3 the probability of choice is not completely specified by the difference between two scale values. The variances or the ratio of the variance must be also specified.

Several difficulties arise in the application of the Law of Comparative Judgment. First, actual data do not fit the theory exactly, so that a fitting procedure must be used. The usual procedure is to average the X_{jk} values of each column in the X_{jk} matrix, the matrix of normal deviates formed from all the possible pairings of a given set of stimuli. (The diagonal entries are left blank.) This fitting procedure has been shown by Mosteller (1951) to satisfy a least squares criterion.

A second difficulty in practice relates to the interpretation of the random variable u_i . If only one subject is used this variable is interpreted as the various momentary values of sensation which are normally distributed over time. Generally, however, the theory is applied to a group of subjects. **Two interpretations are possible.** It may be assumed that the value of u is fixed for each subject and that u has a normal distribution over subjects. One may also assume, however, that each subject has a normal distribution (over time) and that the means of the normal distributions are also normally distributed. The LCU equations in both cases are formally identical. A difference does arise, however, in determining the goodness of fit of the theory. This is discussed more fully in Chapter 6.

Bradley-Terry Model. - In their initial paper (1952b) Bradley and Terry are concerned with the problem of testing the hypothesis that a set of treatments (or stimuli) presented pairwise are perceived as significantly different. In a typical experiment (1952b) pork roasts were presented in pairs to a competent panel of judges each

of whom **was** asked to indicate their preference. The roasts were obtained from three groups of hogs which had been fed on a different diet. The object of the experiment was to determine whether the diet had a noticeable effect on the flavor of the roast.

Using a maximum likelihood procedure, they assume that the probability of treatment $\underline{j}$ being preferred to treatment $\underline{i}$ is v_i/v_i+v_j so that the likelihood function can be written as

$$(22) \quad L = \prod_{ij} \binom{m_{ij}}{n_{ij}} \left(\frac{v_i}{v_i+v_j} \right)^{n_{ij}} \left(\frac{v_j}{v_i+v_j} \right)^{n_{ji}}$$

where n_{ij} is the number of times $\underline{i}$ was preferred to $\underline{j}$ and $m_{ij} = n_{ij} + n_{ji}$. The values of v_i which maximize L in general can only be obtained through an iteration procedure. Bradley and Terry do not give the iteration procedure, but Ford (1957), working independently, states the following iteration scheme:

$$(23) \quad \left. \begin{aligned} v_i^{(n+1)} &= \frac{n_i}{\sum_j \frac{m_{ij}}{v_i^{(n)} + v_j^{(n)}}} \\ v_j^{(n+1)} &= v_j^{(n)} \quad j \neq i \end{aligned} \right\} \begin{array}{l} i = 1, 2, \dots, n \\ \text{cyclically} \end{array}$$

where $n_i = \sum_j n_{ij}$.

Dykstra (1956) has suggested a procedure for obtaining the starting values of v_i in this iteration scheme,

$$(24) \quad v_1^{(0)} = \frac{\sum_j p_{1j}}{\left[(n-1)^2 - (n-2) \sum_j p_{1j} \right]}$$

and Bradley (1952b, 1954b) has published tables giving the values of v_i when the number of stimuli does not exceed five and the number of repetitions on a pair is less than 10. The test statistic indicating the probability of the v 's being equal is also tabled.

Since it is clear that the model underlying the Bradley-Terry analysis is the V model of Luce, the Bradley-Terry procedure may be viewed as providing a procedure for obtaining a best fit of the V theory to actual data. The goodness of fit of the model has also been discussed by Bradley (1954a) although for large m his statistic is substantially the same as the usual χ^2 test based on the squared difference between observed and theoretical frequencies.

Other pair comparison analyses. - Other methods for analyzing choices among pairs of stimuli are proposed from time to time, but in most cases the model underlying the analysis is not clearly stated or it does not seem to have direct implication for understanding the basis of an individual's choice. Guttman (1946), utilizing a discriminant function analysis, attempts to determine scale values by minimizing the sum of squares within an individual as compared with the sum of squares between individuals. The method is equivalent to maximizing correlation ratios and, as Guttman shows, will generally not yield the same solution as the Thurstone Law of Comparative Judgment solution. The relationship between the probability of choice for an

individual and certain parameters of the analysis do not seem to be indicated in Guttman's treatment so that a probabilistic theory of choice is not easily extracted from his method.

Kendall's method (1955) involves powering an (n_{ij}) or D matrix where n_{ij} gives the number of times i was chosen over j . (The diagonal entry is $\frac{1}{2}$.) The scale value of an alternative is obtained by summing the appropriate row. Kendall shows that if the matrix is squared and the row sums computed, the values computed for S_j , for example, will refer to the number of two step transitivity. An m step transitivity is defined as a preference relation between n stimuli having the form $S_1 > S_2 > S_3 \dots S_m$. A matrix that is raised to the n th power will give the number of n step transitivity relations for each stimulus. Kendall asserts that there will exist some finite value of n such that the rank of the row sums (scale values) will not change. Thompson (1958) has proposed a modification of the Kendall Method. He suggests the matrix D be powered and summed as follows:

$$(25) \quad D + rD + r^2 D + r^3 D + \dots,$$

where r is a positive number for which the sum converges. In fact, the series will converge if and only if

$$0 < r < 1/k$$

where k is the principal characteristic root of D . The two methods will agree if $r = 1/k$.

The choice theory implicit in the Kendall and Thompson methods would seem to imply that an individual attempts to maximize his

transitivities. However, transitivities are maximized when the choices are completely consistent, so that a probabilistic theory of **choice** is not implied by these methods.

Behavioral theory approach

Theories of behavior, containing implications for a theory of choice, abound in the psychological literature. Most theories, however, do not explicitly relate a choice probability to the parameters of the theory. The theories usually state only that the response with the larger performance vector (Tolman, 1955), or behaviorial potential (Rotter, 1954), or resultant motivation (Atkinson, 1957), or force (Lewin, 1951) will occur. A few theories are more explicit and these will be discussed below.

Model V. - Estes (1958) assumes that an individual confronted with two stimuli, orients successively towards **these** stimuli, first one then the other. When the individual is oriented toward one stimulus he is in a position **for** choosing it. If he does not choose it, he orients then to the other stimulus. It is assumed that a stimulus can be chosen only when one is oriented toward it. If p_i is the conditional probability of choosing S_i when the individual is orienting toward S_i , Estes shows that in the limit the probability of choosing S_i when the individual is initially oriented to S_i is:

$$(26) \quad \frac{p_i}{p_i + p_j - p_i p_j}$$

The probability of choosing S_i when initially oriented to S_j is:

$$(27) \quad \frac{p_i (1-p_j)}{p_i + p_j - p_i p_j}$$

If it is also assumed that initially the individual is equally likely to orient to S_i as to S_j then

$$(28) \quad p_{ij} = \frac{1}{2} \left[\frac{p_i}{p_i + p_j - p_i p_j} + \frac{p_j(1-p_i)}{p_i + p_j - p_i p_j} \right] = \frac{p_i(1-\frac{p_j}{2})}{p_i + p_j - p_i p_j} .$$

Furthermore since

$$(29) \quad \left(\frac{p_{ij}}{p_{ji}} \right) = \frac{p_i(1-\frac{p_j}{2})}{p_j(1-\frac{p_i}{2})}$$

we have,

$$(30) \quad \left(\frac{p_{ij}}{p_{ji}} \right) \left(\frac{p_{jk}}{p_{kj}} \right) = \left(\frac{p_{ik}}{p_{ki}} \right) ,$$

which is exactly the triple condition. Since the choice probabilities must satisfy the triple condition, it can be concluded that the V scale will exist, that its values will be determined up to a multiplicative constant, and that the values of v will precisely generate the choice probabilities p_{ij} according to the V theory, namely:

$$(31) \quad p_{ij} = v_i / (v_i + v_j) .$$

Thus starting from certain behavioral considerations Estes has shown that the choice probabilities will be constrained according to the V theory.

A learning theory developed by Thurstone (1930) and extended by Guilksen (1953) employs as a response strength a term which has the property of a V scale. Luce (1959) has also shown that the learning theories of Bush and Mosteller (1953) and Estes (1950) will, under certain conditions, have the properties of the V model.

Model U. - Cartwright and Festinger (1943), using the concepts of Lewin, propose a theory that assumes that choice between a pair of alternatives is the resultant of three forces: a force directed toward S_i , one directed toward S_j and one called the restraining force. Each force is assumed to be normally distributed with equal variance. The probability of preferring S_i to S_j is equal to the probability that the force toward S_i will exceed the sum of the other two forces. The use of restraining forces in this theory presents some difficulties in interpreting the theory. The concept of a restraining force produces a correlation between the two resultant forces, one toward S_i and one toward S_j , and prevents the two probabilities p_{ij} and p_{ji} from summing to one. If the restraining force term is neglected the theory requires

$$(32) \quad f_i - f_j > f_k - f_m \Leftrightarrow p_{ij} > p_{km}$$

where f_i denotes the mean value of the force toward S_i . This relationship is, of course, the basis of theory U.

The learning theory of Hull (1943) assumes that a response is determined by the momentary effective reaction potential, $\frac{\dot{E}}{S_{R_1}}$ which in turn is a function of $\overline{E_{R_1}}$, a constant for a given state of learning, and an oscillation function $S_{R_1}^0$, i.e.,

$$(33) \quad \frac{\dot{E}}{S_{R_1}} = \overline{E_{R_1}} - S_{R_1}^0.$$

The oscillation function is normally and independently distributed for different responses, with a fixed variance, σ^2 , and a mean 3σ less than $\overline{E_{R_1}}$. If for simplicity we let $u_i = \frac{\dot{E}}{S_{R_1}}$, and $U_i = \overline{E_{R_1}} - 3\sigma$, the density function of u_i is then $N(U_i, \sigma)$. A threshold concept is also introduced in the theory. A response can occur only if u_i exceeds some fixed number L . The probability that R_i will occur rather than R_j , i.e., that S_i will be chosen instead of S_j , is equal to the probability that u_i exceeds u_j and also that u_i exceeds the threshold value L . If the threshold term is ignored, the familiar LCJ equation occurs so that the U model is certainly implied in the learning theory, that is,

$$(34) \quad P_{ij} > P_{km} \Leftrightarrow U_i - U_j > U_k - U_m.$$

However (34) will not result if the threshold concept is employed so that in this case theory U will not be implied.

Theory U seems to be implicitly used in many learning theories. However, to be certain that theory U is the basis of a theory of

behavior, it should be evident in the behavior theory (1) that the probability of choice depends only on the difference between two values, U_i and U_j , where U_i is associated with S_i , and (2) that the larger the value of U_i in comparison to U_j the more probable the choice of S_i . If the u 's are random variables the probability must be shown to depend only on the mean values of distributions having equal variances. Many theories of learning postulate a difference between two variables as the determiner of the response, but it is not always clear whether the above conditions are also present.

Summary

On the basis of this limited review, the following generalizations would seem appropriate:

1) The U and V theories are perhaps basic to most theories of choice. Certainly in view of the estimation procedures and goodness of fit tests which have been devised for the U and V theories, these theories are the ones **most** thoroughly developed. In view of the axiomatic treatments, by economists and mathematicians, these models also appear to be the ones most thoroughly understood.

2) The U and V models have been applied both to what **have** been called preferential choices and to discriminative choices. The applicability of the models to either type of choice has not been questioned.

3) SST is implied by both models.

4) Although not implied in the above discussion, it should be mentioned that there have been few, if any, attempts to test the

quadruple condition or the triple condition directly. Scaling studies in the past, using a least squares criterion to fit a model to their data, frequently report merely the absolute deviations of predicted probabilities and obtained proportions (Guilford, 1936), or they report their correlation (Shepard, 1957), or in some cases, merely their graph with the comment "how well the theoretical curves fit the empirical data may be judged by the reader for himself" (Cartwright and Festinger, 1943, p. 615). In all cases the experimenter has neglected to consider the number of parameters he has had to fit to obtain the predictions, but, as the saying goes, a few parameters will fit an elephant.

More recent studies have used a χ^2 goodness of fit technique. The test however must be exceedingly weak, since the fit, as indicated by the χ^2 value, is frequently better than one would expect if the model were in fact correct. Apparently only one published study (Guilksen and Tukey, 1958) found a significant χ^2 (i.e. rejected the model) using the LCJ method. Bradley (1952b, 1954a) reports the results of a series of small experiments using the V model and finds only one χ^2 approaching the 5 per cent level of significance. This χ^2 value is attributed to the poor approximation resulting from low expected frequencies. (The χ^2 goodness of fit test will be discussed more fully in Chapter VI.)

An indirect and somewhat weaker test of the U and V model would involve testing the SST condition. These experiments are discussed next.

Experimental tests of SST

Papandreou et al (1957) attempted to test the SST condition by using 18 subjects in one experiment (E_1), 5 in another (E_2) and using as stimuli admission tickets to theatrical or sports events. In E_1 , 135 pairs were presented five times to each subject and in E_2 , 63 pairs eight times. Of the 805 triples in E_1 , 704 fit the SST condition and E_2 , 89 out of 102 satisfied SST. Papandreou concludes that "the evidence provided by the data (in support of SST) is ... overwhelming by any criterion " (1957, p. 7).

May (1954) had each of 65 students make three pair comparison judgments involving three different (but related) descriptions of women. Seventeen out of the 65 (27%) gave responses which were not transitive. Edwards (1954) has criticized the experiment on the grounds that the S's were not allowed an indifferent judgment, suggesting that such subjects would be expected to have intransitive judgments 25% of the time. May suggests that intransitives can occur when more than one dimension underlies the stimuli.

Edwards (1953) reports in an experiment using bets with equal expected value as stimuli that intransitive triads regularly occurred in about 20% of the triads.

Davidson and Marschak (1958) also employed bets as stimuli but with several modifications. The expected value of the two bets were not necessarily equal although the difference was at most $4\frac{1}{2}$ cents; the two events which comprise each bet had a probability of occurrence equal to $\frac{1}{2}$; and no pair of bets was replicated. There were 17

subjects, 76 triples, 26 of which were presented in the last experimental session. According to Davidson and Marschak, if certain rectangular assumptions are made, one would expect 19 of the 26 triples to be intransitive; if the rectangular distribution is confined to a space consistent with WST, the expected number of intransitive triples is 14.25; under SST the number is 10.45. The average number of intransitives for all subjects is 10.4 and only two subjects exceeded 14.25. It is therefore concluded that all but two of the subjects confirm the validity of the SST condition.

Davidson (1959) observed in another experiment, utilizing eleven gambles to form 55 pairs, each of which was presented 6 times in 6 sessions, that the number of intransitivities decreased uniformly from the first session to the last. However, in all sessions he found fewer than 41 intransitivities which, according to Landau (1951), is less than the number expected on the basis of random responses.

Davidson and Suppes (1957) use a linear programming model which requires determining a utility ϕ such that inequalities of the form:

$$(35) \quad \phi(x) + \phi(y) \geq \phi(u) + \phi(v)$$

are satisfied. If the preference judgments are not completely transitive, it is necessary to introduce an additional term θ , (called the threshold preference) whose value is to be minimized. The six stimuli, LP phonograph records, were presented in 15 non-replicated pairs. In most cases, the θ was non zero, indicating a lack of transitivity.

Coombs (1958) analyzed the judgments obtained from 4 subjects and 12 gray chips by first distinguishing between unilateral

and bilateral triples. Assuming a Q model, a unilateral triple is defined as one consisting of stimuli all located on the same side of the individual's ideal. A bilateral triple has one stimulus on one side of the ideal and the other two stimuli on the other side. Coombs suggests that the SST condition should not necessarily hold for bilateral triples, although it may be expected to hold for unilateral triples. (This is discussed more fully in Chapter III.) The stimuli, which varied in grayness, were presented in groups of four, each pair being replicated the equivalent of 66 times. The analysis gives χ^2 values at the .01 level indicating the expected relationship between the SST-nonSST dimension and bilateral-unilateral dimension. Luce (1959) has questioned whether SST has been tested in this experiment since it required estimating P_{ij} from triadic judgments by methods which do not always give correct estimates of P_{ij} .

Experimental comparison of U and V theories

The test of SST will not distinguish between the U and V theories since both are consistent with the SST condition. There are apparently only two experiments that empirically attempt to compare the two theories. The customary procedure in scaling theories, as has been mentioned, is to show that the fit is better than an alternative (non-random) theory.

Bliss et al (1956) performed separate U and V analyses of the preference judgments involving 4 stimuli (sprays used on apples) presented pairwise to 25 tasters (human subjects). The six pairs were replicated over 6 days. The U analysis, for example, consists

of obtaining u values for each day--- u therefore is assumed to be normally distributed over subjects---and the u value for a given stimulus is then averaged over days. The scale values in Table 3 were obtained by normalizing the averages so that the u value of S_2 equalled 1.

TABLE 3
COMPARISON OF U AND V SCALES

	S_1	S_2	S_3	S_4
U Theory	.108	1.000	-.409	-.699
V Theory	.039	1.000	-.399	-.640

They conclude that "the scale rankings on a preference scale were substantially the same" (1956, p. 401), although the goodness of fit **of neither model is** reported. (The scale values of the Kendall and the Rankit method of Bliss also gave the same ranking of the stimuli.)

Jackson and Fleckenstein (1957) used 12 colored slides of the same scene developed through different colored filters. Each subject's preferences were replicated twice for each pair. The experimenters conclude that "in terms of goodness of fit only the Bradley-Terry [v] model predicted the original data sufficiently well " (1957, p. 62). However no probability levels are reported so it is difficult to know what sufficiently well means, nor is it clear what test statistic they used.

The present study

The experimental issue of SST is approached somewhat indirectly in the present study. The difficulties with the previous experimental tests of SST are either their lack of a clear criterion defining what constitutes a significant number of intransitive triples, or their use of a criterion involving arbitrary assumptions (for example, the Davidson and Marschak assumption of a uniform distribution). This study will attempt to fit the parameters of the Q model as well as the U and V models. If the Q model reproduces or describes more of the data than the U and V models it will be argued that there is good evidence that some of the triples do not satisfy SST. The study will therefore determine which one of the three models (U , V , or Q) best accounts for the data, rather than test whether there are enough observations to reject a model at some given level of significance, **or to test what Mosteller (1958) has called** the "nullest of null hypotheses" (1958, p. 224).

CHAPTER III

DEVELOPMENT OF A PROBABILISTIC QUADRANT I MODEL

In this chapter, the equations for a probabilistic Q theory are developed and their properties explored.

Assumptions of the theory

The basic assumptions of the theory are as follows.

- 1) The stimulus point q_j and the ideal point c_i associated with stimulus S_j and individual i , respectively, are random variables. It is further assumed that the variables are independently distributed so that the correlation between q_j and q_k is zero, as well as the correlation between c_i and q_j .
- 2) Both variables, q_j and c_i , are distributed normally with equal and constant variances, for all i and j . The corresponding density functions are

$$(1) \quad N(q_j, \sigma) \quad \text{and} \quad N(c_i, \sigma) .$$

- 3) The probability of individual i preferring S_j to S_k is equal to the probability that his ideal point will be nearer to the stimulus point of S_j than to S_k ¹. That is,

¹ It is assumed here that when the individual judges S_j and S_k , c_i has the same value for both stimuli.

$$(2) \quad p_{jk} = \text{Prob} \left[|c_1 - q_j| < |c_1 - q_k| \right] .$$

Assumption 2) is perhaps the least tenable one. To assert that the ideal point of each person has the same variance is to assume that each person locates his ideal point or region with the same precision. If the variance of the ideal point can be associated inversely with the strength of preference, a not too implausible interpretation, then we are assuming in effect that the members of the population prefer their ideal stimulus with equal strength. If a homogeneous population is selected there will be some grounds for accepting this assumption as, at least, a first approximation. To introduce a separate variance term for each individual would drastically increase the number of parameters that would have to be estimated.

The assumption of equality between the stimulus variance and the ideal variance is more severe. This assumption may be assumed valid in the limit, however, when an individual's preferences have optimal precision. It can be expected that a person can do no better than to discriminate his ideal stimulus from other stimuli as well as he can discriminate other stimuli from themselves. However, he may do worse, if he has less certainty or conviction about his preferences. In psychological terms this assumption probably requires subjects in an experiment to be optimally motivated toward, or sensitive to, the stimuli to be judged.

Development of equations

Two distinct forms of the p_{jk} equation, called equation A and B, are developed in this section.

Equation A. - Since the joint density function of the variables $(c_i - q_j)$ and $(c_i - q_k)$ will be needed it will be necessary first to determine their correlation.

Let us define

$$t_{ij} = c_i - q_j$$

(3) and

$$t_{ik} = c_i - q_k$$

In the following discussion, i is constant since only one individual is involved; the i subscript of t is therefore generally suppressed. Under the conditions of the independence of the variables c_i , q_j and q_k , the correlation between t_{ij} and t_{ik} is as follows:

$$(4) \quad r_{t_j t_k} = \frac{\sigma_{c_i}^2}{\sqrt{\sigma_{c_i}^2 + \sigma_{q_j}^2} \sqrt{\sigma_{c_i}^2 + \sigma_{q_k}^2}}$$

If $\sigma_{q_j}^2 = \sigma_{q_k}^2 = \sigma_q^2$ then

$$(4') \quad r_{t_j t_k} = \frac{\sigma_{c_i}^2}{\sigma_{c_i}^2 + \sigma_q^2}$$

and finally, if $\sigma_{c_i}^2 = \sigma_q^2$ then

$$(4'') \quad r_{t_j t_k} = \frac{1}{2}$$

Equation (4'')¹ gives the required correlation term under the assumptions of the theory. However, in the discussion of group results it will be necessary to use (4').

The proof of equation (4) may be indicated as follows.

By definition

$$(5) \quad r_{t_j t_k} = \frac{E[(t_j - T_j)(t_k - T_k)]}{\sigma_{t_j} \sigma_{t_k}},$$

where T_j and T_k denote the mean values of t_j and t_k , respectively. Using primes to denote deviations from the mean,

(5) may be rewritten as

$$(6) \quad r_{t_j t_k} = \frac{E[(c'_j - q'_j)(c'_k - q'_k)]}{\sigma_{t_j} \sigma_{t_k}}.$$

The numerator term is equal to

$$(7) \quad E[(c'_j)^2] - E(q'_j c'_j) - E(c'_j q'_k) + E(q'_j q'_k),$$

but from the assumption of independence, the last three terms above are zero. Since

$$(8) \quad E[(c'_j)^2] = \sigma_{c_j}^2,$$

¹This result is due to Professor Coombs.

$$(9) \quad r_{t_j t_k} = \frac{\sigma_{c_1}^2}{\sigma_{t_j} \sigma_{t_k}},$$

or, as in (4)

$$(10) \quad r_{t_j t_k} = \frac{\sigma_{c_1}^2}{\sqrt{\sigma_{c_1}^2 + \sigma_{d_j}^2} \sqrt{\sigma_{c_1}^2 + \sigma_{d_k}^2}}$$

Since t_j is normally distributed (it is the difference between two normally and independently distributed variables), the joint density, $f(t_j, t_k)$, is assumed to be bivariate normal. From (4'') and setting $\sigma_{t_j} = \sigma_{t_k} = 1$ it can be expressed as

$$(11) \quad N(T_j, T_k, 1, 1, \frac{1}{2}),$$

or more explicitly as

$$(12) \quad f(t_j, t_k) = \left(\frac{1}{\sqrt{3} \pi} \right) e^{-\frac{2}{3} \left[(t_j - T_j)^2 - (t_j - T_j)(t_k - T_k) + (t_k - T_k)^2 \right]}$$

The probability of individual i preferring S_j to S_k is then

$$(13) \quad \text{Prob} \left[|t_j| < |t_k| \right]$$

or

$$(14) \quad P_{jk} = \int \int_{|t_j| < |t_k|} f(t_j, t_k) dt_j dt_k$$

Since (13) may be expressed as

$$(15) \quad \text{Prob} \left[-|t_k| < t_j < |t_k| \right] ,$$

equation (14) may be written as

$$(16) \quad p_{jk} = \int_{-\infty}^{\infty} \int_{-|t_k|}^{|t_k|} f(t_j, t_k) dt_j dt_k .$$

This equation is simplified as follows.

$$(17) \quad \begin{aligned} p_{jk} &= \int_0^{\infty} \int_{-t_k}^{t_k} f(t_j, t_k) dt_j dt_k + \int_{-\infty}^0 \int_{t_k}^{-t_k} f(t_j, t_k) dt_j dt_k \\ &= \int_0^{\infty} \int_{-t_k}^{t_k} f(t_j, t_k) dt_j dt_k - \int_{-\infty}^0 \int_{-t_k}^0 f(t_j, t_k) dt_j dt_k \end{aligned}$$

which is abbreviated as

$$p_{jk} = F_1 - F_2 .$$

Equation (17), also called equation A, can be written as

$$(18) \quad p_{jk} = 2 \int_0^{\infty} \int_{-t_k}^{t_k} f(t_j, t_k) dt_j dt_k - \int_{-\infty}^0 \int_{-t_k}^{t_k} f(t_j, t_k) dt_j dt_k$$

or by definition as

$$p_{jk} = 2F_1 - F_3 .$$

The second pair of integrals in (18), F_3 , may be reduced to a single integral by a tedious substitution of variables. To

justify the effort involved in producing this simplification, it should be remarked, first, that numerical integration will eventually be necessary, since it is well known that integrals of normal distributions cannot be evaluated in terms of a finite number of elementary functions; and secondly that the time required to evaluate a double integral is approximately equal to the square of the time necessary to evaluate a corresponding single integral.

Reduction of F_3 . - It will be necessary to reverse the order of integration and this requires having fixed limits of integration. This is accomplished by substituting

$$(19) \quad x_j = \frac{t_j}{t_k} \quad \text{and} \quad x_k = t_k$$

into F_3 . Thus

$$(20) \quad F_3 = \int_{-\infty}^{\infty} \int_{-1}^1 x_k f(x_j x_k, x_k) dx_j dx_k$$

By completing the square, the exponent in the integrand may be written as

$$(21) \quad -\frac{2}{3} \left\{ \left[x_k \sqrt{x_j^2 - x_j + 1} - \frac{x_j A + B}{\sqrt{x_j^2 - x_j + 1}} \right]^2 - \left[\frac{(x_j A + B)^2}{x_j^2 - x_j + 1} \right] + T \right\}$$

where

$$(22) \quad \begin{aligned} A &= t_j - \frac{t_k}{2}, \\ B &= t_k - \frac{t_j}{2}, \\ T &= T_j^2 - T_j T_k - T_k^2 \end{aligned}$$

Set

$$(23) \quad w_k = x_k \sqrt{x_j^2 - x_j + 1} - \frac{x_j A + B}{\sqrt{x_j^2 - x_j + 1}}$$

$$w_j = x_j,$$

substitute into (20), and obtain

$$(24) \quad F_3 = \frac{1}{\pi \sqrt{3}} e^{-\frac{2}{3}T} \int_{-1}^1 \exp \left[\frac{2}{3} \frac{(w_j A + B)^2}{w_j^2 - w_j + 1} \right] \cdot$$

$$\int_{-\infty}^{\infty} \left\{ \frac{3w_k e^{-w_k^2}}{2(w_j^2 - w_j + 1)} + \frac{\sqrt{3}(w_j A + B) e^{-w_k^2}}{\sqrt{2}(w_j^2 - w_j + 1)^{3/2}} \right\} dw_k dw_j$$

Consider the integration in (24) with respect to w_k . The integral of the first term in brackets is zero and the second term is

$$(25) \quad \left[\sqrt{\frac{3}{2}} \frac{w_j A + B}{(w_j^2 - w_j + 1)^{3/2}} \right] \sqrt{\pi}$$

Equation (24) therefore reduces to

$$(26) \quad F_3 = \frac{1}{\sqrt{2\pi}} e^{-\frac{2}{3}T} \int_{-1}^1 \left[\frac{w_j A + B}{(w_j^2 - w_j + 1)^{3/2}} \right] e^{\left[\frac{2(w_j A + B)^2}{3(w_j^2 - w_j + 1)} \right]} dw_j$$

which gives the required reduction of F_3 to a single integral.

Equation A refers to (17) where $f(t_j, t_k)$ is given in (12); Equation A' refers to (18) with F_3 given in (26).

Approximate and numerical solutions to equation A are discussed in a later section.

Equation B. - An alternative formulation of the present theory is obtained by partitioning the region in the c_1, q_j, q_k sample space having $|c_1 - q_j| < |c_1 - q_k|$ into four regions. The probability of individual i preferring S_j to S_k will be expressed as the sum of four probabilities,

$$(27B) \quad P_{jk} = P_1 + P_2 + P_3 + P_4 .$$

In region 1, $c_1 < q_j < q_k$. The probability of R_1 is then

$$(28) \quad P_1 = \int_{-\infty}^{\infty} N(C_1, 1) dc_1 \int_{c_1}^{\infty} N(Q_j, 1) dq_j \int_{q_j}^{\infty} N(Q_k, 1) dq_k$$

In region 2, $q_j < c_1 < q_k$ and $(q_k - c_1) > (c_1 - q_j)$. The probability of R_2 is

$$(29) \quad P_2 = \int_{-\infty}^{\infty} N(C_1, 1) dc_1 \int_{-\infty}^{c_1} N(Q_j, 1) dq_j \int_{2c_1 - q_j}^{\infty} N(Q_k, 1) dq_k$$

In region 3, $q_k < c_1 < q_j$ and $(c_1 - q_k) > (q_j - c_1)$. The probability of R_3 is

$$(30) \quad P_3 = \int_{-\infty}^{\infty} N(C_1, 1) dc_1 \int_{c_1}^{\infty} N(Q_j, 1) dq_j \int_{-\infty}^{2c_1 - q_j} N(Q_k, 1) dq_k$$

And in region 4, $q_k < q_j < c_1$. The probability of R_4 is

$$(31) \quad P_4 = \int_{-\infty}^{\infty} N(C_1, 1) dc_1 \int_{-\infty}^{c_1} N(Q_j, 1) dq_j \int_{-\infty}^{q_j} N(Q_k, 1) dq_k .$$

Each one of the triple integrals in (28), (29), (30) and (31) can be reduced to a double integral by performing the integration over c_1 , after suitable changes in the q_j and q_k variables. However, this tour de force will not be performed since our interest in equation B will be confined to obtaining certain asymptotic properties of the Q theory. The $\underline{P}$ terms in (27) will then be negligible or easily simplified.¹

Special conditions

Although it has been assumed that the variance of c_1 is equal to the variance of q_j , it is of some interest to examine the properties of the Q theory under two special cases: in case (1) c_1 is constant and $\sigma_{q_j} > 0$ and in case (2) q_j is constant and c_1 is a random variable. The equations for these two cases are given in this section and discussed more fully in a later section.

Case 1: $\sigma_{c_1} = 0$ and $\sigma_{q_j} > 0$ for all j . - Under these assumptions the correlation between t_j and t_k is zero so that the joint distribution of t_j and t_k is

¹An alternative approach to obtaining the p_{jk} equation is suggested by the relation

$$(32) \quad \text{Prob}(|t_j| < |t_k|) = \text{Prob}(t_j^2 < t_k^2).$$

It would seem that by obtaining the joint distribution of t_j^2 and t_k^2 , the p_{jk} equation could be written with at most one pair of integrals. This approach is deceptive. The joint distribution of t_j^2 and t_k^2 , when obtained from a change of variable in the bivariate normal distribution $f(t_j, t_k)$, separates into four terms. The result is an equation very similar to equation B.

$$(33) \quad r(t_j, t_k) = N(T_j, \sigma_q) N(T_k, \sigma_q) .$$

If (33) instead of (12) is substituted into equation A, equation A will be valid under these conditions. It may be noted, however, that equation A does not reduce to the LCJ equation in this case. The essential character of the probabilistic Q theory is maintained.

Equation B is modified in case (1) by eliminating the integration over c_1 in each term, leaving two integrals in each term. No term vanishes since there remains a finite probability associated with each of the four regions.

Case 2: $\sigma_{q_j} = 0$ for all j and $\sigma_{c_1} > 0$. - Under these conditions equations A and B are not applicable. Instead, the p_{jk} equation is expressed, for $q_k > q_j$, as

$$(33) \quad p_{jk} = \int_{-\infty}^{\frac{q_k + q_j}{2}} N(c_1, \sigma_{c_1}) dc_1 .$$

Equation (33) is simplified by a change of variable to

$$(34) \quad p_{jk} = \int_{-\infty}^{\frac{q_k + q_j - 2c_1}{2\sigma_{c_1}}} N(0,1) dx$$

or simply as

$$(35) \quad x_{jk} = \frac{q_k + q_j - 2c_1}{2\sigma_{c_1}} .$$

Equation (35) bears a direct similarity to the LCJ equation and, as will be shown, to equations A and B under certain asymptotic

conditions, conditions in which the distributions of c_1 , q_j , and q_k do not overlap.

Asymptotic properties

If C_1 , Q_j , and Q_k approach plus or minus infinity, equations A and B approach a certain asymptotic expression. Two cases may be distinguished: (1) Q_j and Q_k approach plus and minus infinity and C_1 remains fixed at some finite value, and (2) Q_j and Q_k remain finite and C_1 approaches plus or minus infinity. Coombs (1958) has termed the first condition bilateral and the second unilateral. A bilateral condition may be loosely described as one in which the three distributions, q_j , q_k , and c_1 do not overlap---or at least, not significantly---and C_1 lies between Q_j and Q_k . A unilateral condition is described by saying that C_1 lies to the left of Q_j and Q_k ($C_1 < Q_j$ and Q_k) or to the right of Q_j and Q_k ($C_1 > Q_j$ and Q_k) and that the c_1 distribution does not overlap either of the other two distributions.

Unilateral condition. - Since the c_1 distribution does not overlap the stimulus distributions it can be expected that the preference probability will not be influenced by the particular form of the c_1 distribution. In fact it is reasonable to expect that if C_1 is far enough to the left of Q_j and Q_k that p_{jk} will depend only on the discrimination between S_j and S_k , or, that is, between the separation of Q_j and Q_k . The momentary positions of the individual's ideal should not effect his preferences. He should prefer S_j to S_k whenever $q_j < q_k$. The probability of q_j being less than q_k should then be equal to p_{jk} . These consideration suggest that the p_{jk} equation under the

unilateral conditions can be written as

$$(36) \quad \begin{aligned} P_{jk} &= \int_{-\infty}^0 N(Q_j - Q_k, 1) dx && \text{for } C_1 < Q_j \text{ and } Q_k \\ P_{jk} &= \int_0^{\infty} N(Q_j - Q_k, 1) dx && \text{for } C_1 > Q_j \text{ and } Q_k \end{aligned}$$

After the usual change of variables equations (36) can also be written as

$$(37) \quad \begin{aligned} x_{jk} &= Q_k - Q_j && \text{for } C_1 < Q_j \text{ and } Q_k \\ x_{jk} &= Q_j - Q_k && \text{for } C_1 > Q_j \text{ and } Q_k \end{aligned}$$

Equation (37) bears a formal similarity to the LCJ equation---the equations are not quite equivalent---and for this reason (37) has been termed the "unilateral Law of Comparative Judgment" (Coombs, Greenberg, Zinnes, 1959).

The argument used in the preceding paragraph to derive (37) is more heuristic than rigorous. It would be desirable to check our intuitions by deriving (37) from equations A or B. A derivation from equation B readily follows so that it will be briefly described.

Derivation of unilateral equation from equation B. - Let C_1 approach minus infinity without limit and Q_j and Q_k remain fixed. It should be evident on inspection that P_2 , P_3 and P_4 will be zero since each one refers to a region which has a negligible probability under this assumption. P_2 , for example, refers to the region in which $c_1 > q_j$ and this inequality should have a negligible probability if C_1 approaches negative infinity and Q_j remains finite.

A more rigorous argument would consist of manipulating directly the expression for P_2 . This would involve first changing the variables of integration to produce constant limits of integration, integrating over c_1 , and then showing that the expression goes to zero as C_1 approaches minus infinity.

The P_1 expression can be simplified by performing the integration over c_1 . This integration can be performed independently of the remaining two integrals, since these integrals do not contain c_1 . This integration is, of course, equal to one. P_1 may then be written as

$$(39) \quad P_1 = \int_{-\infty}^{\infty} N(Q_j, 1) dq_j \int_{q_j}^{\infty} N(Q_k, 1) dq_k.$$

Performing a change of variable ($q_k - q_j = y_k$ and $q_j = y_j$) and reversing the order of integration in (39) yields

$$(40) \quad P_1 = \int_0^{\infty} \left\{ \int_{-\infty}^{\infty} N(Q_j, 1) N(Q_k, 1) dy_j \right\} dy_k$$

[The notation in (40) should show that $N(Q_k, 1)$ is a function of $y_k + y_j$.] The integration indicated in brackets in (40) can be carried out. It is equal to

$$(41) \quad N(Q_k - Q_j, \sqrt{2}),$$

so that

$$(42) \quad P_1 = \int_0^{\infty} (Q_k - Q_j, \sqrt{2}) dy_k.$$

Equation (42) can also be written as

$$(43) \quad P_1 = \int_{-\infty}^0 (Q_j - Q_k, \sqrt{2}) dy_k,$$

which agrees with the unilateral left equation in (36), except for the variance term. The difference in variance, however, is only a difference in the unit of measurement and this is arbitrary.

The unilateral equations will be useful in obtaining an approximate solution to the scale values of the stimuli Q_j and Q_k . To determine $Q_j - Q_k$, it is only necessary to locate a subject who is far to the right or left of this pair. This procedure is discussed in more detail in a later section.

Bilateral condition. - In the bilateral condition, the three distributions q_j , q_k , and c_1 are assumed to be non-overlapping. If Q_j is to the left of C_1 , and Q_k to the right of C_1 , then at all times

$$(44) \quad |c_1 - q_j| - |c_1 - q_k| = (c_1 - q_j) - (q_k - c_1) = 2c_1 - q_j - q_k$$

Since the sum or difference of normally distributed variables is normal, p_{jk} can immediately be written as

$$(45) \quad p_{jk} = \int_{-\infty}^0 N(2c_1 - q_j - q_k, \sqrt{2}) dx \quad \text{for } q_j < c_1 < q_k$$

where the variance of $c_1 - q_j$ is set equal to one. Equation (45) can also be written in the usual form as¹

¹This result due to Professor Coombs.

$$(46) \quad x_{jk}\sqrt{2} = q_j + q_k - 2c_1 \quad \text{for } q_j < c_1 < q_k.$$

Equation (46) can also be derived from equation B by showing first that $P_1 = P_3 = P_4 = 0$ and then simplifying the expression for P_2 . This will not be carried out.

If $(q_j - q_k)$ has been determined from the unilateral equation and if q_j and q_k are far enough apart and on opposite sides of c_1 , then it would seem that (46) could be used to locate c_1 . This approximate method of solution is discussed in the next section.

Approximate solution

The simplicity of the unilateral and bilateral equations suggests that solutions of c_1 , q_j , and q_k could easily be obtained from the p_{jk} values of a group of subjects. The procedure would be as follows.

1) Determine the I scale for each subject as indicated in Table 1, Chapter I. Assuming that some of the stimulus and ideal distributions overlap---the non-overlapping case is trivial since it is treated by the unfolding technique---then it can be expected that the I scale of a subject will not give a simple order of the s_{1j} values. However, one would expect that even a very rough or weak order of the stimuli would be adequate.

2) Determine the order of the stimuli on the J scale, that is, the q_j values of the stimuli. Again it cannot be expected that the number of I scales will be less than $\binom{n}{2} + 1$ (n equals the number of stimuli) or that the I scales will be ordered as required by condition 3 (see Chapter I). In brief, the I scales will not

necessarily fit a perfect J scale. However, one could adopt the criterion of selecting the J scale consistent with the largest number of I scales. Here too, only a rough order of the stimuli would seem to be required so that the above criterion need not specify a unique J scale.

3) With a weak order of the stimuli determined on the J scale and with the subjects approximately located on this scale, it should be possible, for a particular S_j and S_k , to select subjects who are far to the left or right of Q_j and Q_k . These subjects would then determine a P_{jk} , which could be used in the unilateral equations to give $Q_j - Q_k$. To obtain the scale values of every pair of stimuli it would be necessary to have a distribution of subjects that was widely dispersed over the J scale.

4) Determine C_1 from the bilateral equation. This requires selecting for each subject a pair of stimuli satisfying the bilateral conditions. It would be necessary to replicate this pair of stimuli quite often to obtain a stable P_{jk} for each individual. (The unilateral equation required only a P_{jk} for a group, not for a single subject.)

There are obviously many loopholes in these four steps which make one question the feasibility of the method. Is it feasible, for example, to determine a J scale for a large number of distinct I scales? How non-unique a J scale will the method tolerate? Can unilateral subjects far to the right or left be selected from a population having only weak order I scales? And precisely how far to the right or left do the subjects have to be located in order to permit solutions with a given degree of accuracy? Concerning

the bilateral equation, is it feasible to replicate a single pair of stimuli for an individual? There is also the question of determining the goodness of fit of the model. The estimate of goodness of fit is complicated by the fact that the method only uses part of the data---only certain unilateral and bilateral judgments---to obtain a solution.

The plan of this study is to answer some of these questions experimentally, using the method with actual data and obtaining a solution; other questions will be discussed more theoretically. The basic procedure in the experimental application is one of making certain tentative assumptions and then checking back, after a solution is obtained, to see whether these assumptions were met. If the assumptions were met at each stage in the analysis then the application is at least internally consistent. Furthermore, if a J scale is, in fact, obtained, then it may be argued that the method is feasible.

Some of the theoretical issues of this approximate method are discussed in the following sections of this chapter.

Accuracy of the unilateral equations

The approximate method of solution described in the previous section requires selecting subjects who are far to the right or left of a given pair of stimuli. The question to be answered in this section is: how far from the given pair of stimuli should the subjects be located to justify the use of the unilateral equations at some prescribed level of accuracy? This question can be answered by computing the value of p_{jk} for different values of C_j , keeping $Q_j - Q_k$ constant. (The method of computation and its accuracy are described in Appendices 1 and 2.) The results of three such series

of calculations are given in Table 6, and in Figure 1 the values of p_{jk} are plotted against T_j . Positive values of T_j in this figure indicate how far to the right of Q_j the subject is located.

Except for the computing inaccuracies in series I ($Q_j - Q_k = 2$), it can be seen that the three curves in Figure 1 uniformly approach asymptotes which agree with the estimates from the corresponding unilateral equation. For subjects located approximately two units to the right of Q_j , Table 7a gives the deviations between their p_{jk} value and the asymptotic value. (The location of T_j in series I is given at 1.5 units because of the computational errors in p_{jk} for $T_j = 2.0$.) Considering the inaccuracies of computation, that the p_{jk} values are probably underestimated by at least .001, it is reasonable to conclude from Table 7a that the selection of subjects 2σ units to the right or left of a given stimulus pair will give nearly three digit estimates of the asymptotic p_{jk} values. Estimates of $Q_j - Q_k$ obtained from the unilateral equations based on these subjects should therefore be accurate to approximately three digits.

Table 7a gives the deviation of p_{jk} from asymptotic values for subjects located 1.5 units to the right of Q_j . Except for series I the p_{jk} values based on these subjects have approximately two digit accuracy. The fact that the accuracy of p_{jk} is better in series I suggests that the unilateral equations will approximate large $Q_j - Q_k$ intervals better than small ones.

Alternative estimates of each $Q_j - Q_k$ interval can be obtained by taking the difference between two other intervals, for example $(Q_1 - Q_k) - (Q_1 - Q_j)$. The usual procedure in obtaining an ICJ

TABLE 4

ACCURACY OF NUMERICAL INTEGRATION FOR $p_{jk} = .5$

Series	$Q_j - Q_k$	p_{jk}	h
I	2.0	.4985	.40
II	1.0	.4986	.35
III	0.4	.4986	.32

TABLE 5

ACCURACY OF NUMERICAL INTEGRATION FOR ASYMPTOTIC VALUES OF p_{jk}

Series	p_{jk}	Asymptotic p_{jk}	Deviation	h
I	.9763	.9773	.0010	.65
II	.8389	.8413	.0024	.70
III	.6552	.6554	.0002	.60

TABLE 6
VALUES OF p_{jk} FOR DIFFERENT VALUES OF C_1

Series I: $Q_j = 1, Q_k = -1$			Series II: $Q_j = \frac{1}{2}, Q_k = -\frac{1}{2}$			Series III: $Q_j = .2, Q_k = -.2$		
T_j	p_{jk}	h	T_j	p_{jk}	h	T_j	p_{jk}	h
-1.0	.4985	.40	- .5	.4986	.35	- .2	.4986	.32
- .8	.5856	.42	- .3	.5610	.37	- .1	.5130	.33
- .6	.6682	.44	- .1	.6200	.39	0	.5272	.34
- .4	.7425	.46	.1	.6731	.41	.1	.5410	.35
- .2	.8058	.48	.3	.7183	.43	.2	.5542	.36
0	.8570	.50	.5	.7549	.45	.3	.5667	.37
.2	.8963	.52	.7	.7830	.47	.4	.5784	.38
.4	.9251	.54	.9	.8034	.49	.5	.5892	.39
.6	.9450	.56	1.1	.8175	.51	.6	.5990	.40
.8	.9585	.58	1.3	.8266	.53	.8	.6157	.42
1.0	.9670	.60	1.5	.8321	.55	1.0	.6284	.44
1.5	.9763	.65	2.0	.8370	.60	1.2	.6376	.46
2.0	.9750	.70	2.5	.8373	.65	1.4	.6440	.48
2.5	.9653	.75	3.0	.8389	.70	1.6	.6482	.50
3.0	.9502	.80				2.1	.6531	.55
						2.6	.6552	.60
Asymptote: .9773			Asymptote: .8413			Asymptote: .6554		

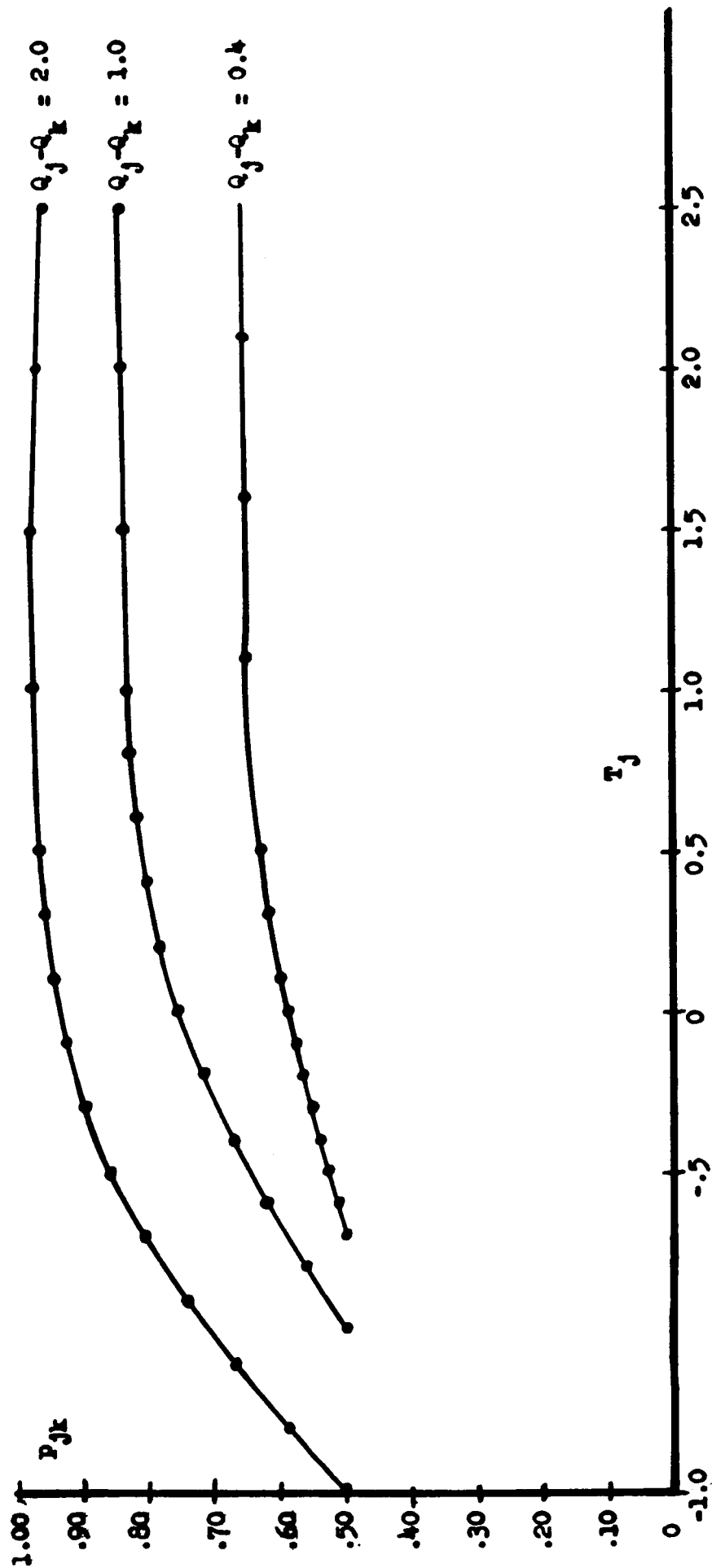


Fig. 1 - The relationship between p_{jk} and T_j for three values of $Q_j - Q_k$.

TABLE 7
ACCURACY OF THE UNILATERAL EQUATIONS

Series	$Q_j - Q_k$	$C_1 - Q_j$	Deviation of p_{jk} from asymptotic value
I	2.0	1.5	.001
II	1.0	2.0	.004
III	.4	2.1	.002

(a)

I	2.0	.001
II	1.0	.009
(III	.4	.011)
III	.4	.007)

(b)

I	2.0	.010
(II	1.0	.024)
II	1.0	.038)
III	.4	.027

(c)

scale is to average the separate estimates of $Q_j - Q_k$ where S_j and S_k are adjacent on the scale. If this method is to be used in conjunction with the unilateral equations (to obtain the Q values on the J scale), then it should be ascertained first whether the average of the separate estimates is, in fact, a better estimate of the $Q_j - Q_k$ interval than the estimate based on p_{jk} only. The accuracy of the separate estimates of a given interval can be gauged from the following numerical example.

Table 8 consists of a p_{jk} matrix which is compatible with the Q theory. The probabilities in this table were generated by numerically integrating F_1 and F_2 in equation A, using as values of Q_j , Q_k , and C_1 those given in Figure 2. Consider the stimuli: A,B,C,D, which are to the left of C_1 on the hypothetical J scale. Using the unilateral equation and the probabilities associated with the four stimuli, A,B,C,D, three estimates of the intervals $Q_A - Q_B$, $Q_B - Q_C$ and $Q_C - Q_D$ are possible. Consider the three estimates of $Q_A - Q_B$.

$$P_{AB}: Q_B - Q_A = .2637$$

$$\begin{aligned} P_{BC}, P_{AC}: Q_B - Q_A &= (Q_B - Q_C) - (Q_A - Q_C) \\ &= -.1611 - (-.4207) = .2596 \end{aligned}$$

$$\begin{aligned} P_{BD}, P_{AD}: Q_B - Q_A &= (Q_B - Q_D) - (Q_A - Q_D) \\ &= (-.4070) - (-.6557) = .2487. \end{aligned}$$

Since the correct value of $Q_A - Q_B$, as given by the hypothetical J scale in Figure 2, is equal to 0.3, the best estimate of this interval is given by P_{AB} . This example suggests that only one estimate of adjacent intervals should be used. However, the generality of this conclusion may be suspect, since the alternative estimates of $Q_B - Q_A$

TABLE 8

P_{jk} MATRIX GENERATED FROM HYPOTHETICAL J SCALE

	E	F	D	G	C	B	H	A
E		.5218	.5565	.6137	.6473	.6977	.7045	.7706
F			.5311	.6053	.6126	.6573	.7123	.7228
D				.5586	.6007	.6583	.6472	.7441
G					.5211	.5659	.6239	.6331
C						.5635	.5672	.6627
B							.5229	.6038
H								.5456
A								

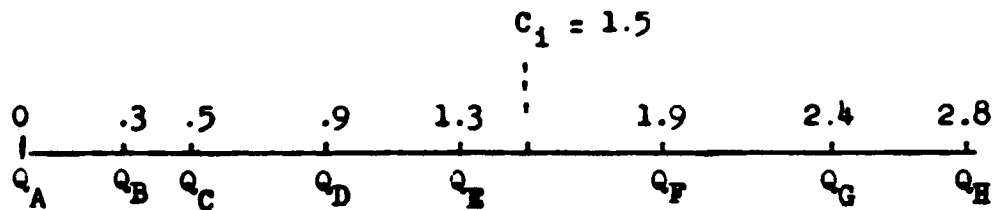


Fig. 2 - Hypothetical J scale.

were based on stimuli closer to C_1 than B or A. Another example should therefore be considered which does not have this property.

Consider the interval $Q_D - Q_C$ and the three estimates of this interval based on A, B, C, and D. Two of the estimates will be based on A and stimuli further from C_1 than C and D. The three estimates are as follows

$$\begin{aligned} P_{DC}: Q_D - Q_C &= .2559 \\ P_{DA}, P_{CA}: Q_D - Q_C &= (Q_D - Q_A) - (Q_C - Q_A) \\ &= .6557 - .4207 = .2350 \\ P_{DB}, P_{CB}: Q_D - Q_C &= (Q_D - Q_B) - (Q_C - Q_B) \\ &= .4070 - .1611 = .2459. \end{aligned}$$

The correct value of $Q_D - Q_C$ in this example is .4. The best estimate therefore is provided by P_{DC} , indicating, it would seem, that the direct estimate of a given interval is probably more accurate than the estimates based on differences between two larger intervals.

When estimates of an interval are based on stimuli quite far removed from the ideal, far enough, that is, for the unilateral equations to yield considerable accuracy, then it would seem appropriate to average the several estimates based on these stimuli. However, when the accuracy of the unilateral equations are somewhat less than desired, then it would seem appropriate to maximize this accuracy by using only the direct estimate. Of course, in using only the direct estimate of an interval, one may be throwing away data already collected.

Strong stochastic transitivity

In Chapter I it was asserted that the present probabilistic

Q theory did not imply the SST condition and, therefore, that this theory was incompatible with what has been called in Chapter II the U and V theories. In this section the non-SST properties of the general theory will be demonstrated. First, however, certain special conditions will be considered, conditions which one might expect would satisfy SST.

The SST property for fixed stimulus points and a variable ideal point.- This special condition, $\sigma_{c_1} > 0$, $\sigma_{q_j} = 0$, was previously labelled case 2; the relevant equation is therefore (34) or (35). Equation (35) can be written as

$$x_{jk} = \left[(q_j - c_1) - (c_1 - q_k) \right] \frac{1}{2\sigma_{c_1}} \\ = - (t_j + t_k).$$

Let us consider an example in which $q_j < q_k$, $q_k < q_1$, and

$$(47) \quad p_{jk} > \frac{1}{2} \\ p_{k1} > \frac{1}{2}.$$

Adding $(q_1 - 2c_1)$ to each side of the inequality

$$q_j < q_k,$$

gives $(q_j + q_1 - 2c_1) < (q_k + q_1 - 2c_1)$

which by equation (35) and assumption (47) implies that

$$x_{j1} < x_{k1} \\ \text{or} \\ (49) \quad p_{j1} < p_{k1}.$$

But (49) contradicts SST. It must be concluded, therefore, that SST is not implied by case 2.

SST for non-overlapping distributions. - The assumption of non-overlapping distributions implies that the distribution of c_1 does not significantly overlap the q_j or q_k distribution, although the two stimulus distributions may overlap each other. This is the assumption basic to the unilateral and bilateral conditions. The region in which non-SST probabilities are obtained under this condition can be indicated with the following example.

Let $q_j < q_k < c_1 < q_1$, and $\sigma_{q_j} = \sigma_{q_k} = \sigma_{q_1} = \sigma_{c_1} = \sigma$ (see Figure 3). Assume that

$$(50) \quad \begin{aligned} p_{kj} &> \frac{1}{2} \\ p_{j1} &> \frac{1}{2} . \end{aligned}$$

SST will be contradicted if

$$(51) \quad p_{k1} < p_{kj},$$

and (51) will occur, according to the unilateral and bilateral equations, if

$$(52) \quad \begin{aligned} &x_{k1} < x_{kj} \\ \text{or} \\ &\frac{(q_1 - c_1) - (c_1 - q_k)}{\sqrt{6}\sigma} < \frac{q_k - q_j}{\sqrt{2}\sigma} . \end{aligned}$$

Inequality (52) can be written as

$$(53) \quad (q_1 - c_1) < (\sqrt{3} - 1)(q_k - q_j) + (c_1 - q_j).$$

However, from (50) and the unilateral equation for p_{j1} ,

$$(54) \quad \begin{aligned} &x_{j1} = \frac{(q_1 - c_1) - (c_1 - q_j)}{\sqrt{6}\sigma} > 0 \\ \text{or} \\ &(q_1 - c_1) > (c_1 - q_j). \end{aligned}$$

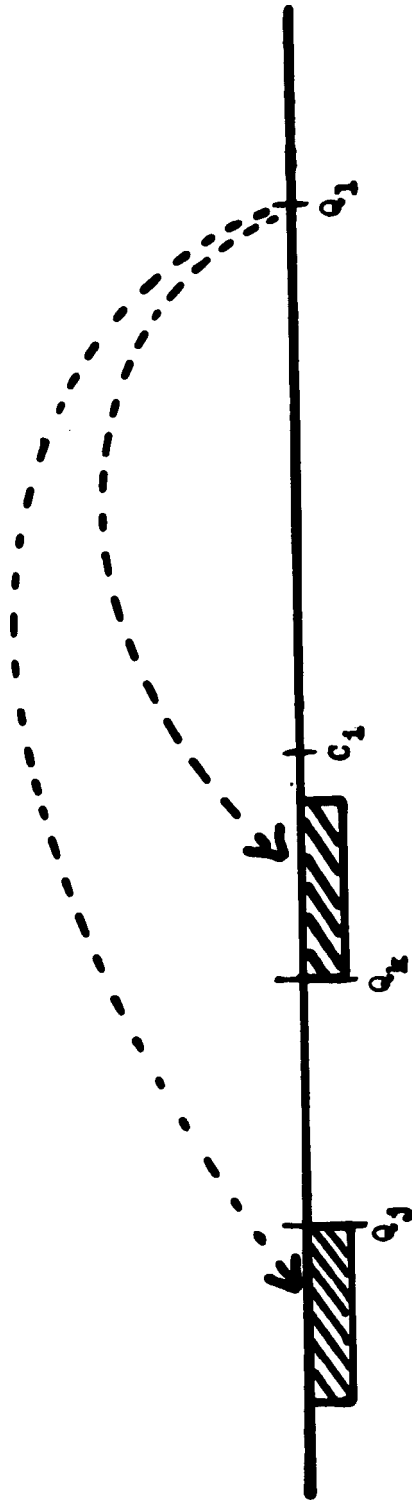


Fig. 3 - Non-SST property for non-overlapping distributions. The shaded area, which is equal to $(\sqrt{3} - 1)(Q_k - Q_j)$, indicates the locations of Q_l (on the folded scale) that result in non-SST probabilities.

Combining (53) and (54) gives

$$(55) \quad (c_1 - q_j) < (q_1 - c_1) < \left[(\sqrt{3} - 1)(q_k - q_j) + (c_1 - q_j) \right].$$

Equation (55) gives the range of values of q_1 in which SST will not be satisfied.

It is also possible to show by using

$$(56) \quad \begin{aligned} p_{1k} &> \frac{1}{2} \\ p_{kj} &> \frac{1}{2}, \end{aligned}$$

instead of (50), that another non-SST region exists, namely

$$(57) \quad \left[(1 - \sqrt{3})(q_k - q_j) + (c_1 - q_k) \right] < (q_1 - c_1) < c_1 - q_k.$$

Both regions (55) and (57), can be represented graphically in terms of a "folded" scale. If the J scale containing q_j, q_k, c_1 and q_1 is folded at c_1 , then the shaded regions in Figure 3 indicate locations of the folded q_1 point in which non-SST probabilities will be obtained. It should be noted that for the condition of non-overlapping distributions, non-SST probabilities can only occur when a bilateral pair of stimuli is involved.

SST for the general case.- The general case refers to the condition in which both stimulus point and ideal point vary and in which the c distribution is allowed to overlap the stimulus distributions. Since the theory in this case is described by the cumbersome A or B equations, investigating the SST properties of the theory is difficult. However, the non-SST character of the theory can be illustrated by using the hypothetical data in Table 8, generated from the Q and C values given in Figure 2. The intervals between the Q values in this example were made relatively small

so that appreciable overlap of the stimulus distributions would occur. Also C_1 was placed approximately in the center of the scale so the distribution of the ideal point would appreciably overlap many of the stimulus distributions. The variance of t_j was set equal to one.

Inspection of Table 8 indicates two non-SST triples: (B,D,H) and (D,F,A). The first of these has

$$P_{DB} = .6583 > \frac{1}{2}$$

$$P_{BH} = .5229 > \frac{1}{2}$$

but, contrary to SST,

$$P_{DH} < P_{DB}$$

since

$$P_{DH} = .6427.$$

The second triple has

$$P_{FD} = .5311 > \frac{1}{2}$$

$$P_{DA} = .7411 > \frac{1}{2}$$

and

$$P_{FA} = .7228 .$$

Clearly both triples do not satisfy SST. It may be observed that while the (B,D,H) triple satisfies the requirements of a non-SST triple described in Figure 3 for non-overlapping distributions, the (D,F,H) triple does not satisfy these requirements. The regions indicated in Figure 3 therefore, do not apply in the general case.

Case 1: fixed ideal point and varying stimulus points.-

Equation A with the distribution function described in (33) is appropriate to this case. Equation A with this function is still

unwieldy so that numerical methods are necessary. Although these evaluations have not been carried out, it is reasonable to surmise, in view of the results for the three conditions described above, that non-SST triples would occur in this condition also.

Estimation of I scales

The initial step in the approximate method of solution using the unilateral equations is to estimate the I scale for each subject, estimate, that is, the order of s_{ij} for all j . If each pair of stimuli is replicated a few times---to obtain more stable estimates of the I scale---the question arises as to how the several judgments for each pair can be amalgamated to give the best estimate of the I scale. The problem is complicated somewhat by the non-SST triples which are predicted by the theory. If the theory satisfied SST, then a reasonable procedure would be to use a vote count method. The number of "votes" a stimulus receives is the total number of times the stimulus is preferred to all other stimuli; this corresponds to the row sum in an (n_{jk}) matrix. The vote count method simply ranks the stimuli according to the total number of votes each receives. However, this method is not applicable with the present theory, since p_{jk} (or n_{jk}) is not monotonically related to the difference between s_{ij} and s_{ik} . This means that for three stimuli S_j, S_k, S_l , not satisfying SST, that if S_j receives more votes than S_k it does not follow that $s_{ij} > s_{ik}$.

An alternative method is to use what Coombs (1958) has called the individual's stochastic preference. If the proportion of choices in favor of S_j over S_k exceeds one half, then S_j is

said to be stochastically preferred to S_k . An I scale can then be estimated by ranking the stimuli according to the number of pairs in which they are stochastically preferred. The validity of this method depends on WST, so that its application to the present theory requires first a demonstration that the theory satisfies WST. A proof of WST easily follows from the special cases---for example, the case involving non-overlapping distributions. It is also apparent from the hypothetical data matrix in Table 8 that WST is satisfied. (That WST is implied in the matrix is indicated by the fact that all the values of p_{jk} in the upper half of the matrix are greater than one half.) However, a general proof from equation A or B has not been obtained. It will be assumed, nevertheless, that the theory may satisfy WST so that the method of stochastic preference outlined above can be used.

Group results

The discussion of the approximate method of solution has largely been confined to considerations of the unilateral equations. The bilateral equations were not dealt with because the preliminary experimental results indicated that its application was not feasible. For reasons to be discussed in Chapter VI it was not practical to replicate a single stimulus pair the required number of times. There are also some theoretical considerations for not using the bilateral equations. Since the plan of this study is to compare the fit of the Q theory to the U and V theory, the number of free parameters in each theory should be approximately equal and furthermore they should be estimated from the same data. Using a large number of replications to estimate C_1 for each subject in the Q

theory would give this theory an unfair advantage.

The procedure employed in this study is to assume that C_1 is normally distributed over subjects. The effect of this assumption is to reduce considerably the number of parameters to be estimated. Instead of having to estimate C_1 for each subject, only the average C_1 (denoted by C') and the variance of C_1 ($\sigma_{C_1}^2$) have to be estimated. However, it should be observed that the Q theory will still have two free parameters more than the U and V theory. (For n stimuli, the U and V theory have $n-1$ free parameters.)

Estimates of C' and σ_{C_1} can be made after the unilateral equations have been applied and the Q values obtained. It is only necessary to fit a normal curve to the histogram of subjects on this J scale.

There remains the problem of "postdicting" the p_{jk} values---to determine the degree of fit of the theory---from the values of $Q_1 \dots Q_n$ and C' and σ_{C_1} . Although equation A was developed for a single individual, it has the remarkable property of being directly applicable to a group when the normal distribution assumption of C_1 is made. The only modifications necessary in equation A are modifications of the parameters appearing in the bivariate normal distribution. The justification for these remarks lies in the well known theorem that if c_1 is normally distributed for fixed $\underline{1}$,

$$N(C_1, \sigma_{C_1}),$$

and C_1 is normally distributed over $\underline{1}$,

$$N(C', \sigma_{C_1})$$

then the distribution of c ---the subscript $\underline{1}$ can be dropped---in the

superpopulation composed of within subjects and between subjects variability is

$$N(c', \sqrt{\sigma_{c_1}^2 + \sigma_{c_1'}^2}).$$

This theorem follows from the fact that the density function of c , written as

$$(58) \quad \frac{1}{2\pi\sigma_{c_1}\sigma_{c_1'}} \int_{-\infty}^{\infty} \exp\left[-\frac{(c_1 - c')^2}{2\sigma_{c_1}^2}\right] \exp\left[-\frac{(c_1 - c_1')^2}{2\sigma_{c_1'}^2}\right] dc_1$$

can be reduced, performing the integration over c_1 , to

$$(59) \quad \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{c_1}^2 + \sigma_{c_1'}^2}} \exp\left[-\frac{(c - c')^2}{2(\sigma_{c_1}^2 + \sigma_{c_1'}^2)}\right].$$

The variables of the bivariate normal function appropriate for group data are

$$\begin{aligned} t'_j &= c - q_j \\ t'_k &= c - q_k, \end{aligned}$$

where c refers to the momentary location of an ideal point selected from the "superpopulation". The corresponding mean values are then

$$(60) \quad \begin{aligned} T'_j &= C' - Q_j \\ T'_k &= C' - Q_k. \end{aligned}$$

There are two more parameters to determine before the function

$f(t'_j, t'_k)$, appearing in equation A, is completely specified: $r_{t'_j t'_k}$ and $\sigma_{t'_j t'_k}^2$.

The variance of t'_j is

$$\sigma_{t'_j}^2 = \sigma_q^2 + \sigma_{c_1}^2 + \sigma_{c_1}^2 \quad \text{for all } j$$

or since $\sigma_{q_j}^2 + \sigma_{c_1}^2 = 1$

$$(61) \quad \sigma_{t'_j}^2 = 1 + \sigma_{c_1}^2 \quad \text{for all } j.$$

The correlation term, which was previously equal to one half, requires modification. The previous result was based on (4'') which assumed that $\sigma_{c_1}^2 = \sigma_{q_j}^2$. For the group result

$$(62) \quad \sigma_{c_1}^2 + \sigma_{c_1}^2 \neq \sigma_q^2$$

so that (4') must be used instead of (4''). Equation (4') gives

$$(63) \quad r_{t'_j t'_k} = \frac{\sigma_{c_1}^2 + \sigma_{c_1}^2}{\sigma_{c_1}^2 + \sigma_{c_1}^2 + \sigma_q^2}$$

which under the present assumptions is equal to

$$(64) \quad r_{t'_j t'_k} = \frac{\sigma_{c_1} + .5}{1 + \sigma_{c_1}^2}.$$

If $r(t'_j, t'_k)$ is used in equation A with the bivariate parameters given in (60), (61) and (64), then equation A will predict the group results and it will be compatible with the unilateral equations. This means that if the Q values have been estimated from the unilateral equations, then equation A will reproduce the P_{jk} values for the entire group.

Since the above group results appear simpler than would have

been hoped for, it seems desirable to check on these results by deriving them more directly from the theory. This can be done by performing the following integration

$$(65) \quad \frac{1}{\sqrt{2\pi}\sigma_{c_1}} \int_{-\infty}^{\infty} \exp\left[-\frac{(c_1-c')^2}{2\sigma_{c_1}^2}\right] (F_1-F_2) dc_1.$$

The integration over c_1 can be performed after the usual change of variable procedure. The results, after considerably rearranging of terms, is compatible with the conclusions arrived at above. That is, (65) reduces to equation A with the parameters given in (60), (61) and (64).

CHAPTER IV

THE SIMILARITY OF PROBABILISTIC THEORIES OF CHOICE

In Chapter III it was shown that the probabilistic Q theory does not imply SST and in Chapter II that the U and V theories do imply SST. The remarks made in Chapter I, concerning the incompatibility of the Q theory with current theories, are therefore justified. However, one can ask how incompatible these theories are. Can the U or V theory be made to give a close, although necessarily imperfect, fit to a p_{jk} matrix generated from the Q theory? Or will the deviations be large enough to reject the U and V theories for a given number of replications?

The answers to these questions are considered in this Chapter. A related question concerns the difference between the U and V theories. While both theories imply SST, the two theories are not equivalent. How different they are, particularly in terms of fitting a p_{jk} matrix generated from the Q theory, will also be discussed in this Chapter.

U and V theories

The fact that the U and V theories are not equivalent is easily demonstrated, for, as Luce (1959) points out, if the two theories were equivalent then the integral of the normal function could be expressed in terms of a finite number of elementary

functions. It is well known, however, that the integral of the normal does not have this property. Although not equivalent, the U and V theories can be made to have a certain formal similarity.

The P_{jk} equation of the V theory can be written in the following form.

$$(1) \quad P_{jk} = \frac{1}{t} \int_{-\infty}^{\log \frac{v_j}{v_k}} \operatorname{sech}^2 \frac{y}{2} dy$$

$$\text{where } \operatorname{sech}^2 \frac{y}{2} = \frac{4}{e^y + e^{-y} + 2}.$$

Equation (1) is directly analogous to the ICJ equation, where the normal function has been replaced by the squared hyperbolic secant function and where $U_j - U_k$ has been replaced by $\log v_j - \log v_k$. However, the similarity is more than formal. According to Torgerson (1958) the two functions---the normal and squared hyperbolic secant---are "virtually linearly related over the range of proportions ordinarily used" (1958, p. 202). A plot between $U_j - U_k$ and $\log \frac{v_j}{v_k}$ given in Torgerson confirms this assertion. The curve is quite linear between $\pm 2\sigma$, the range of $U_j - U_k$ values given in the figure.

Numerical values presented in Luce (1959, p.56) also indicate close equivalence of the U and V theories. Luce gives the values of P_{j1} as computed from various values of P_{jk} and P_{k1} using both the U and V theories. The average discrepancy in these computations is .0082. The largest discrepancy between the P_{j1} values as determined by the U and V theories occurs when

$$P_{jk} = P_{k1} = .8.$$

The p_{j1} value for the U and V theories are then

$$p_{j1} = .941 \text{ for the V theory,}$$

$$p_{j1} = .954 \text{ for the U theory,}$$

which is a discrepancy of .013. Somewhat larger discrepancies can be obtained if more extreme values are used for p_{jk} or p_{k1} . (Luce used values between .6 and .9.) For example: if $p_{jk} = .3085$, $p_{k1} = .9332$, then

$$\text{and } p_{j1} = .8618 \text{ for the V theory,}$$

$$p_{j1} = .8413 \text{ for the U theory.}$$

The discrepancy between U and V theories in this example is .0205.

It would appear from the above examples that, for probabilities not too extreme, the predictions from the U and V theory would agree to two digits. In actual practice, however, the agreement is much closer to three digits. This greater similarity results from using a fitting procedure which minimizes deviations by placing predicted p_{jk} values somewhat between the obtained p_{jk} values. This effect can be illustrated with the p_{jk} matrix in Table 9. This table consists of p_{jk} values which were generated from the V theory by using the following scale values.

v_1	v_2	v_3	v_4	v_5	v_6
100	93	84	70	32	20 .

The p_{jk}^U terms in Table 10 are the values predicted from the U theory. These values were obtained by fitting the ICJ scale, using the least squares method described by Mosteller (1951), to the

TABLE 9

HYPOTHETICAL DATA GENERATED FROM V THEORY

	(1)	(2)	(3)	(4)	(5)	(6)
(1)		.518	.543	.588	.758	.833
(2)			.525	.571	.744	.823
(3)				.545	.724	.808
(4)					.686	.778
(5)						.615
(6)						

TABLE 10

SIMILARITY OF U AND V THEORIES

	(1)	(2)	(3)	(4)	(5)	(6)
(1) p*		.517	.542	.586	.757	.836
p		.518	.543	.588	.758	.833
d		-.001	-.001	-.002	-.001	.003
(2) p*			.525	.569	.743	.825
p			.525	.571	.744	.823
d			.000	-.002	-.001	.002
(3) p*				.545	.723	.808
p				.545	.724	.808
d				.000	-.001	.000
(4) p*					.684	.776
p					.686	.778
d					-.002	-.002
(5) p*						.610
p						.615
d						-.005

(6)

Note: p* denotes a probability reproduced by the U theory, p denotes a probability generated from the V theory, d denotes (p* - p).

p_{jk} matrix in Table 9. The ICJ scale is then used to "reproduce" the pairwise probabilities. The differences between the probabilities derived from the V theory, p_{jk} , and the probabilities predicted from the U theory, p_{jk}^* , are also given in Table 10. These differences, it will be seen, are all smaller than the average deviations obtained in the Luce examples. The average deviation in Table 10 is .0018, suggesting that, in fact, the agreement between the two theories is generally even closer than the two digit accuracy indicated by Luce.

The degree of similarity between the U and V theories can be expressed alternatively by performing the goodness of fit test described by Mosteller (1951). This test involves computing the angles θ_{jk} and θ_{jk}^* (in degrees) from p_{jk} and p_{jk}^* by

$$\theta_{jk} = \arcsin \sqrt{p_{jk}} \quad \text{and} \\ \theta_{jk}^* = \arcsin \sqrt{p_{jk}^*},$$

and then determining the residual sum of squares, SS_D , expressed in terms of θ_{jk} and θ_{jk}^* ; that is

$$SS_D = \sum_{j,k} \frac{n(n-1)}{2} (\theta_{jk} - \theta_{jk}^*)^2.$$

If SS_D is divided by $\frac{820.7}{n}$ where n is the number of replications of each stimulus pair, the resulting expression has a χ^2 distribution with $\frac{(n-1)(n-2)}{2}$ degrees of freedom. Since the data in Table 9 are hypothetical, n does not have a specific value. However, it is of interest to determine the value of n which would produce a χ^2 term equal to 23.21, the value of χ^2 at the .01 level for 21 degrees of freedom. This value of n , in the present example, is

$$n = \frac{(820.7)(23.21)}{SS_D} \approx 100,000.$$

This result can be interpreted as follows. If the hypothetical data in Table 9 were based on at least one hundred thousand replications per pair, the U theory could be rejected at the .01 level.

Not only do the U and V theories have similar predictions, but since they have the same character---both imply SST---it can be expected that if one theory overestimates an obtained P_{jk} value the other theory will also tend to overestimate it. This means that both theories will tend to respond similarly to a given P_{jk} matrix, so that distinguishing the two theories experimentally will be a difficult task. This conjecture is confirmed in the next section using a P_{jk} matrix generated from the Q theory.

The fit of U and V theories to Q theory data

An indication of the similarity of the U and V theories to the Q theory can be obtained by fitting the two theories to the P_{jk} matrix given in Table 8. This matrix, it will be recalled, consists of P_{jk} values generated from the Q theory. These P_{jk} values can therefore be completely described or accounted for by the Q theory.

Table 11 gives the results of fitting the U theory to the matrix in Table 8. The values reproduced from the U theory, P_{jk}^* , are obtained from the ICJ scale which produces a least squares fit (as defined by Mosteller, 1951) to the data matrix. The fact that the U theory does not precisely predict these P_{jk} values---the deviations in this matrix are all greater than zero---is prima facie evidence once again for the incompatibility of the Q and the U theory. The

TABLE 11

SIMILARITY OF U AND Q THEORIES AS INDICATED BY REPRODUCTION METHOD

	E	F	D	G	C	B	H	A
p*		.527	.552	.622	.641	.692	.716	.766
E p		.522	.557	.614	.647	.698	.705	.771
d		.005	-.005	.008	-.006	-.006	.011	-.005
p*			.525	.597	.616	.668	.693	.745
F p			.531	.605	.613	.657	.712	.723
d			-.006	-.008	.003	.011	-.019	.022
p*				.572	.591	.644	.670	.724
D p				.559	.601	.658	.647	.744
d				.013	-.010	-.014	.023	.020
p*					.520	.575	.602	.660
G p					.521	.566	.624	.633
d					-.001	.009	-.022	.027
p*						.555	.583	.642
C p						.564	.567	.663
d						-.009	.016	-.021
p*							.528	.589
B p							.523	.604
d							.005	-.015
p*								.561
H p								.546
d								.015

A

Note: p* denotes a value reproduced by the U theory
p denotes a value generated by the Q theory.

average absolute deviation between p_{jk}^* and p_{jk} in Table 11 is .0121, with six deviations at least as large as .020. The largest deviation is .027. These deviations, it will be observed, occur with p_{jk} values that are not extreme, the values lying between .766 and .234. Most of the large deviations occur in the neighborhood of the two non-SST triples: (D,B,H) and (F,D,A). The Mosteller goodness of fit test performed on the θ_{jk} and θ_{jk}^* corresponding to the p_{jk} and p_{jk}^* in Table 9 gives a residual sum of squares equal to 19.39. This residual would be significant at the .01 level if there were 1648 replications on each p_{jk} . That is, it would take 1648 replications to reject the U theory at the .01 level.

Table 12 presents predictions of the U theory to the p_{jk} matrix in Table 8 based on an alternative, less efficient, fitting procedure. The LCJ scale values were obtained by using only the seven pairs along the diagonal of Table 8. The predicted values, p_{jk}^* , for the entire matrix are then obtained from this scale. It will be seen in Table 12 that the p_{jk} and p_{jk}^* values along the diagonal are equal, which is as expected. The reason for using this less efficient fitting procedure is to obtain some indication of the actual predictability of the theory to "new" p_{jk} values. The least square procedure based on the entire p_{jk} matrix indicates the degree to which the theory can postdict or reproduce the probabilities and this, as was shown in the previous section, can be appreciably higher than actual predictions. The alternative fitting procedure is also of some interest because of its use in the experimental part of this study. The less efficient procedure will be denoted by prediction method; the least squares procedure by reproduction method.

TABLE 12

SIMILARITY OF U AND Q THEORIES AS INDICATED BY PREDICTION METHOD

	E	F	D	G	C	B	H	A
E p*		.522	.553	.611	.631	.690	.710	.748
p		.522	.557	.614	.647	.698	.705	.771
F p*			.531	.589	.610	.670	.691	.730
p			.531	.605	.613	.657	.712	.723
D p*				.559	.580	.641	.663	.704
p				.559	.601	.658	.647	.744
G p*					.521	.585	.607	.651
p					.521	.566	.624	.633
C p*						.564	.587	.631
p						.564	.567	.663
B p*							.523	.569
p							.523	.604
H p*								.546
p								.546

Note: p* denotes a value predicted by the U theory
p denotes a value generated by the Q theory

Table 13 indicates the fit of the V theory to the data matrix in Table 8. The prediction method was also used in Table 13 so that the diagonal entries show perfect agreement. (The equations for obtaining the scale values of the V theory with the prediction method are given in Appendix 3.) The deviations of the U and V theories, implied in Table 12 and 13, are given explicitly in Table 14. It will be noted that in every case the sign of corresponding deviations in both theories agree. This confirms the previous conjecture, that the U and V theories have a similar character, at least with respect to data generated from the Q theory.

The average absolute deviations of the U and V theories according to Table 14 are .0126 and .0128 respectively. The value of SS_D for the U theory equals 29.41, which indicates that the number of replications required to reject the U theory at the .01 level is 1,086; SS_D for the V theory is 30.93 and n equals 1,033.

The discussion in this section suggests the following conclusions:

- 1) The U and V theories will be extremely difficult, if not impossible, to distinguish experimentally.

- 2) It appears more feasible to compare the Q theory with the U and V theory. This does not mean that it will be possible, with a limited number of replications, to obtain a probability level of .01 for the U and V theories and a probability level of .99 for the Q theory. However, it should be possible to detect substantial differences in the degree of fit of the U and V theories compared to the Q theory. This optimistic statement is based on the fact that the SS_D of the U or V theory in the previously cited example, while not

TABLE 13

SIMILARITY OF V AND Q THEORIES AS INDICATED BY PREDICTION METHOD

	E	F	D	G	C	B	H	A
E	p*	.522	.553	.610	.630	.688	.707	.744
	p	.522	.557	.614	.647	.698	.705	.771
F	p*		.531	.589	.609	.669	.689	.727
	p		.531	.605	.613	.657	.712	.723
D	p*			.559	.580	.641	.662	.702
	p			.559	.601	.658	.647	.744
G	p*				.521	.585	.607	.650
	p				.521	.566	.624	.633
C	p*					.564	.587	.631
	p					.564	.567	.663
B	p*						.523	.569
	p						.523	.604
H	p*							.546
	p							.546
A								

Note: p* denotes a value predicted by the V theory
p denotes a value generated by the Q theory

TABLE 14
THE SIMILARITY OF THE U AND V THEORIES WITH RESPECT
TO DATA GENERATED FROM THE Q THEORY

	E	F	D	G	C	B	H	A
E		.000 .000	-.004 -.004	-.004 -.003	-.017 -.016	-.010 -.008	.002 .005	-.027 -.023
F			.000 .000	-.016 -.016	-.004 -.003	.012 .013	-.023 -.021	.004 .007
D				.000 .000	-.021 -.021	-.017 -.017	.015 .016	-.042 -.040
G					.000 .000	.019 .019	-.017 -.016	.017 .018
C						.000 .000	.020 .020	-.032 -.032
B							.000 .000	-.035 -.035
H								.000 .000
A								

Note: upper number in each cell denotes the deviations
of the V theory; lower number, the U theory.

large enough to permit a relatively easy rejection of the two theories---about 1000 replications were needed---was nevertheless considerably greater than the SS_D of the Q theory. The latter value was, of course, zero.

It should also be worthwhile to consider the degree of fit in terms other than the magnitude of the deviations. Two curves fitted to the same set of points may have equally large deviations but nevertheless have different "profiles", one of which is similar to the profile described by the data. If, for example, a data matrix contained several triples not satisfying SST, then it should be possible to observe whether the Q theory tends to "follow" the non-SST pattern of these triples as compared to the U or V theory. The latter theories, fitting a monotonic function to these triples, cannot follow the non-SST pattern, although they may do as well as the Q theory in minimizing the deviations.

CHAPTER V

REVIEW OF THE EXPERIMENTAL LITERATURE ON COLOR PREFERENCES

With this chapter the experimental aspects of the study begin. It will be recalled that one purpose of the experiment is to test the probabilistic Q theory. For this reason the approximate method of solving for the Q theory parameters was discussed in the preceding two chapters. The second aim of the experiment is to test the independence assumption as described in Chapter I.

The preferential choices studied in the experimental part of this study involve choices of color. It is pertinent, therefore, to review in this chapter certain features of the previous studies of color preferences.

Early studies of color preference

Early experimental studies of color preferences were largely concerned with establishing the "true" order of aesthetic values for each color or in some cases establishing the "golden" colors. The results of these experiments, as reviewed by Pressey (1921) and later by Chandler (1934), were essentially negative or inconsistent. Not only did subjects disagree amongst each other, concerning the true value of each color, but their own preferences changed from time to time.

To some extent the negative findings of the early studies, as indicated by Norman and Scott (1952), can be attributed to their failure to control or specify the exact character of the colors used

and their neglect to treat the data statistically. The former defect would explain to some extent why one experimenter reports a predominant number of subjects whose most preferred color was red, and another experimenter reports only few preferences for red. The two experimenters may of course be using different reds, one highly saturated and the other unsaturated. The second defect, the lack of statistical methods, tended to obscure subtle regularity or consistency of the data that might not be apparent to the experimenters on inspection.

There is still another major defect with these early studies and to some extent with the recent ones as well; they ignored individual differences. In their attempt to study color preferences scientifically, the early experimenters apparently felt it necessary to assume that all persons had the same or very similar color preferences. To assume that people differed from one another and that they changed over times was, in effect, to invite chaos. The laws of color preference, like the law of gravity, had to be universal. The negative and inconsistent findings of the early experimenters were therefore looked upon with despair.

Recent studies of color preference

One of the first series of color preference studies to control the experimental stimuli and obtain systematic results were the Guilford experiments (1934, 1939, 1940). The 316 colors used by Guilford were specified in terms of the three Munsell dimensions, hue, chroma (saturation) and value (brightness). Each subject rated each color on a ten point scale of preferability, and the average rating of each color was used to plot isochedon lines in the Munsell color space. The isochedon lines, defined by Guilford as lines of equal

preferability or equal average rating, were found to be approximately concentric circles in a value-chroma plane; circles of larger radii denoted more preferred colors.

The preference behavior of the average subject, as indicated by the isochedon lines, stands in direct opposition to that expected from the Q theory. The average subject in these experiments acts as though he had a non-ideal point. His criterion of preference is that a color be maximally dissimilar to the non-ideal color, but all colors equally dissimilar to the non-ideal color are equally acceptable. It is as if the average subject knew what color he disliked most but was relatively unconcerned about his first choice. If the preferential choices of the average subject are typical choices of individuals, then the application of the Q theory to color preferences would be inappropriate. However, the choice pattern of the average subject may result from certain artifacts and does not necessarily reflect individual choice behavior.

Consider the following example. Assume that there are two types of individuals, those preferring bright, highly saturated colors and those preferring shaded, unsaturated colors. If the relationship between saturation and ratings of preferability produced a concave, monotonic curve for each type of subject (as illustrated in Figure 4), the average ratings of these subjects would exhibit the same pattern of choices as the average subject in the Guilford experiments. In this case, the preference pattern of the average subject does not reflect the preference pattern of either type of subject. In fact, in terms of the Q theory, the average subject will be a typical subject only if the distribution of ideal points is unimodal. It is pertinent then to

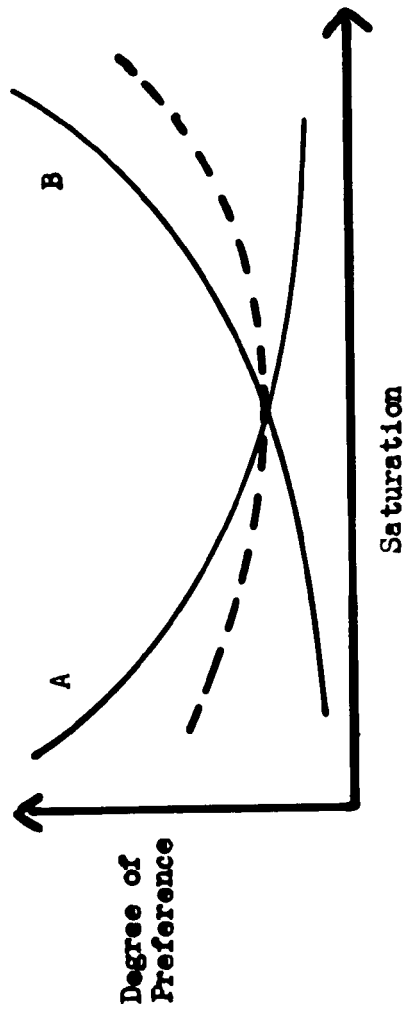


Fig. 4 - Example of two types of observers. Observer A prefers low saturated colors while B prefers highly saturated colors. (Hue and brightness are assumed constant.) The average or standard observer, represented by the dotted line, prefers both highly saturated and low saturated colors.

ask whether there are in fact two types of subjects, or, more generally, whether the distribution of ideals is unimodal. The experimental results of Eysenck (1941a) have some bearing on this question.

The study of Eysenck employs a factor analytic method known as the Q technique. Each subject in this experiment ranked the set of colors with respect to their preferability and the correlations which were factored, were obtained by correlating the rankings of pairs of subjects. (A high positive correlation indicates that the two rankings are quite similar.) A factor analysis of these correlations presumably uncovers the attributes or dimensions which determine the color preferences of the observers.

The first factor of Eysenck's analysis is, as customary, a general factor; it accounts for whatever agreement exists between subjects. It also correlates with a general aesthetic ability factor found in a previous study (1941b). The second factor is more interesting. It is a bipolar factor which divides the subjects into those who preferred pure saturated colors and those who preferred unsaturated colors. The existence of this factor suggests that the distribution of subjects in Guilford's experiments may very well have been bimodal, and this distribution, it was shown, could explain the anomalous behavior of the average subject.

Eysenck has also re-examined some of the older studies reporting negative findings. By using factor analytical methods, Eysenck finds considerable regularity in their data and rather close agreement with his own findings.

Unfortunately the trend started by Eysenck's studies (the use of more sophisticated statistical methods and explicit treatment of individual differences) does not seem to have persisted. The recent extensive research of Granger (1955) carried out in Eysenck's laboratory is in the tradition of the older studies. In these studies the experimenter attempts to demonstrate once again that individuals have substantially the same color preferences. The rankings of the subjects are analyzed by Kendall's w , the coefficient of concordance, and the conclusion is drawn that the rankings of the 50 observers are more compatible or consistent with each other than one would expect if the observers had acted entirely randomly. Kendall's w however, is a very weak test of consistency. If, for example, the subjects divided neatly into two classes as suggested by Eysenck, Kendall's w would nevertheless indicate large agreement between subjects, that is, significant departure from randomness.

The average rankings reported by Granger are at variance with Guilford's results. Granger finds that the preferences of chroma with other variables held constant, are ordered as follows.

Increasing chroma:	/2	/4	/6	/8	/10
Rank	5	3	2	1	4

These results can be interpreted in terms of the Q theory if chroma 8 is identified as the ideal. The non-ideal color and bimodal distribution, introduced to interpret Guilford's data, are not called for with these data.

Implications for the Q theory

In Chapter III it was shown that group data could easily be

treated by equation A if the distribution of C_i over individuals was assumed to be normal. However, even if the histogram of subjects on the J scale indicates that the C_i distribution is bimodal, equation A can be employed. In this case, it is necessary to fit the bimodal histogram with two normal distributions and apply equation A to each distribution separately. The final p_{jk} values will then be a weighted average of corresponding pairs of p_{jk} terms. There is, of course, the question of whether the bimodal distribution can be fitted by two normal curves, but this can be tested, by the usual goodness of fit procedures, before the analysis involving equation Q is carried out.

Unfortunately, the U and V theories are not as flexible as the Q theory in treating group data. Although in principle the ICJ method could be carried out in two parts if the distribution of U_j over individuals were bimodal, there are practical difficulties. Unlike the C_i variable in the Q theory, a histogram of the U_j variable cannot be plotted before the application of the ICJ equations. The parameters of the two normal distributions composing the bimodal distribution, therefore, cannot be determined.

In order to carry out the ICJ analysis of the experimental data in the subsequent chapters, it will be necessary to assume, without access to confirming evidence, that U_j is normally distributed over subjects for all S_j .

The assumptions which must be made for the V theory to be applicable to group data are discussed further in the following chapter.

CHAPTER VI

EXPERIMENTAL METHOD

Stimuli

The stimuli used in the experiment were Munsell standard colors (Munsell, 1945). The Munsell designations of the stimuli and their tristimulus values are given in Table 15.

It will be noted that the colors lie on a diameter perpendicular to the brightness axis in the Munsell space. The line lies in the brightness 4 plane and contains the two complementary hues, 5BG and 5R (a blue-green and a red hue). Since the line passes through the origin in the constant brightness plane, a neutral gray is also included. It was necessary to employ a one dimensional continuum of colors to test the Q theory, since the mathematical development in Chapter III pertains exclusively to the one dimensional case.

There were two criteria for selecting the colors on this line. It was, first of all, necessary to select colors which were not easily distinguishable. Since the Q theory being tested is a probabilistic theory, some inconsistent judgments must be obtained and this requires having somewhat indiscriminable stimuli.

The second criterion concerns the range of stimuli. If most persons are unilateral to all the stimuli, then the distinction between the Q theory and the U theory is minimized. In order to provide an adequate test of the Q theory, the difference between the two

TABLE 15
DESIGNATIONS OF THE STIMULI

Munsell Designations ^a	Abbreviated Designation	Tristimulus Values		
		x	y	z
5R 4/8.3	R8	.1591	.1060	.0660
5R 4/6.2	R6	.1752	.1280	.0969
5R 4/4.2	R4	.1420	.1139	.0989
5R 4/2.0	R2	.1328	.1217	.1250
5R 4/1.0	R1	.1268	.1218	.1364
5R 4/	R	.1201	.1232	.1437
5BG 4/6.0	G6	.0775	.1140	.1645
5BG 4/4.0	G4	.0979	.1263	.1731
5BG 4/2.0	G2	.1122	.1293	.1633
5BG 4/1.0	G1	.1099	.1180	.1430

^aThe saturation values were obtained by interpolation from the tables given in Newhall, Nickerson, and Judd, 1943.

theories should be maximized and this can be accomplished by selecting a range of stimuli wide enough to include the ideal color (the most preferred color) of most subjects. For the red colors, preliminary experimental results indicated that a saturation of 8 was high enough to satisfy the second criterion; the ideal red color on this line for most persons was between a zero saturated red (a neutral gray) and a red of saturation 8. Unfortunately, the highest saturation obtainable for the blue-green colors is 6 and preliminary experimental results suggested that this saturation was not sufficiently high to fully meet the second criterion. Although most ideal points in the preliminary experiment clustered about a saturation of 4, there were, nevertheless, a significant number of preferences for a saturation of 6.

Two hues were selected instead of one (as is customary in recent scaling studies) to provide a more stringent and interesting test of both the independence assumption and the Q theory. Assume for example, that a subject selects as his most preferred color on the 5BG-R line a blue-green hue with a saturation of 6. According to the Q theory and assuming the validity of the independence assumption---in this case, that the Munsell space is the appropriate discrimination space for preferential choices---this subject should prefer all blue-green hues on this line to any red on the line. However, from a priori considerations this seems unreasonable. It seems unreasonable to assume that the subject would prefer the most unsaturated, washed out blue-green (in the limit, a neutral gray) to any red, in particular to a moderately saturated red.

Two alternative explanations appear possible. One could reject the Q theory or perhaps modify the Q theory to include separate ideal points for different hues. This explanation, aside from being cumbersome to deal with, does not readily account for choices between different hues. A second alternative explanation is to assume that the Munsell space is inappropriate for preferential choices involving more than one color, that is, to reject the independence assumption. This means that the subject's behavior could be explained by maintaining the Q theory in its simple form (one ideal point for all colors) but drastically revising the color space. If, for example, the subject chose saturated colors in preference to unsaturated colors, irrespective of the hue, then the appropriate color space would be obtained by merely folding the Munsell line at the origin onto itself. The folded line would then represent the J scale upon which the subject would be located.

If only one hue were used in the experiment the independence assumption and the Q theory would be more easily verified since the complications described above could not occur.

Subjects

The subjects were 64 undergraduate women. Those recruited from the elementary psychology course were unpaid subjects; others, recruited at large, were paid. Fourteen of the subjects participated in the pretest and of the 50 who participated in the main experiment one was rejected because of inadequate color vision as indicated by the Farnsworth-Munsell 100 Hue Test (Farnsworth, 1957).

Apparatus and viewing conditions

The stimuli were presented pairwise using the display board illustrated in Figures 5 and 6. This display board consists of two pairs of openings, one pair on the movable panel which is in front of the display board and one pair directly behind the panel on the board itself. The openings on the board are 3 by 5 inches while the openings on the sliding panel are $1\frac{1}{2}$ by 5 inches. The display board was painted a light gray having a reflectance, according to a Macbeth illuminometer, of .48.

A pair of colors inserted from behind the display board are viewed by the subject when the panel slides across the face of the board. The motion of the panel is controlled by sliding a knob which is behind the display board. As the experimenter moves the knob to the right or left the panel moves in the corresponding direction with the same speed as the knob. When the panel, for example, moves continuously from left to right, the stimuli appear in rapid succession, first the right stimulus and then the left. When the sliding panel moves in the reverse direction, right to left, the stimuli are presented in reverse order also. Each stimulus appears when the opening on the sliding panel coincides with the opening on the board. The approximate time of exposure of each stimulus was .30 seconds.

The experimental room was windowless and had black walls and ceiling. The illumination was provided by three Macbeth Examolite fixtures, which provide light approximating north skylight (7300 K), placed

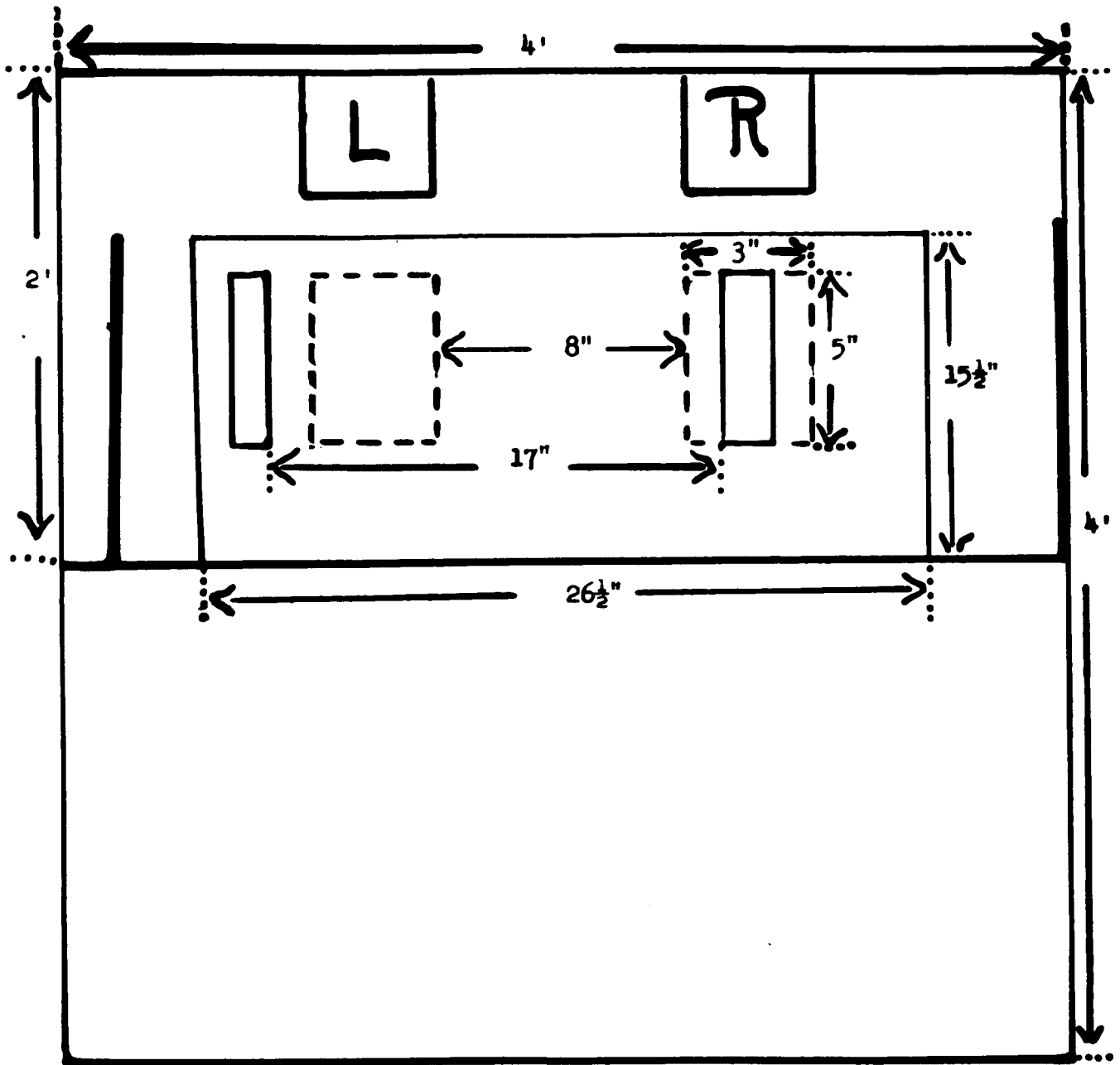


Fig. 5 - The display board.

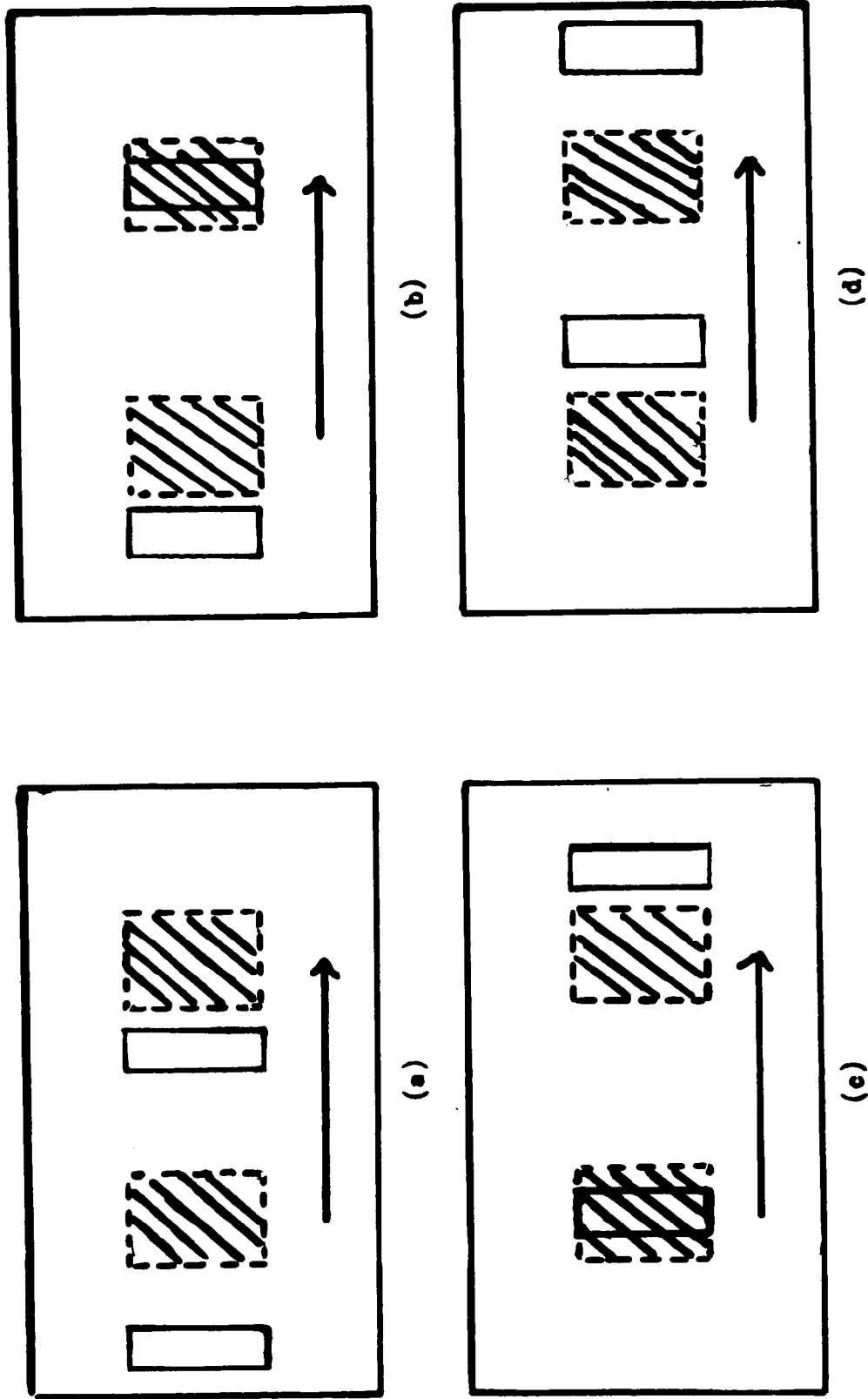


Fig. 6 - Successive positions of the panel with respect to display board. (a) Neither stimulus is visible to subjects. (b) Right stimulus is visible to subjects. (c) Left stimulus is visible to subjects. (d) Neither stimulus is visible to subjects.

directly above the display board. The board was set at approximately 45 degrees and received an illumination of 69 ft. candles. The subjects sat between $8\frac{1}{2}$ to 10 feet from the colors presented on the board so that the angle subtended by each color was about .8 degrees in the horizontal direction and 2.7 degrees in the vertical direction. The reason for using a relatively small viewing angle and rapid exposure was to increase the number of inconsistent judgments. Preliminary tests indicated that the subjects were generally remarkably consistent when the viewing angle and exposure were somewhat greater.

The Munsell papers were cut to 4 by $5\frac{3}{4}$ inches and mounted on illustration board. Since openings of the display board were 3 by 5 inches, there was sufficient margin to permit easy handling of the non-viewed portion of the stimulus. Easy handling of the stimulus cards was also facilitated by the use of a box having separate compartments for each stimulus card. The edges of each compartment were covered with velvet to minimize scratching the surface of the stimulus cards when the cards were put in or taken out of the compartment. To minimize the possibility of identifying a stimulus by certain distinguishing marks which might appear on the stimulus cards, each stimulus was represented by more than one card. Also any card which had inadvertently become marred or defective through handling was replaced.

Procedure

Between three and seven subjects participated in each experimental session. The initial task given each subject was a color vision test

known as the Farnsworth-Munsell 100 hue test for the examination of color discrimination (Farnsworth, 1957). This test consists of four sets of color chips varying primarily in hue. The subject's task is to arrange the colors in each set so that they form a graduated series. The score on the test is determined by the number of errors the subject makes, an error being defined as the number of pairs which are incorrectly ordered by the subject. The correct order of the colors is easily determined by the numbers written on the reverse side of the color chips.

The subjects were told that the purpose of the test was to determine how well they could see colors. They were told to take their time and when finished to turn over the color chips and determine their own score. After a subject had recorded her own score, the set of colors was returned to the experimenter who, after randomizing the order of the colors, presented the set to another subject. In this manner it was possible to test all subjects in each session in about a half hour.

In the color preference part of the experiment the ten pairs of stimuli were presented randomly, each pair replicated three times. There were, in addition, five practice trials in which all of the colors appeared in at least one pair. After the 140 pairs were presented at the rapid rate of exposure, eleven additional pairs were presented with longer exposure times of about 3 seconds. The eleven pairs were: G1 G2, G2 G4, G4 G6, R1 R2, R2 R4, R4 R6, R6 R8, G1 R1, G2R2, G4 R4, G6 R6. The reason for presenting these eleven pairs was to determine whether preferential choices at the rapid rate of exposure is compatible

with choices made under more usual exposure times.

Since the order of presentation of the colors was random, the three replications of each color were randomly dispersed over the trials so that in fact certain replications appeared consecutively to the subject. The position of each stimulus was also randomized, being on the left or right an equal number of times. Also, the three replications of each pair involved different positions of the two stimuli comprising the pair.

The trial number which was called out about a second before the colors were exposed, acted as a ready signal for the subject. The subjects indicated their choices on a data sheet containing the trial numbers and the letter L and R (for the left and right colors) adjacent to each number.

A five minute rest period was introduced after about half of the pairs were presented. The exact point at which the rest period began was varied in different experimental sessions. This was done because it was observed that subjects tended to have difficulty with the first few trials after the rest period. A few trials were needed to warm up again. Varying the location of the rest period, therefore, had the effect of varying the stimuli which occurred during the brief warm up period.

Instructions to the subjects

The instructions given to the subjects after the color vision test was completed were stated as follows.

In this next part of the experiment I am going to present a pair of colors on this board in front of you. For each pair

of colors. I want you to imagine that you are in a paint store selecting a color of paint to be used for your entire dormitory room (or living room, if you live at home). Select the color in each pair which you would prefer for your room. Be sure to keep in mind that you are selecting a color for your room, not a color which you may prefer for other reasons. Remember also that the color you select is to be used to paint the entire room, not merely one wall or two. Also, the color is not to be part of a pattern or **combination** of colors.

Indicate your selection by placing an 'x' over the 'R' or 'L' which appears on the sheet before you. In some instances you will probably dislike both colors, or perhaps, you will like both colors; however, for each pair you should try to select the color which appears the more preferable.

When this panel is moved across the board the colors will appear in rapid succession, one under the letter 'L' and one under the letter 'R'. If the panel moves to the left like this (panel is **moved** to the left) then the first color you will see will appear under the 'L' and the second color shortly thereafter will appear under the 'R'. The colors will appear rather rapidly, somewhat less than a second, so that you will have to look very quickly from one side to the other. You will have some difficulty in the beginning seeing the colors, but after a while it will probably be quite easy. In any case, please be sure to indicate a choice for every pair. If necessary, guess. Do not worry about being inconsistent from one pair to another.

The presentations are rapid enough so that you are bound to make some mistakes.

Although the colors will appear for less than a second, it will take about 10 seconds to change the colors. During this time feel free to talk to your neighbor or study. Please refrain, however, from talking about the experiment. Let me emphasize that we are interested in your color preferences, no matter how different they may be. In fact, we expect people to differ.

The first five trials will be practice trials. Indicate your choices on these trials with the appropriate 'x'. After these trials, I will answer any questions that you may have and we will proceed with the experiment.

After the first five trials, questions were answered, and the subjects were told that there would be about 150 trials and that it would take about a half hour, including a five minute rest period.

On trial 141 the subjects were told that the following few pairs would be presented quite slowly and, without additional pause, they were begun. After the 151 trials were complete, the subjects were instructed to turn over their paper and to describe on the back of it what system, if any, they had used in making their choices.

Treatment of the data

The U theory.- The LCJ solution is applied to a p_{jk} matrix which is obtained from the combined replicated judgments of all subjects, that is,

$$(1) \quad \frac{n_{jk}}{Nm}$$

where n_{jk} is the total number of judgments of all subjects in which S_j is preferred to S_k . In the experiment, $N = 49$ and $m = 3$ so that (1) reduces to

$$(2) \quad \frac{n_{jk}}{147} .$$

There are two properties of (2) which require some further discussion: its expected value and its sampling variance. If (2) is to be analyzed by the LCJ equation, its expected value should equal a preference probability that satisfies the U theory. This remains to be established. It is also necessary to know the sampling variance of (2) in order to perform a goodness of fit test. Both properties of (2) will be considered after the relevant assumptions have been stated for the group LCJ analysis.

The assumptions for an LCJ analysis of group data are analogous to those stated for the Q theory in Chapter III. However, it will be well to state them explicitly here. It is assumed that for each individual i and stimulus S_j the distribution of $u_{j,i}$ is

$$(3) \quad N(U_{j,i}, \sigma_u),$$

and the distribution of $U_{j,i}$ is

$$(4) \quad N(U_j^!, \sigma_u).$$

Since **these** two distributions are independent, the distribution of u_j is therefore

$$(5) \quad N(U'_j, \sqrt{\sigma_u^2 + \sigma_a^2}).$$

If the usual correlations are set equal to zero,

$$r_{iU_j, iU_k, i} = 0$$

$$r_{u_j, iu_k, i} = 0,$$

which implies that

$$(6) \quad r_{u_k u_j} = 0,$$

and if

$$(7) \quad \sqrt{\sigma_u^2 + \sigma_a^2} = 1$$

then the LCJ equation results for the group data:

$$(8) \quad X_{jk} = U'_j - U'_k.$$

The normal deviate in (8) corresponds to the p_{jk} based on the entire group.

If $p_{jk, i}$ is defined as the preference probability for a single individual, then the group p_{jk} is merely the expected value, over subjects, of $p_{jk, i}$, that is,

$$(9) \quad p_{jk} = E(p_{jk,i}).$$

Equation (9) indicates that if the subjects are selected randomly, then the proportion of preferences of S_j to S_k for $m = 1$ ---that is, n_{jk}/N ---will be an unbiased estimate of p_{jk} . Furthermore, equation (8) asserts that this p_{jk} will satisfy the conditions of the U theory.

Consider now the case of replicated judgments. The question is whether (1) is an unbiased estimate of the group p_{jk} as defined in (9), that is, whether

$$(10) \quad E\left(\frac{n_{jk}}{mN}\right) = p_{jk}.$$

To prove (10) it is simpler to assume, without loss of generality, that the population from which the subjects are drawn is large but finite. The group p_{jk} can then be defined as

$$(11) \quad p_{jk} = \frac{1}{N^*} \sum_{i=1}^{N^*} p_{jk,i}$$

where N^* is the size of the population. In a sample of size N , the probability of selecting a given subject is N/N^* . For m replications per subject, the expected number of preferences of S_j to S_k for individual i is $mp_{jk,i}$. Therefore, the expected number of preferences of S_j to S_k for a sample of size N is

$$(12) \quad E(n_{jk}) = \frac{N}{N^*}(mp_{jk,1}) + \frac{N}{N^*}(mp_{jk,2}) + \dots + \frac{N}{N^*}(mp_{jk,N^*})$$

which can be simplified to

$$(13) \quad E(n_{jk}) = \frac{Nm}{N^*} \sum_{i=1}^{N^*} p_{jk,i}.$$

Using (11), (13) reduces to

$$(14) \quad E(n_{jk}) = Nmp_{jk}.$$

Dividing (14) by Nm , to obtain the expected proportion, (10) is obtained, which was to be proved. It is appropriate, therefore, to use (2) as the basis for the LCJ analysis.

The second property of (2) to be discussed is its sampling variance. The problem is to determine whether the sampling variance of (2) is consistent with the error term in the Mosteller goodness of fit test described in Chapter V. This test assumes that the error variance of the estimate of p_{jk} is the usual binomial variance

$$(15) \quad \frac{p_{jk}(1-p_{jk})}{n_{jk} + n_{kj}}$$

and that the estimate of p_{jk} is independent of the estimate of p_{jl} . However, both assumptions can be questioned if the LCJ analysis is based on (2). Bock (1958) has shown that even when $m=1$ for group data and the correlation indicated in (6) is zero, the normal deviates X_{jk} and X_{jl} will, nevertheless, have a non-zero correlation. (The correlation between deviates not sharing a common stimulus is zero.) This non-zero correlation contradicts the assumption of independence in the Mosteller tests, although the first assumption described above is satisfied for non-replicated group data; that is,

the sampling variance of n_{jk}/N for group data is equal to (15) (see Cramer, 1951, p. 192). When m is greater than one, however, (15) is not the correct sampling variance of (2).

It is well known that a two stage sampling procedure---and this is what is involved in using replicated judgments---results in a higher sampling variance than that obtained from simple random sampling (see Cochran, 1953). This means that the binomial variance (15) is an underestimate of the sampling variance of (2). The correct estimate in a two stage sampling procedure involves estimating the variance of $p_{jk,i}$ between subjects as well as within subjects, and in the present case, these variances are difficult to determine. The analysis by **Bock** however, indicates that the binomial variance is an over-estimate of the sampling variance, and unfortunately the correct value is not readily estimated from his results. Since both types of biases are involved in using (15) as the sampling variance of (2), it is not clear whether (15) is an overestimate or underestimate of the correct expression. However, since the primary purpose of the experiment is to compare the fit of various theories, rather than to reject or accept a null hypothesis, it will probably be sufficient to use the Mosteller test without further modifications to obtain a measure, albeit an approximate one, of the degree of fit of each theory.

V theory analysis. - The assumptions which must be made for the V theory to be applicable to group data are not as easily determined as the corresponding assumptions for the U theory. The problem is as follows. Given that the preference probabilities for each subject

satisfy the V theory, what additional constraints must be placed on these probabilities so that the preference probabilities for the group will also satisfy the V theory. For the U theory it is sufficient to assume, in addition to the equal correlation and equal variances assumption, that the distribution of the scale values of each stimulus is normally distributed over subjects.

The analogous assumption for the V theory restricts the function $\phi(y;z)$ as follows:

$$(16) \quad \int_{-\infty}^{\infty} \phi(y;z) dy \int_{-\infty}^0 \text{sech}^2 g(x;y) dx = \int_{-\infty}^0 \text{sech}^2 g(x;z) dx.$$

If such a function exists and if it has the properties of a frequency function, then it would be sufficient to assume that $v_{j,i}/v_{k,i}$ was distributed over subjects with the function ϕ as its density function. This assumption would guarantee that if the $p_{jk,i}$ of each subject satisfied the V theory, then the group probability, p_{jk} , would also satisfy the V theory. Since it is not known whether the function ϕ exists, the experiment will merely test whether the group p_{jk} satisfies the V theory.

The Q theory. - The normal distribution assumptions which make the Q theory applicable to group data have been discussed in Chapter III and Chapter V. It remains to discuss here the problem of comparing the Q theory results to the U and V theories.

This comparison is complicated by the fact that the Q theory has two more free parameters than either the U or the V theory. Consequently, if all the data were used in the Q theory analysis, then one would expect the Q theory to predict better than the other two theories purely on the basis of the additional parameters. However, the approximate method of solution does not use all the data. Under certain circumstances it may use only the pairs adjacent on the J scale to determine the Q values of the stimuli, and even these pairs will not, in general, involve the judgments of all the subjects.

Since the Q theory analysis does not use all the data to fit the parameters in the theory, it cannot be argued a priori that the Q theory has an advantage over the other two theories. In fact, it appears quite likely that the Q theory is at a disadvantage to the U and V theories when the analysis of these two theories is based on the reproduction method, that is, when these two theories use all the data. One is tempted to put the three theories on an equal footing by using the prediction method for analyzing the U and V theories, but it cannot be said that this is an equitable solution either. While the stimulus values of the Q theory may be obtained from adjacent pairs only, the values of the mean and variance of the distribution of ideals is estimated by obtaining the I scale of each subject, and this entails using all the judgments from each subject. However, since the I scale involves only ordinal relations, little of the information in the data is actually used. It cannot be said, therefore, that a Q theory analysis using adjacent pairs to determine the stimulus values, is at the same level as the prediction

method of U and V theories. However, it also cannot be said that the Q theory analysis uses as much data as the reproduction method. Very likely the Q theory analysis lies somewhat between the prediction method and the reproduction method of the U and V theories. For this reason, both the reproduction and prediction method are used in the analysis. It will be assumed that a residual sum of squares (SS_D) of the Q theory which is equal to or less than the SS_D obtained from analyzing the U or V theory by the reproduction method is evidence for the superiority of the Q theory. If the SS_D of the Q theory exceeds the SS_D of the U or V theory analyzed by the prediction method, the Q theory will be considered inferior to the U or V theory. The third possibility is indeterminate; in this case the SS_D of the Q theory would lie between the residual sum of squares computed by the reproduction method and the prediction method.

CHAPTER VII

EXPERIMENTAL RESULTS

Color vision test

The results of the Farnsworth-Munsell Test indicated that the color discriminations of most subjects were exceedingly precise. According to the manual for this test (Farnsworth, 1957) about 16% of the population (exclusive of color defectives) have been found to make four or less reversals on this test. About 76% of the 50 subjects in the experiment did this well. One subject, however, did considerably worse than this. Although her score did not indicate seriously defective color vision, this subject was excluded from further analyses. The average score for the remaining 49 subjects was 2.85 and the largest error score in this group was 10.

Inconsistency of preferences

The number of inconsistent stimulus pairs (an inconsistent stimulus pair defined as one judged inconsistently on the three replications of the pair) was used to determine whether the subjects were responding in a purely random fashion. If a subject were behaving randomly, the probability of giving three consistent responses to a particular stimulus pair would be one-fourth. Of 45 pairs, the expected number of inconsistent pairs for random behavior, is therefore 33.75; a two sigma interval about this expected value is 28.2-39.4. For three subjects the number of inconsistent pairs was large enough to fall in this interval; they had 29, 30 and 33 inconsistent pairs. These

subjects, however, were not omitted in the subsequent analyses because it was clear from inspecting their I scales that they were not responding randomly. All three subjects had I scales which were perfectly compatible with the obtained J scale described in the subsequent analyses; the I scales could be conceived as a folded J scale.

Figure 7 gives the histogram of inconsistent stimulus pairs for all subjects. The most consistent subject had 4 inconsistent pairs; the average for the 49 subjects is 18.2. In Figure 8 the number of inconsistent stimulus pairs is presented separately for unilateral and bilateral subjects. In this case, unilateral subjects refer to those subjects located near either end of the J scale and bilateral subjects those located near the middle of the J scale. The I scales and the J scale which were used to make this dichotomy between subjects are described in a later section. Here it will suffice to indicate that the most preferred stimulus of the unilateral subjects was one of the following:

N, R1, G1, G6, R6, or R8

and the most preferred stimulus of the bilateral subjects was either

R2, R4, G2, or G4.

One subject is not indicated in Figure 8 because the response pattern of this subject deviated excessively from the J scale.

The reason for separating the unilateral and bilateral subjects was to test certain implications of the Q theory. According to the Q theory, bilateral subjects should be more inconsistent than unilateral ones. This is a consequence of the fact that a bilateral person has to discriminate between stimuli on a folded J scale, and this scale will

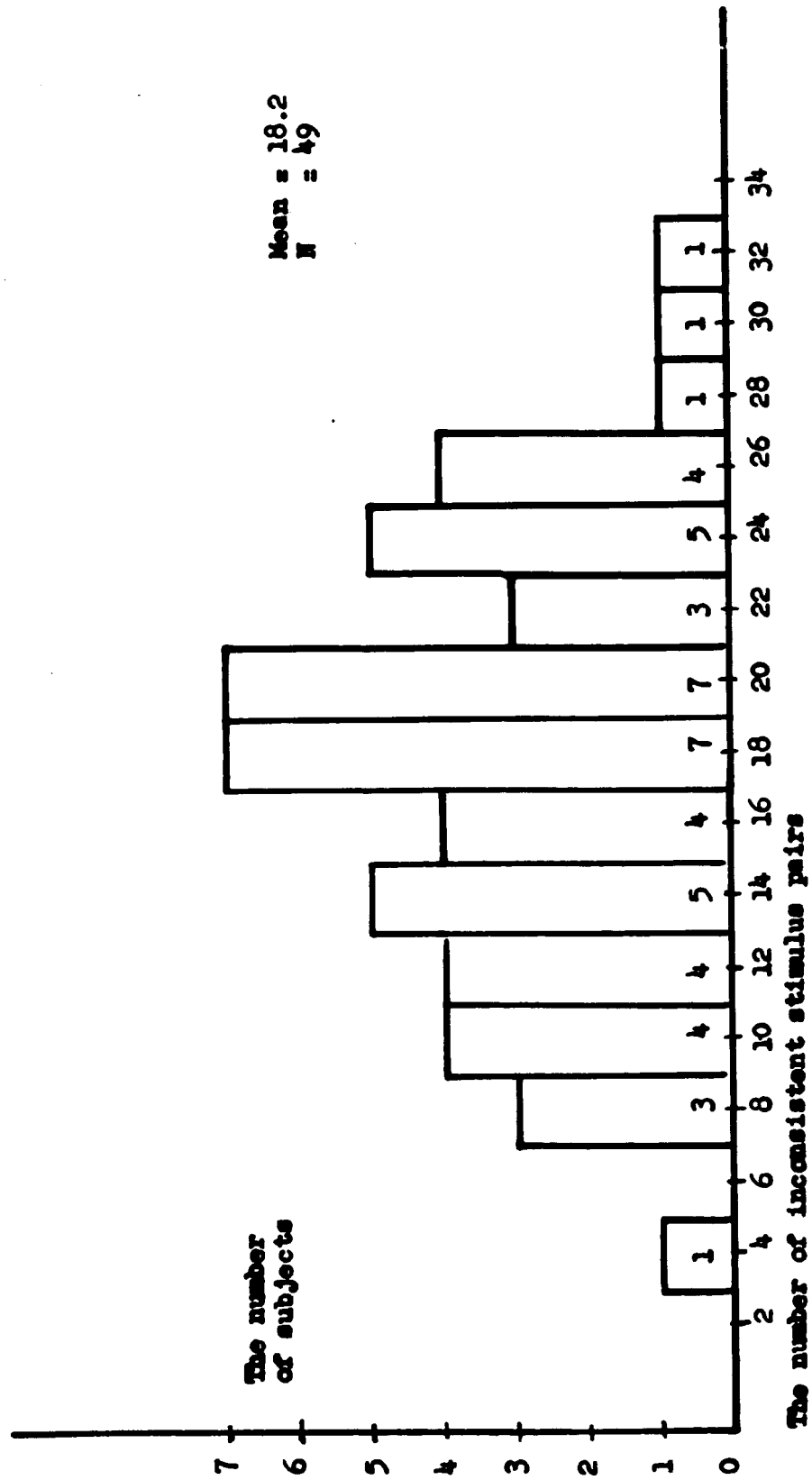
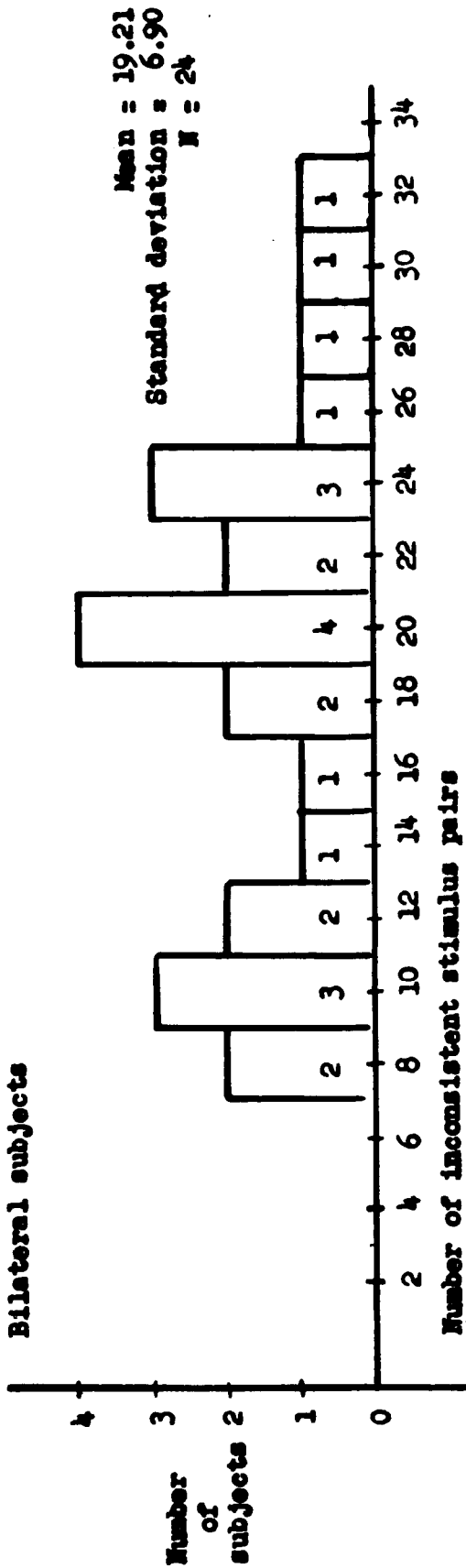


Fig. 7 - The number of inconsistent stimulus pairs for all subjects.



Unilateral subjects

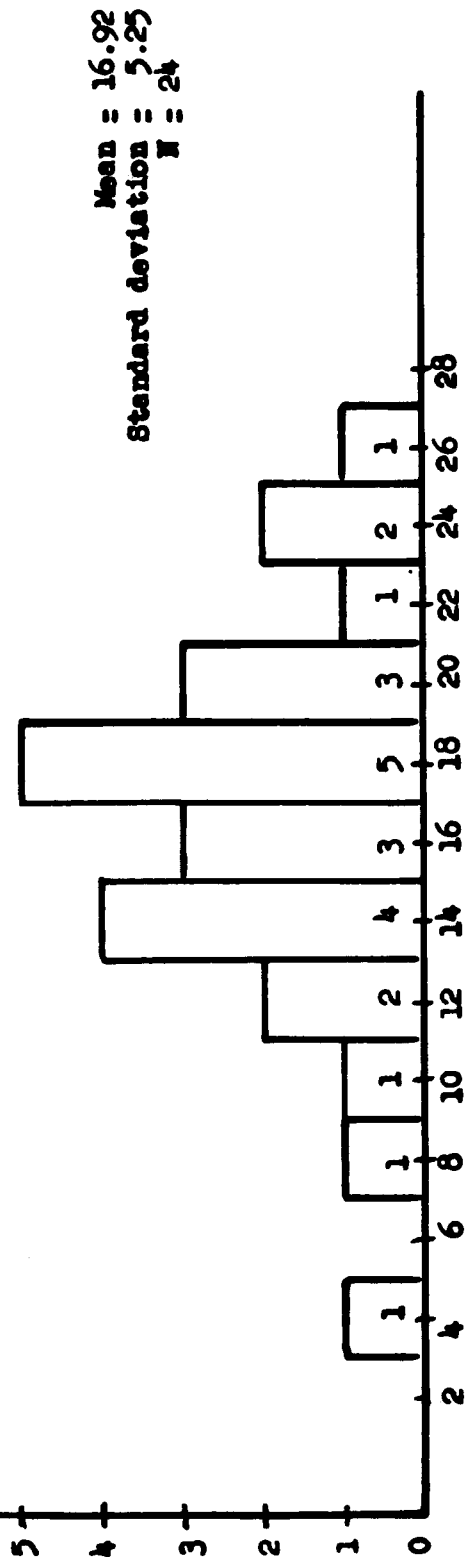


Fig. 8 - The number of inconsistent stimulus pairs for unilateral and bilateral subjects

necessarily have smaller intervals between the stimuli than a non-folded scale. The unilateral person does not fold the scale---or at least, not very much---and therefore has larger intervals to discriminate. The discrimination judgments of the unilateral subject should also tend to be less influenced by the variability of their ideal point, and this too should tend to make the unilateral subject more consistent than the bilateral subject. The data in Figure 8 indicate that the mean number of inconsistent pairs for bilateral subjects (DB) is 19.21 and for unilateral subjects DU is 16.92; the corresponding standard deviations are 6.90 and 5.25, respectively. DB is larger than DU as expected, but their difference is not significant at the 5% level. The statistic y (see Pearson and Hartly, 1954, pp. 26-27) is equal to .70 while the value of this statistic at the 5% level (using a one tailed test) is approximately 1.68. (This test of the difference between means does not assume equal variances.)

Relation between color vision and inconsistency of preferences

One of the assumptions implicit in the Q theory is the assumption that the variance of q_j and c_j is equal for all subjects. This assumption implies, among other things, that the subjects have essentially the same color discriminations. This can be tested by comparing the scores on the color vision test with the number of inconsistent preference judgments. If the subjects had, in reality, different color discriminations, and if the color vision test was effective in measuring these individual differences, then one would expect a correlation between the color vision score and the number of inconsistent stimulus pairs. To be more precise, it must be added that it is necessary to control for the different points at which subjects

may fold the J scale, since the discriminability of the stimuli is a function of C_1 . The degree of fold was controlled to some extent by considering unilateral and bilateral subjects separately.

The correlations between the color vision score and the number of inconsistent stimulus pairs for unilateral and bilateral subjects are as follows:

$$r_{\text{unilateral}} = .36, p = .08$$

$$r_{\text{bilateral}} = -.28, p = .18$$

(The probability values were determined by using Fisher's z transformation and a two tailed test.) Neither correlation is significantly different from zero at the 5% level and only one correlation, the unilateral correlation, has the expected sign. The homogeneity of the group and the assumption of equal variance therefore cannot be rejected.

LCJ analysis of the combined set of stimuli¹

Reproduction method (U1).-Table 16 gives the p_{jk} matrix based on all the judgments of all subjects. In principle, each proportion in this matrix should be based on 147 judgments but since certain subjects did not respond on some trials, six proportions are based on 146 judgments and two on 145. However, in performing the goodness of fit test it will be assumed that all proportions are based on 147. The error committed in making this assumption will be seen to be negligible.

Table 17 contains the p_{jk}^* values reproduced from Table 16 using the LCJ solution. (This procedure is called the U1 analysis.) Table 20 contains the Mosteller goodness of fit test and, in addition, analyses developed by Gulliksen and Tukey (1958), which are described in

¹In general, in this and the subsequent chapters the terms p_{jk} , c_1 , σ_{c_1} , q_j , etc. refer to estimates of the corresponding population values.

TABLE 16

OBTAINED p_{jk} MATRIX

	G4	G6	G2	R2	R1	R4	G1	R6	R8	N
G4		.548 ^a	.578	.619	.626	.653	.641 ^b	.714	.769	.728
G6			.517	.599	.619	.692 ^a	.578	.755	.694	.664 ^a
G2				.578	.619	.646	.726 ^a	.626	.714	.721
R2					.529	.456	.599	.565	.435	.728
R1						.476	.500 ^a	.544	.605	.707
R4							.503	.531	.578	.644 ^a
G1								.503	.558	.538 ^b
R6									.571	.578
R8										.524
N										

^aProportion based on 146

^bProportion based on 145

TABLE 17
REPRODUCED p_{jk}^* MATRIX FROM LCJ SCALE (U1)

	G4	G6	G2	R2	R1	R4	G1	R6	R8	N
G4 p^*		.532	.536	.632	.652	.657	.683	.697	.725	.761
G4 p		.548 ^a	.578	.619	.626	.653	.641 ^b	.714	.769	.728
G6 p^*			.503	.601	.621	.627	.654	.668	.697	.735
G6 p			.517	.599	.619	.692 ^a	.578	.755	.694	.664 ^a
G2 p^*				.598	.618	.624	.651	.665	.694	.732
G2 p				.578	.619	.646	.726 ^a	.626	.714	.721
R2 p^*					.521	.527	.556	.571	.603	.645
R2 p					.592	.456	.599	.565	.435	.728
R1 p^*						.506	.535	.550	.582	.625
R1 p						.476	.500 ^a	.544	.605	.707
R4 p^*							.528	.544	.576	.619
R4 p							.503	.531	.578	.644 ^a
G1 p^*								.516	.548	.592
G1 p								.503	.558	.538 ^b
R6 p^*									.532	.576
R6 p									.571	.578
R8 p^*										.544
R8 p										.524
N p^*										
N p										

^aProportion based on 146

^bProportion based on 145

Table 19. In Table 19 it will be seen that the total sum of squares of the angular measure θ have been separated into two components, a residual component (SS_D) which is the basis of the Mosteller test and a linear component (SS_L) which indicates the variance attributable to the LCJ scale. The χ^2_T term in Table 19 associated with SS_T , the total sum of squares indicates the extent to which the obtained p_{jk} values deviate from one half. In Table 20 χ^2_T is significant (beyond the .001 level) so that one can reject the hypothesis that the preference proportions in Table 16 are mere random deviations about one half.

According to Table 20 χ^2_L , associated with the linear component SS_L , is also significant beyond the .001 level so that one may reject the hypothesis that the p_{jk}^* values in Table 17, the preference proportions reproduced from the LCJ scale, are random deviations about one half. The third χ^2 term, χ^2_D described in Table 19 is based on the residual sum of squares, SS_D , and this term is precisely the χ^2 term of the Mosteller test. In Table 20 χ^2_D term is significant at the .001 level indicating that the p_{jk} matrix in Table 16 cannot be assumed to have been generated by an LCJ scale. However, although the LCJ scale can be rejected at the .001 level the scale may nevertheless account for a considerable amount of the data. A χ^2 test does not give this kind of information. To provide some indication of how much is accounted for by the theory, Gulliksen and Tukey (1958) have developed an expression which may be interpreted as a reliability term; the true score variance in this case refers to the variance accounted for by the LCJ scale. Two formulas are given by Gulliksen and Tukey for calculating the reliability: r_s and r_{ss} . Both formulas are presented in Table 19. The formulas do not, in

TABLE 18

PREDICTED p_{jk}^* MATRIX FROM ICJ SCALE (U2)

	G4	G6	G2	R2	R1	R4	G1	R6	R8	N
G4 p^*		.502	.533	.546	[.548]	[.578]	.603	.636	.787	.813
G4 p		.653	.714	.619	[.548]	[.578]	.769	.626	.641	.728
G6 p^*			[.531]	[.544]	.546	.576	.601	.634	.785	.813
G6 p			[.531]	[.544]	.308	.354	.578	.524	.503	.644
G2 p^*				.513	.515	.545	[.571]	.605	.763	.791
G2 p				.435	.245	.374	[.571]	.456	.497	.578
R2 p^*					.502	.532	.558	[.592]	.752	.781
R2 p					.401	.422	.435	[.592]	.599	.728
R1 p^*						.530	.556	.590	.751	.780
R1 p						.517	.694	.619	.578	.664
R4 p^*							.526	.560	[.726]	.757
R4 p							.714	.619	[.726]	.721
G1 p^*								.534	.704	.736
G1 p								.395	.442	.524
R6 p^*									.673	[.707]
R6 p									.500	[.707]
R8 p^*										[.538]
R8 p										[.538]
N p^*										
N p										

Note: Brackets indicate the nine pairs which were used to determine the ICJ scale.

TABLE 19
VARIANCE-COMPONENTS ANALYSIS FOR PAIRED COMPARISONS

Source	df	SS	MS	χ^2	P
Total	$\frac{n(n-1)}{2}$	$SS_T = \sum_{i < j} (\theta_{ij} - 45)^2$	$T = \frac{SS_T}{df}$	$\chi_T^2 = \frac{MS_{SS_T}}{820.7}$	
Linear	n-1	$SS_L = \sum_{i < j} (\theta_{ij}^* - 45)^2$	$L = \frac{SS_L}{df}$	$\chi_L^2 = \frac{MS_{SS_L}}{820.7}$	
Residual	$\frac{(n-1)(n-2)}{2}$	$SS_D = \sum_{i < j} (\theta_{ij} - \theta_{ij}^*)^2$	$D = \frac{SS_D}{df}$	$\chi_D^2 = \frac{MS_{SS_D}}{820.7}$	

$$r_s = \frac{2(L-D)}{n_T}$$

$$r_{ss} = \frac{T-D}{T}$$

Note: Angles are measured in degrees.

TABLE 20
VARIANCE-COMPONENTS ANALYSIS FOR U1

Source	df	SS	MS	χ^2	P
Total	45	2899.722	64.438	519.38	$< .001$
Linear	9	2541.814	282.423	455.27	$< .001$
Residual	36	363.407	10.095	65.091 (64.986) ^a	.001

$$r_s = .845$$

$$r_{ss} = .843$$

^aThis value was computed without assuming that $N_m = 147$ for all stimulus pairs. Comparing the two values in this cell indicates that the assumption produces a negligible error.

TABLE 21
VARIANCE COMPONENTS ANALYSIS FOR U2

Source	df	SS	MS	χ^2	P
Total	45	2895.8	64.35	518.68	$< .001$
Linear	9	4413.9	490.43	790.59	$< .001$
Residual	36	3313.0	92.03	593.41	$< .001$

general, give identical values because the residual sum of squares expressed in terms of θ is not minimized by the usual fitting procedure.¹ The consequence of this is that the components SS_D and SS_L , are not quite orthogonal.

The values of r_g and r_{ss} based on Table 17 are given in Table 20. It will be seen that whereas the χ^2 test indicates that the LCJ scale can be rejected, the reliability of the LCJ scale as given by r_g or r_{ss} indicates that the y theory accounts for a considerable amount of the data in Table 16.

The UI scale, the LCJ scale which was used to generate the p_{jk}^* values in Table 17, is as follows.

(UI scale)

N	R8	R6	G1	R4	R1	R2	G2	G6	G4
0	.1117	.1929	.2321	.3042	.3190	.3718	.6197	.6284	.7093

It will be noted that the R stimuli on this scale have the order of a folded Munsell line. The G stimuli on the LCJ scale may also be conceived as a folded Munsell line, but the combination of the R and G stimuli on the LCJ scale does not have an order which is compatible with the entire Munsell line folded at any point. This suggests that the Munsell line, G6-R8, will probably not agree with the J scale on which the subjects are located, and therefore that the Munsell space is not an appropriate space for the entire set of preferential choices. This will be considered in greater detail when the procedure for constructing the J scale is described.

¹In fact, SS_D expressed in terms of p_{jk} is generally not minimized either. Mosteller (1951) has shown that a least squares criterion is involved in averaging the columns of the X_{jk} matrix, but this criterion is only tangentially related to the residual sum of squares $\sum (p_{jk} - p_{jk}^*)^2$.

Prediction method (U2).—Using the prediction method and the P_{jk} values of the nine pairs of stimuli adjacent on the Munsell line, the following LCJ scale resulted:

(U2 scale)

N	G1	R1	R8	G2	G6	R2	R6	R4	G4
0	.0954	.5446	.6311	.6962	.7724	.7773	.8100	.8878	.8930

While this scale does not give the same order of stimuli as the previous scale (U1) it does give an order which is consistent with a folded Munsell scale for the R and G stimuli considered separately.

Table 18 contains the predicted p_{jk}^* values based upon the above LCJ scale, and Table 21 gives the corresponding analysis of the sum of squares. The three χ^2 terms in Table 21 are each statistically significant so that it may be concluded that both the p_{jk} values and the p_{jk}^* are not random deviations about one half—this property of the p_{jk} values was already shown in the previous analysis—and that the LCJ scale does not adequately account for the obtained p_{jk} values. Comparing the residual sum of squares between the prediction and reproduction analyses, U2 and U1, indicates that the former gives a considerably inferior fit to the data than the latter; the SS_D based on the prediction method is about nine times larger than the SS_D based on the reproduction method. The reliabilities, r_s and r_{ss} , cannot be determined from Table 21 because the prediction method which is used in the U2 analysis gives a residual and linear sum of squares which is larger than the total sum of squares. Of course, this is the consequence of using such an inefficient fitting procedure.

LCJ analysis of the R stimuli

Tables 22 and 23 give the results of an LCJ analysis of the R

TABLE 22

REPRODUCED P_{jk}^* MATRIX (R STIMULI ONLY) FROM ICJ SCALE (U3)

	R4	R2	R1	R6	R8	N
R4 P^*		.506	.512	.552	.581	.657
R4 P		.544	.524	.531	.578	.644
R2 P^*			.507	.547	.576	.652
R2 P			.592	.565	.435	.728
R1 P^*				.540	.569	.646
R1 P				.544	.605	.707
R6 P^*					.529	.608
R6 P					.571	.578
R8 P^*						.579
R8 P						.524
N P^*						
N P						

TABLE 23

VARIANCE-COMPONENTS ANALYSIS FOR U3

Source	df	SS	MS	χ^2	p
Total	15	571.37	38.091	102.34	<<.001
Linear	5	377.45	75.490	67.64	<<.001
Residual	10	157.40	15.740	28.19	.001
$r_s = .523$ $r_{ss} = .587$					

stimuli only using the reproduction method (denoted by U3) and Tables 24 and 25 using the prediction method (denoted by U4). The residual sum of squares are significant in both analyses indicating that the LCJ theory does not adequately account for the preference proportions of the R stimuli. This conclusion should be tempered since, as it was pointed out in Chapter VI, the Mosteller analysis is only approximate. The two reliability values in Table 23, $r_g = .523$ and $r_{ss} = .587$, while not as high as the reliabilities for the combined stimuli in the U1 analysis, indicate nevertheless that the LCJ scale does account for at least one half of the variance.

Unfortunately, due to the difference in degrees of freedom it is difficult to tell from the χ^2 values whether the LCJ theory fits the R stimuli better than it does the combination of R and G stimuli. The reliability values also do not accurately indicate the relative difference in fit when the degrees of freedom are not equal or, as it is more usually described, when the range of talent is restricted. This characteristic of the reliability coefficient could explain why r_{ss} for U1 is greater than r_{ss} for U3 (.845 compared to .587). However, it appears reasonable to conclude that the prediction method gives a better fit to the R stimuli than it does to the combined stimuli. This conclusion is borne out by the fact that the SS_D for the U4 analysis is approximately three times the SS_D for the U3 analysis while the corresponding residual sums of squares for the U2 and U1 analyses are related by a factor of nine.

The similar residual sum of squares for the U3 and U4 analyses suggest that the corresponding scales would agree. However, the two scales, which are given below, do not appear particularly compatible.

TABLE 24

PREDICTED P_{jk}^* MATRIX (R STIMULI ONLY) FROM ICJ SCALE (U^4)

	R4	R6	R2	R8	N	R1
R4	P^* P	$\begin{bmatrix} .531 \\ .531 \end{bmatrix}$	$\begin{bmatrix} .544 \\ .544 \end{bmatrix}$	.601 .578	.624 .644	.634 .524
R6	P^* P		.513 .435	$\begin{bmatrix} .571 \\ .571 \end{bmatrix}$	.595 .578	.605 .456
R2	P^* P			.558 .435	.582 .728	$\begin{bmatrix} .592 \\ .592 \end{bmatrix}$
R8	P^* P				$\begin{bmatrix} .524 \\ .524 \end{bmatrix}$	.534 .395
N	P^* P					.510 .293
R1	P^* P					

Note: Bracket indicates the pairs which determined the ICJ scale.

TABLE 25

VARIANCE-COMPONENTS ANALYSIS FOR U^4

Source	df	SS	MS	χ^2	P
Total	15	571.37	38.091	102.34	<<.001
Linear	5	308.20	61.640	55.20	<<.001
Residual	10	494.23	49.423	88.52	<<.001

(U3 scale)

N	R8	R6	R1	R2	R4
0	.1996	.2733	.3741	.3908	.4051

(U4 scale)

R1	N	R8	R2	R6	R4
0	.0263	.0865	.2327	.2654	.3432

It will be noted that except for the positions of N in U4, both scales are folded Munsell lines.

To compare the U1 and U2 scales with the U3 and U4 scales, the U1 and U2 scales are reproduced below for the R stimuli only.

(U1 scale)

N	R8	R6	R4	R1	R2
0	.1117	.1929	.3042	.3190	.3718

(U2 scale)

N	R1	R8	R2	R6	R4
0	.5446	.6311	.7773	.8100	.8878

Except for the order of stimulus R4, the order of the stimuli on the scales in U1 and U3 agree; U2 and U4 agree in order except for the stimulus pair, N, R1. In terms of the metric relations on these scales, only the N, R1 reversal is serious since the N, R1 interval in the U2 scale is the largest adjacent interval on this scale.

V theory analysis

The analysis of the V theory proceeds analogously to the U theory analysis so that it will suffice to summarize the results without further discussion of the procedure. The predicted or reproduced $p\hat{f}_k$ values corresponding to the V1, V2, V3, and V4 analyses are given in Tables 26, 27, 28, and 29, respectively. There is no need to present in detail the Gulliksen and Tukey variance-components analysis for the V

TABLE 26

REPRODUCED P_{jk}^* MATRIX FROM V SCALE (V1)

[illegible]

TABLE 27

PREDICTED p_{jk}^* MATRIX FROM V SCALE (V2)

		G4	R4	R6	R2	G6	G2	R8	R1	G1	N
G4	p^*		.653	.714	.619	[.548]	[.578]	.769	.626	.641	.728
	p		.502	.533	.547	[.548]	[.578]	.604	.636	.784	.808
R4	p^*			[.531]	[.544]	.308	.354	.578	.524	.503	.644
	p			[.531]	[.544]	.545	.576	.601	.634	.782	.807
R6	p^*				.435	.245	.374	[.571]	.456	.497	.578
	p				.514	.515	.546	[.571]	.605	.760	.787
R2	p^*					.401	.422	.435	[.592]	.599	.728
	p					.501	.532	.558	[.592]	.751	.778
G6	p^*						.517	.694	.619	.578	.664
	p						.531	.557	.591	.749	.777
G2	p^*							.714	.619	[.726]	.721
	p							.526	.561	[.726]	.755
R8	p^*								.395	.442	.524
	p								.535	.704	.735
R1	p^*									.500	[.707]
	p									.675	[.707]
G1	p^*										[.538]
	p										[.538]
N	p^*										
	p										

Note: Bracket denotes a pair which was used to determine the V scale.

TABLE 28

REPRODUCED p_{jk}^* MATRIX (R STIMULI ONLY) FROM V SCALE (V3)

	R4	R2	R1	R6	R8	N
R4 p^*		.508	.514	.552	.582	.665
R4 p		.544	.524	.531	.578	.644
R2 p^*			.506	.545	.574	.658
R2 p			.592	.565	.435	.728
R1 p^*				.539	.569	.653
R1 p				.544	.605	.707
R6 p^*					.533	.617
R6 p					.571	.578
R8 p^*						.588
R8 p						.524
N p^*						
N p						

TABLE 29

PREDICTED p_{jk}^* MATRIX (R STIMULI ONLY) FROM V SCALE (V4)

	R4	R6	R2	R8	N	R1
R4 p^*						
R4 p		[.531]	[.544]	.601	.624	.634
		[.531]	[.544]	.578	.644	.524
R6 p^*			.514	[.571]	.595	.605
R6 p			.435	[.571]	.578	.456
R2 p^*				.558	.582	[.592]
R2 p				.435	.728	[.592]
R8 p^*					[.524]	.534
R8 p					[.524]	.395
N p^*						.511
N p						.293
R1 p^*						
R1 p						

Note: Bracket denotes a pair which was used to determine the V scale.

theory, since the results are essentially identical to the U theory analysis. This may be seen in Table 30 in which the residual sum of squares, SS_D , are compared. In each case, corresponding SS_D terms between U and V theories are quite close, although there appears to be a slight tendency for the V theory to fit the combined set of stimuli better than the U theory, while the opposite tendency holds for the analyses based on the R stimuli only. However, these tendencies should probably not be taken seriously since they could have easily arisen by "rounding off" errors. In any case the difference between the two theories is, as was shown in Chapter IV, essentially trivial.

TABLE 30
COMPARISON OF SS_D BETWEEN U AND V THEORIES

Analysis	U Theory	V Theory
1.	363.41	354.12
2.	3313.03	3247.43
3.	157.40	157.49
4.	494.23	495.12

There is also no need to repeat the four V scales in the four V analyses since the order of the stimuli on these scales completely agree with the corresponding U scales. The numerical values of the stimuli differ, as is to be expected, since the functions which map the scale values of a pair of stimuli into a preference probability differ.

Q theory analysis of R stimuli

Qa analysis.-The I scales for the R stimuli, given in Table 31, are obtained by the method of stochastic preference. If R2 is higher than R1 in a given I scale, it means that R2 was stochastically preferred by the subject more often than was R1. It should be mentioned,

however, that the preference judgments used in determining these I scales do not include judgments involving the G stimuli; had these judgments been used, fewer ties would have been obtained, but the analysis would have confounded the two hues.

It is evident on inspection of the I scales in Table 31 that the J scale which is compatible with the predominant number of I scales is equal to the Munsell order, that is, the order

N, R1, R2, R4, R6, R8.

The order of the stimuli in forty of the forty-nine I scales is compatible with this scale, except for at most a single adjacent pair of stimuli, which are reversed in order. Of the remaining nine I scales, only one, the I scale of subject 8, is clearly incompatible with a folded Munsell scale. This subject acted as though R4, the middle stimulus, and the extremes, N and R8, were nearly equally acceptable. The analysis will, however, include this subject since his I scale could have resulted by chance from preferences which are compatible with the above scale.

The numerical values of the J scale were obtained by using the p_{jk} matrix given in Table 32. The entries in this table were obtained by using the unilateral judgments only. Precisely how this is accomplished in the present case can be illustrated by the following example.

According to the I scale of subject 2 in Table 31, the ideal stimulus for this subject is in the vicinity of R2. The unilateral judgments for this subject will then consist of the pairs

(N, R1), (R4, R6), (R4, R8), and (R6, R8).

It will be noted that only the stimulus R2 is omitted from the four pairs above. For the pair N, R1 the number of times (out of three)

this subject preferred R1 to N was tabulated and formed the basis of the p_{jk} value in the R1, N cell of Table 32. For the remaining pairs, (R4, R6), (R4, R8), and (R6, R8), the preference for the less saturated stimulus was tabulated. The general procedure for each subject, therefore, was to use all pairs unilateral to the most preferred stimulus (not including the ideal stimulus) and to tabulate for each pair the number of preferences for the stimulus which was nearer to the ideal stimulus. When two or more stimuli were tied and had the highest rank in the I scale, a corresponding number of stimuli was deleted from the above analysis. Subject 4, for example, has R1, R2, and N tied so that only the judgments of pairs (R4, R8), (R6, R8), and (R4, R6) were tabulated. It will be seen that the matrix in Table 31 based on the above tabulation scheme does not represent a preference matrix, but rather a discrimination matrix. The proportion in the N, R1 cell essentially indicates the number of times R1 was greater than N, the positive direction being defined as the direction of greater saturation.

The matrix of normal deviates corresponding to Table 32 is given in Table 33. The separate estimates of each adjacent interval were obtained by taking the difference between the entries in adjacent columns of Table 33. These differences are presented in Table 34 along with the appropriate weighting factors. The weighting factors were necessary to take account of the different reliabilities of each estimate which resulted from the differences in the number of judgments used to determine each proportion in Table 32 (Thurstone, 1928). The weights are determined by

$$\frac{2N_1N_2}{N_1+N_2}$$

TABLE 31

I SCALES FOR R STIMULI

Subject Number	Order of the Stimuli from most Preferred to Least Preferred				Subject Number	Order of the Stimuli from most Preferred to Least Preferred			
1	R6	(R4	R2)	R1	(R8	R6	N)	R4	R1
2	R2	R1	N	R4	R8	R2	R8)	(R1	R4
3	R8	R6	N	R4	R2	R1	R2	R4	R1
4	(R1	R2	N)	R4	R8	R6	(R6	N	R4
5	R4	(R1	R2)	(R6	R8)	R1	(R6	R2	R1
6	R6	R8	R4	R2	R1	R4	R8	R1	R4
7	R8	R6	R4	R2	R1	R4	R8	R1	R4
8	N	(R8	R6	R1)	R2	R4	N	R6	R1
9	R4	(R1	R6)	(R2	R8)	R1	(R6	R2	R1
10	R2	R1	(R4	N)	(R6	R8)	(R6	R2	R1
11	R1	R2	(R8	N)	(R4	R6)	R2	R1	R4
12	R6	R4	(R1	R2	R8)	R1	(R1	R6	R4
13	R1	R2	R6	(R4	R8)	R2	(R6	R2	R4
14	R1	R2	R4	N	R6	R1	(R2	R2	R4
15	R1	R2	(R6	N)	(R4	R8)	(R2	R2	R4
16	R2	R1	N	R4	R6	R1	R1	R2	R4
17	R2	R1	R4	R6	R2	R1	R1	R2	R4
18	R8	R6	R4	R1	R8	R2	(R1	R2	R4
19	R2	R1	N	R4	R6	R1	R6	R2	R4
20	R1	(R4	N)	(R2	R8)	R1	R1	R2	R4
21	N	R1	R2	R4	R6	R1	R1	R2	R4
22	R1	N	(R2	R6	R8)	R1	R1	R2	R4
23	R4	R1	(R2	R6	R8)	R1	R1	R2	R4
24	(R2	R6)	R4	N	R6	R1	R1	R2	R4
25	(R4	R6	R8)	R2	N	R6	R1	R2	R4
26	R8	R6	R2	R1	R4	R6	R1	R4	R2
27	R6	R2	R6	R1	R4	R6	R1	R4	R2
28	R8	R1	R2	R4	R6	R1	R1	R4	R2
29	R1	R4	(R2	R4)	R1	R4	(R6	R2	R1
30	R4	R2	(R1	N)	R4	R6	(R6	R2	R1
31	R2	R2	R4	R1	R4	R6	R8	R1	R4
32	R2	R6	R1	R4	R6	R1	R8	R1	R4
33	R6	R1	R8	R4	R6	R1	R8	R1	R4
34	N	(R2	R4)	(R2	R8)	R1	(R6	R2	R1
35	(R2	R4)	R6	R1	R4	R6	(R6	R2	R1
36	R4	R4	(R2	R8)	R1	R4	R2	R1	R4
37	(R2	R1	R4)	(R1	R8)	R1	(R1	R2	R4
38	(R1	R6	R6)	R2	R8)	R1	(R2	R2	R4
39	R6	R8	R4	R1	R4	R6	(R2	R2	R4
40	R8	R8	R6	R1	R4	R6	(R2	R2	R4
41	R8	R8	R6	R1	R4	R6	(R2	R2	R4
42	N	R4	R1	R4	R6	R1	R1	R2	R4
43	R8	R8	R6	R1	R4	R6	(R2	R2	R4
44	N	R4	R1	R4	R6	R1	R1	R2	R4
45	R4	R8	R6	R1	R4	R6	(R2	R2	R4
46	R8	R8	R6	R1	R4	R6	(R2	R2	R4
47	R8	R8	R6	R1	R4	R6	(R2	R2	R4
48	R2	R2	R1	R4	R6	R1	R1	R2	R4
49	R8	R8	R6	R1	R4	R6	(R2	R2	R4

Note: Parentheses denote tied stimuli.

TABLE 32

P_{jk} MATRIX (R STIMULI ONLY) FROM UNILATERAL JUDGMENTS (Q_a)

	N	R1	R2	R4	R6	R8
N		.750 (108)	.787 (75)	.868 (53)	.848 (33)	--
R1			.690 (87)	.848 (66)	.778 (45)	.750 (12)
R2				.701 (87)	.758 (66)	.813 (33)
R4					.771 (96)	.714 (63)
R6						.656 (93)

Note: Upper number in each cell is a proportion; lower number is the total number of judgments on which proportion is based.

TABLE 33

x_{jk} MATRIX (R STIMULI ONLY) FROM UNILATERAL JUDGMENTS (Q_a)

	N	R1	R2	R4	R6	R8
N		.6745 (108)	.7961 (75)	1.1170 (53)	1.0279 (33)	--
R1	- .6745 (108)		.4959 (81)	1.0279 (66)	.7655 (45)	--
R2	- .7063 (75)	- .4957 (87)		.5273 (87)	.6999 (66)	.9078 (33)
R4	-1.1170 (53)	-1.0279 (66)	-.5273 (87)		.7421 (96)	.5651 (63)
R6	-1.0279 (33)	- .7655 (45)	-.6999 (66)	-.7421 (96)		.4016 (93)
R8	--	--	-.9078 (33)	-.5651 (63)	-.4016 (93)	

where N_1 and N_2 are the number of judgments involved in the two proportions which determine a given estimate. For the direct estimate of an interval by a single p_{jk} value—these terms lie adjacent to the diagonal in the matrix in Table 33,— N_2 may be considered infinite, so the weight is equal to $2N$. If $N_1 = N_2$, then the weight reduces to N . The direct estimate is, therefore, generally weighted at least twice as much as the indirect estimates of a given interval. The scale determined by the average estimates in Table 34 is as follows:

(Qa scale)

N	R1	R2	R4	R6	R8
0	.4521	.7850	1.1748	1.5187	1.7486

With the numerical values on this J scale determined, it is possible to determine the validity of the procedure used in deriving this scale. It was assumed in this procedure that a subject whose most preferred stimulus was R2 could be used in determining the R4, R6 interval. In general, it was assumed that the unilateral equation would give sufficient accuracy if a subject were separated from a given pair by only one adjacent interval. However, the average distance between adjacent stimuli in the Qa scale above is only .35, and from the considerations of Chapter III little accuracy can be expected from the unilateral equations under these conditions. It will be recalled that the accuracy can be improved by using only the direct estimate of each adjacent interval; the indirect estimates tend to underestimate the actual value. In fact, inspection of Table 34 shows that the direct estimate is one of the largest estimates of each interval. It would appear, therefore, that the separation between adjacent stimuli could be increased by using only the direct estimate of each interval, and

TABLE 34

SEPARATE ESTIMATES OF EACH ADJACENT INTERVAL ON THE Qa SCALE

		Adjacent Intervals				
		N R1	R1 R2	R2 R4	R4 R6	R6 R8
Separate Estimates	{	.6745(216) ^a	.0318(88.5)	.4107(62.1)	-.0891(40.7)	.2079(44)
		.2160(80.6)	.4959(174) ^a	.5320(75.1)	-.2624(53.5)	.1770(76.1)
		.0891(58.8)	.5006(75.1)	.5273(174) ^a	.1726(75.1)	.4016(186) ^a
		.2624(38.1)	.0656(53.5)	-.0422(78.2)	.7421(192) ^a	
				.3427(45.4)	.1635(75.1)	
Average Estimates	{	.4521	.3329	.3898	.3439	.2299

^a The direct estimate of the interval.

TABLE 35

p*
jk MATRIX FROM THE Qb ANALYSIS

	R4	R2	R1	R6	R8	N
R4 ^{p*} _p		.522 .544	.581 .524	.569 .531	.632 .578	.682 .644
R2 ^{p*} _p			.572 .592	.542 .565	.596 .435	.698 .728
R1 ^{p*} _p				.489 .544	.538 .605	.657 .707
R6 ^{p*} _p					.584 .571	.605 .578
R8 ^{p*} _p						.552 .524

this modification would increase the accuracy of the unilateral equations. This analysis is carried out in the next section.

Qb analysis.-Using only the direct estimates of each adjacent interval the following scale was obtained.

(Qb scale)

N	R1	R2	R4	R6	R8
0	.6745	1.1704	1.6977	2.4398	2.8414

The average separation on this scale, .57, while larger than the separation on the Qa scale, is not nearly as large as would be desirable. However, the alternative procedure of selecting subjects who are at least two adjacent intervals from each stimulus pair would drastically reduce the number of judgments which would form the basis of the analysis. Since it is difficult to weigh these two factors (the accuracy of the unilateral equations versus the amount of data to be used) both analyses will be carried out. In this section the analysis of the Qb scale will be continued.

Figure 9 contains the histogram of subjects or, more precisely, ideals on the Qb scale. The width of each bar in this histogram is determined by the midpoints between adjacent intervals. The ideal point of each subject cannot be located more accurately than this. (When the most preferred stimulus in an I scale is tied, the average of the scale values of the tied stimuli was plotted.) A normal curve was fitted to the histogram in the following manner. The mean of the normal distribution was determined by the median of the histogram. The mean of the histogram could not be used to determine this parameter because the two end intervals of the histogram are open or unbounded intervals. Determining the value of the standard deviation was somewhat more

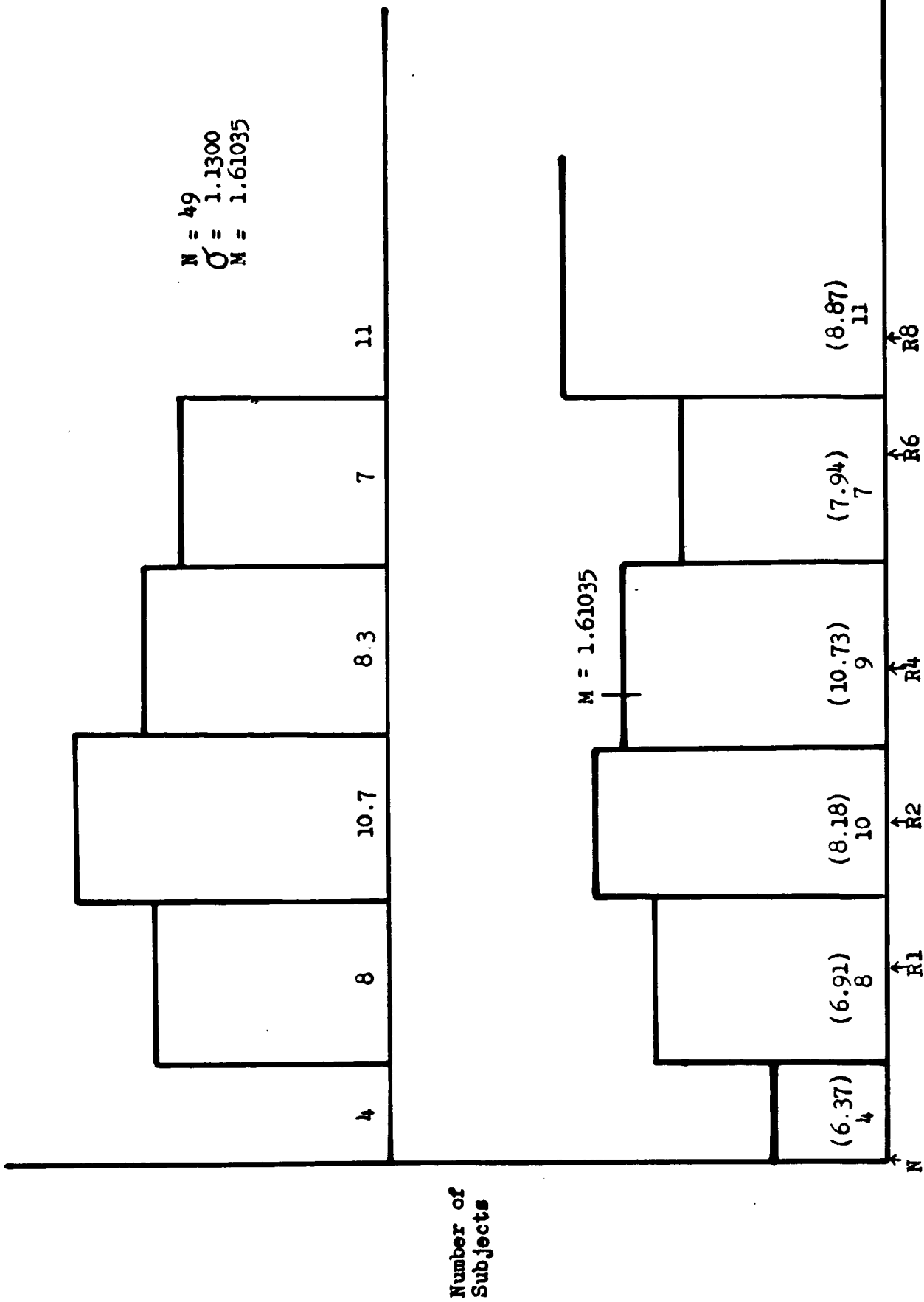


Fig. 9 - Histogram of subjects on the Qb scale. Upper histogram was obtained by equalizing the intervals in the lower histogram (assuming a uniform distribution of subjects within an interval). Numbers in parentheses are predicted values from the normal curve.

complicated. The general procedure was to select the σ which minimized the value of the goodness-of-fit- χ^2 . Three interval estimates of σ were used to determine the final estimate. One estimate of σ was obtained by averaging the two values of σ which would accurately predict the number of subjects in the end intervals of the histogram. These two values of σ are .9148 and 1.3578 and their average is 1.1363. The χ^2 based on this average σ is 2.3785. The two other estimates of σ which were used and their corresponding χ^2 values are:

$$\sigma_2 = 1.1987, \quad \chi_2^2 = 2.6036$$

$$\sigma_3 = 1.0000, \quad \chi_3^2 = 3.1881$$

The final estimate of σ was obtained by fitting the three values of σ and χ^2 to a quadratic curve, $a_1\sigma^2 + a_2\sigma + a_3 = \chi^2$, and then solving the quadratic equation for the minimum value of χ^2 . The value of the quadratic coefficients obtained were

$$a_1 = 48.0502$$

$$a_2 = -108.5896,$$

and using as the minimum point

$$\sigma = - \frac{a_2}{2a_1}$$

the final estimate of 1.1300 was obtained. The χ^2 value of this final estimate is 2.3603 which, in fact, is less than the previous χ^2 values. (The probability level of the final χ^2 for three degrees of freedom is approximately equal to .50.)

The mean and variance of the normal distribution fitted to the histogram in Figure 9 is assumed to be an estimate of the mean and

variance of C_1 in the population from which the subjects are drawn.¹ Consequently the values of the bivariate normal parameters, as described in equations (60), (61), and (64) in Chapter III, are as follows.

$$\left. \begin{aligned} \sigma_{t'_j}^2 &= 1 + \sigma_{C_1}^2 = 2.2769 \\ r_{t'_j t'_k} &= \frac{\sigma_{C_1}^2}{1 + \sigma_{C_1}^2} = .78040 \\ T'_j &= 1.6104 - Q_j \end{aligned} \right\} j = 1 \dots 6$$

The values of Q_j are given in the Qb scale.

Using the above parameter values in equation A, the p_{jk} values predicted by the Q theory (using the Qb analysis) were obtained. These predicted values are contained in Table 35. The residual sum of squares, SS_D , implied by this table is 163.79. Since the SS_D values for the U_3 and U_4 analyses were 157.40 and 494.23, respectively, it is clear that the Qb scale is far superior in accounting for the data than the U_4 scale, but it is not quite as good as the U_3 scale. (The

¹This is not strictly accurate. The variance of the histogram is the result not only of the variability of C_1 over subjects but also of the variability of c_1 within subjects. The total observed variance is therefore an estimate of

$$\left(\frac{n-1}{n} \right) \sigma_{c_1}^2/n + \sigma_{C_1}^2 \left(\frac{N-1}{N} \right)$$

where n , in this case, is the number of observations of each subject's ideal point. The number is difficult to estimate, however. At most it is equal to 135, the number of judgments made by each subject. Since σ_C is approximately equal to 1, σ_{c_1} to $\frac{1}{2}$, a value of n greater than 50 would make the $\sigma_{c_1}^2/n$ term negligible compared to $\sigma_{C_1}^2$. It appears reasonable then to assume in the following analysis that this term is negligible. Since N equals 49, it is also reasonable to assume that the coefficient $(N-1)/N$ can be ignored.

conclusions, of course, are similar for the V3 and V4 scales.) It is difficult to say, however, on the basis of these error terms that the Q theory is more or less adequate than the U or V theories. Nor is the Q theory made more plausible after a consideration of the non-SST triples in the matrix of Table 35. Of the two triples predicted by the Qb analysis which do not satisfy SST only one of these triples failed to satisfy SST in the obtained p_{jk} matrix. The obtained p_{jk} matrix, however, contains eleven triples (out of twenty) which do not satisfy SST.

Qc analysis.—The J scale in this analysis was obtained by deleting for each subject three stimuli, one on each side of the most preferred stimulus. This means that if a subject's most preferred stimulus was R2, then only unilateral stimulus pairs not containing either R1, R2, or R4 were considered for this subject. In this case, only the subject's judgments of the pair R6, R8 would be used. If the subject's most preferred stimulus was an extreme stimulus—that is, either N or R8—generally only two stimuli were deleted. The p_{jk} matrix based on this analysis is given in Table 36.

It is obvious from this table that a considerable amount of the original data has been neglected in the analysis. In fact, of the 2205 judgments ($3 \times 45 \times 15$) in the data, only 282 judgments (or 12.8%) are used in the Qc analysis. (The Qb method used 471 judgments or 21.3% of the data.)

The Q scale determined by Table 36 is as follows:

(Qc scale)

N	R1	R2	R4	R6	R8
0	.5388	1.0205	1.6424	1.9982	2.4687

The average separation between stimuli on this scale is .49, and since

TABLE 36

P_{jk} MATRIX (R STIMULI ONLY) FROM UNILATERAL JUDGMENTS (Q_c)

	N	R1	R2	R4	R6	R8
N		.705 (78)				
R1			.685 (54)			
R2				.733 (45)		
R4					.639 (36)	
R6						.681 (69)
R8						

Note: Numbers in parentheses denote the number of unilateral judgments.

TABLE 37

P_{jk}^* MATRIX (R STIMULI ONLY) FROM THE Q_c ANALYSIS

	R4	R2	R1	R6	R8	N
R4 P^*		.521	.583	.532	.604	.671
P		.544	.524	.531	.578	.644
R2 P^*			.576	.509	.571	.687
P			.592	.565	.435	.728
R1 P^*				.449	.509	.633
P				.544	.605	.707
R6 P^*					.582	.634
P					.571	.578
R8 P^*						.570
P						.524
N P^*						
P						

the stimulus on each side of the most preferred stimulus was deleted in the Qc analysis, each subject was therefore about .98 units from a unilateral pair of stimuli. This distance in the Qb scale was .57 so that the unilateral equations in the Qc analysis should be considerably more accurate.

The histogram of subjects, plotted in Figure 10, was fitted with a normal curve using the same procedure described in the Qb analysis. The mean was set at the median (1.4672) and the three estimates of σ with their χ^2 values were:

$$\begin{array}{ll} \sigma_1 = .9352 , & \chi_1^2 = .7626 \\ \sigma_2 = .920 , & \chi_2^2 = .8319 \\ \sigma_3 = .945 , & \chi_3^2 = .7458 \end{array}$$

The interpolated σ , obtained by minimizing a quadratic function, was .9476. Since the χ^2 value of this sigma was .7540 (.80 < p < .90), which is slightly higher than χ_3^2 , the interpolated estimate was reduced slightly to .9470.

On the basis of these estimates, the parameters of the bivariate normal function were determined, and equation A was used to predict each preference probability. These p_{jk}^* terms are presented in Table 37. The residual sum of squares determined from this table is 196.44. The Qc scale, therefore, accounts for less of the data than the Qb scale ($SS_D = 163.79$) and the U3 scale ($SS_D = 157.40$). However, the Qc scale is considerably better than the U4 scale ($SS_D = 494.23$), and taking into account the fact the U4 analysis uses 2.6 times as much data as the Qc analysis, the results are somewhat surprising.

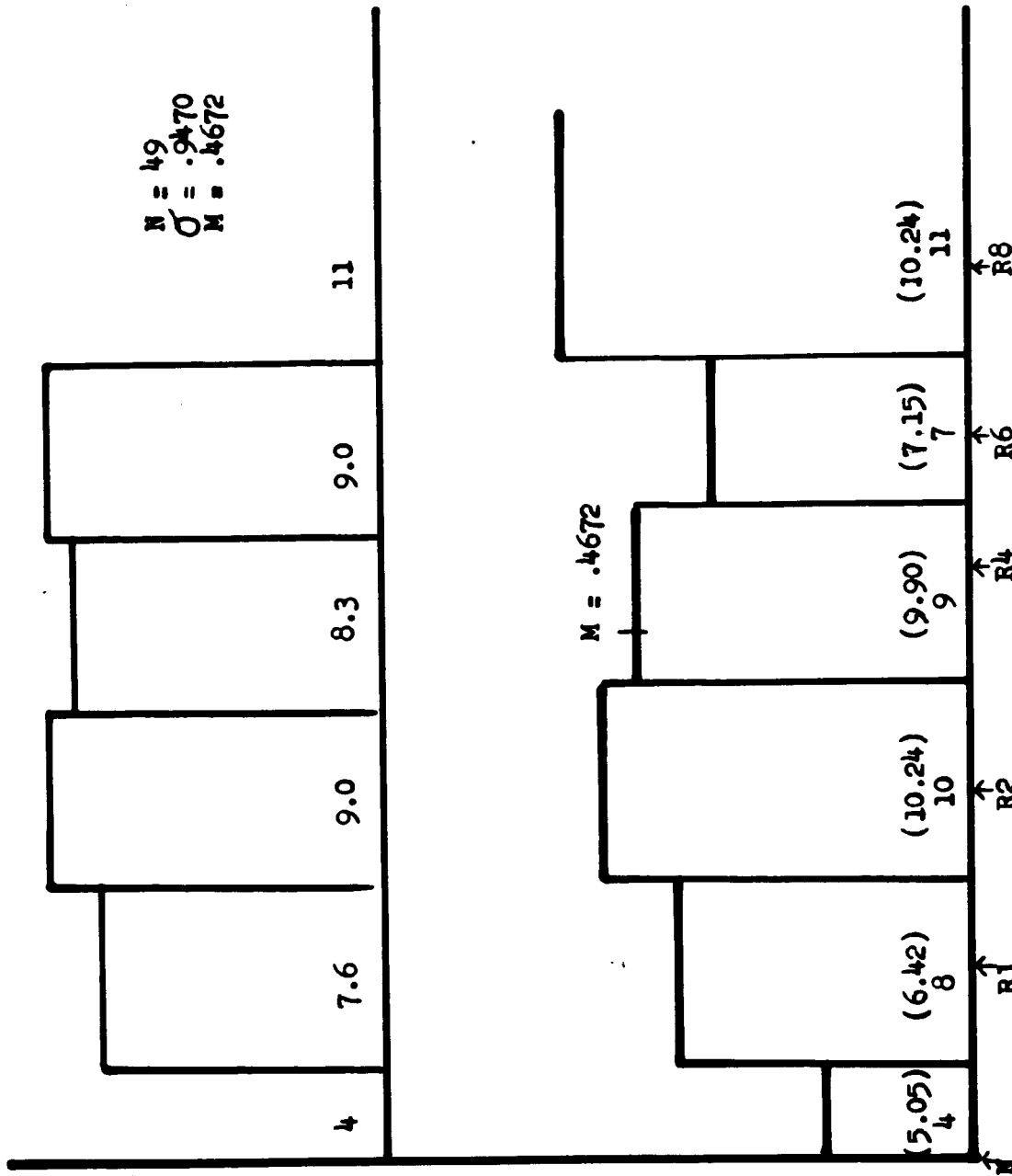


Fig. 10 - Histogram of subjects on the Qc scale. Upper histogram was obtained by equalizing the intervals in the lower histogram (assuming a uniform distribution of subjects within an interval). Numbers in parentheses are predicted values from the normal curve.

Q theory analysis of the combined set of stimuli

The major source of difficulty in analyzing the combined set of stimuli is to determine the order of the stimuli on the J scale from the I scales given in Table 38. While the order of the R stimuli or the G stimuli separately is easily obtained from the I scales—the order is the Munsell order—it is not immediately clear what the combined order is. This order was derived from the following step-wise procedure.

Determination of the J scale.— 1) To determine the extreme stimuli of the J scale, the number of times each stimulus was least preferred was tabulated as follows.

G1	G2	G4	G6	R1	R2	R4	R6	R8	N
7-5/6	1½	0	4	7	½	½	6½	8-1/3	12-5/6

These results suggest that G1, R1, R8, and N are near the extremes of the scale, but the tabulation does not indicate which stimuli are near the same end of the scale.

2) In Table 39 the number of times a stimulus ranked second in the I scale, when a given stimulus was ranked first, is tabulated. The purpose of this tabulation is to determine which pairs of stimuli are juxtaposed on the J scale. The results of this tabulation, however, are clearly disappointing. According to these results, if a subject most prefers a G stimulus then his second choice is generally also a G stimulus. Similarly for the R stimuli. This analysis suggests that the G stimuli and the R stimuli are grouped separately on the J scale, which is, of course, characteristic of the Munsell dimension. However, from other considerations, enumerated below, the Munsell line does not seem to be appropriate for the combined set of stimuli.

3) An alternative but similar tabulation to Table 39 is presented

TABLE 38

I SCALES FOR THE COMBINED STIMULI

Subject Number	Order of the Stimuli for most Preferred to Least Preferred	Subject Number	Order of the Stimuli from most Preferred to Least Preferred
1	G6 G4 (G2 R6) R2 R4 (R1 N) R8 G1	26	G6 (G4 G2) (F8 G1) R6 (R4 N R2) R1
2	R2 G4 G6 R1 G2 G1 N R4 R8 R6	27	G4 G2 G6 G1 R6 R2 R4 (R1 R8) N
3	G6 G2 G4 R8 G1 R6 N R4 R2 R1	28	G6 G4 (G2 R8) R6 (R1 R4) G1 R2 N
4	G2 G6 G4 (R2 N) (G1 R1) R4 R8 R6	29	R1 R4 (R2 N) R6 (G4 R8) G2 G1 G6
5	G6 G4 G2 R4 (R6 R1) (R2 G1) R8 N	30	G2 (G1 R4) G4 (G6 N) (F1 R2) R8 R6
6	(G6 R6 R8) R4 G4 (G2 R2) (F1 G1) N	31	G1 (R2 R4) N G2 (G4 G6) R1 R8 R6
7	G4 G6 G2 R8 R6 R4 G1 R2 R1 N	32	(R1 R2) G2 (G6 R4) G4 (G1 R6) (R8 N)
8	(R8 N) (F6 R1) R2 (G6 G4 G1) (G2 R4)	33	G4 (R6 G6) (F8 G4) R4 (N R2 G1) R1
9	G6 G4 G2 G1 R4 (R1 R6) (R2 R8) N	34	G1 G2 N (G4 R2 R1) R4 R6 R8) G6
10	(F1 R2 G2) (G6 G4) G1 N R4 (R6 R8)	35	R2 R4 (R1 N) G1 (R8 R6 G2) G4 G6
11	R1 R2 N (R4 R6 R8) G1 G4 G6 G2	36	G6 G4 R4 G2 R6 (G1 R8) (R1 R2) N
12	R6 R4 G2 R8 (G4 R1 R2) G6 (G1 N)	37	G4 G6 (G2 R4) R2 (R6 R8) (F1 N) G1
13	R1 R2 R6 (R4 R8) (G4 N) G6 G2 G1	38	R4 R2 (R1 G1) G2 N (R6 G4 G6) R8
14	R1 G2 (G6 R2) G4 (R4 G1) N R6 R8	39	G6 (G4 G2) (R1 R6) (F2 R4 R8) G1 N
15	R1 (G2 R2) G4 (R4 N) (R6 G1) R8 G6	40	G4 (G6 G2 G6 G4 R4 G2 R2 R1 N G1
16	R2 R1 (N G1) (G4 G2 R4) G6 R6 R8	41	(G6 G4 G1 R8) (G2 R6) R4 N R2 R1
17	R2 R1 R4 R6 R8 G2 R6 (G1 R2) N	42	G6 G4 R8 R2 (G2 N) (R4 R6 G1) R1
18	G6 G4 R8 G2 R6 (G1 R4) R1 (R2 N)	43	G4 (G2 G1 N) R1 G6 R2 R4 R6 R8
19	(G6 G4 R2) G2 R1 N G1 R4 R6 R8	44	R4 R2 R6 R8 (G6 G4 R1) G2 N G1
20	G2 G4 (N R1) G1 (G6 R4) R2 R8 G6	45	(G6 G4 G2) R8 R6 (R2 R4) (G1 R1) N
21	N R1 R2 G1 G2 R4 (R6 R8 G4) G6	46	G2 (G4 G1) (F1 R8 G6) (R2 R6) R4 R8
22	G4 G6 R1 (R2 R6 R8 N) G2 R4 G1	47	G1 G2 (R2 R1 N G4 G6) R4 R6 R8
23	(G2 G1 R4) R1 R2 N G4 G6 R6 R8	48	G2 G4 (G1 G6 R8) (R6 R4) R1 R2
24	(G4 G2) (G6 R2 R4) R6 G1 (R8 N) R1	49	
25	R6 (G6 G4) (G2 R8) R4 G1 R2 N		

TABLE 39

THE RELATIONSHIP BETWEEN FIRST AND SECOND
CHOICES AS INDICATED IN THE I SCALES

	G6	G4	G2	G1	R1	R2	R4	R6	R8	N
G6		2.5*	1					.5		
G4	9*		2.5*			1		.5		
G2	2	1.3		2*	1.5					
G1		.3	1							
R1						2*				1*
R2				.5	2.5*		2*			
R4			.5	.5	1	1		1*		
R6		.5							1*	
R8										
N		.3								

Note: Entries denote the number of I scales in which column stimulus was ranked first and the row stimulus second. Asterisks denote the modal value in each column.

in Table 40. This table shows, for example, the number of times a given R stimulus was the most preferred R stimulus when the subject's first choice was a particular G stimulus. The purpose of this tabulation was to discover what interaction or correlation existed between the different hues.

The relationship between the R and G hues indicated in Table 40 is striking. If a subject's first choice is a highly saturated R stimulus, then his first choice among the G stimuli tends to be a highly saturated G stimulus. These results suggest that the R and G stimuli are order on the J scale approximately as follows.

R8, G6, R6, G4, R4, G2, R2, R1, G1 .

This order was used as a first approximation for the remaining part of the analysis.

4) The order of the pair G1, R1, as given in the tentative J scale shown above, was verified as follows. Subjects were selected whose ideal stimulus was not in the region of G1, R1; these subjects have as a first choice R8, G6, R6, G4, or R4. In the I scales of these subjects G1 is ranked over R1 nineteen times; R1 over G1 twenty-four times, and there were three ties. These results confirm the tentative order of the stimuli indicated above, namely that G1 is more extreme on the J scale than R1.

5) It would appear that stimulus N, the neutral gray, would be located near the desaturated end of the scale, near, that is, G1 and R1. This was confirmed by tabulating, for subjects located near the highly saturated end of the tentative J scale, the number of times G1, R1, and R2 were ranked over N. The results were as follows.

TABLE 40

THE RELATIONSHIP BETWEEN THE R AND G STIMULI AS INDICATED IN THE I SCALES

	G6	G4	G2	G1		R1	R2	R4	R6	R8
R1	.5	2	1.5	1	G1	2	2	1	0	.3
R2	0	.25	1.25	1.5*	G2	2.5*	1.5	0*	1	0
R4	3	1.75	1.25	.5	G4	2	1	.5	.5	.3
R6	1.5	2.5	0	0	G6	0	0	.5	.5	1.3*
R8	5	1	1.5	0						

Note: Entries denote the number of I scales in which the column stimulus was first and the row stimulus was the most preferred stimulus which differed in hue from the first choice. Asterisks denote the approximate location of the median value in each column.

TABLE 41

P_{jk} MATRIX FROM UNILATERAL JUDGMENTS (J1)

[illegible]

N > G1 - 8	N > R1 - 9	N > R2 - 4
G1 > N - 16	R1 > N - 17	R2 > N - 20
Tied - 4	Tied - 2	Tied - 4

Since N is the less preferred stimulus in each case, it is evident that N is, in fact, the extreme stimulus at the desaturated end of the scale. This confirms the conclusions reached in section 1) above.

6) The order of the stimuli near the saturated end of the J scale was verified by using the I scale of subjects located at the opposite extreme of the J scale. The tabulation was as follows.

G6 > R8 - 20	G6 > R6 - 19	R6 > R8 - 16
R8 > G6 - 9	R6 > G6 - 9	R8 > R6 - 8
Tied - 2	Tied - 4	Tied - 7

These results indicate that the order of the stimuli from the saturated end of the J scale is R8, R6, G6, which is a modification of the tentative J scale order given above.

7) Using the subjects located near the extremes of the J scale, the order of the middle stimulus was determined by the usual scheme. The results of this tabulation were:

G4 > R4 - 18	G4 > G2 - 16	G4 > R2 - 20
R4 > G4 - 9	G2 > G4 - 5	R2 > G4 - 1
Tied - 1	Tied - 3	Tied - 3
R4 > G2 - 10	R4 > R2 - 17	G2 > R2 - 17
G2 > R4 - 15	R2 > R4 - 3	R2 > G2 - 5
Tied - 0	Tied - 4	Tied - 2

The order of the middle stimuli from the saturated end of the J scale as given by this tabulation is G4, G2, R4, R2.

The order of the stimuli on the J scale as given in the seven step procedure thus far is

(J1 scale)

N G1 R1 R2 R4 G2 G4 G6 R6 R8

This order of the stimuli, a modification of the original tentative J scale given in step 3) will be denoted by J1.

8) Using the J scale in step 7) (J1) and the unilateral judgments, the P_{jk} matrix in Table 41 was constructed. The unilateral judgments that were used in this table did not include at least three stimuli for each subject, usually the three stimuli that were ranked highest in the subject's I scale. Some of the exceptional cases can be illustrated with the following examples. The first three choices of subject 2 as indicated by his I scale in Table 38 are R2, G4, G6. For this subject the three stimuli omitted were R2, R4, G2, so that for this subject the unilateral judgments involved only the pairs (R8, R6), (R8, G6), (R8, G4), (R6, G6), (R6, G4), (G6, G4), (R1, G1), (R1, N) and (G1, N). Subject 10 has three stimuli tied in first position in his I scale, R1, R2, G2, and for this subject the stimuli R1, R2, and R4 were deleted.

Table 41 can be used as the basis for determining the numerical values of the stimuli on the J1 scale, but it leads to difficulties. The average of the estimates of the interval (G1, R1), for example, is negative, indicating that the order of G1, R1 on the J1 scale should be reversed. However, if the order of G1, R1 is reversed then the matrix in Table 41 is no longer valid and must be retabulated.

9) In this step the order of G1, R1 was reversed, giving as the J2 scale

(J2 scale)

N R1 G1 R2 R4 G2 G4 G6 R6 R8

The p_{jk} matrix based on the unilateral judgments corresponding to the J2 scale is given in Table 42, and the numerical values which result from Table 42 (using the weighting scheme discussed previously) is as follows.

(J2 scale)

N	R1	G1	R2	R4	G2	G4	G6	R6	R8
0	.3580	.4999	.5429	.9141	.9409	1.3029	1.5437	1.6407	1.7814

The average separation between adjacent stimuli on this scale is .19, and therefore the distance between a subject's location and a unilateral pair of stimuli is approximately .38. This separation between subject and unilateral stimulus pair is, of course, entirely too low to obtain appreciable accuracy from the unilateral equations.

10) Greater accuracy can be obtained with the unilateral equations if the direct estimate of each adjacent interval is used instead of the average of the indirect estimates. However, if the direct estimates are used in Table 42 the stimulus pair G1, R2 will have to be reversed since the proportion in the G1, R2 cell is less than one half. Reversing this pair of stimuli means, of course, retabulating the entire matrix. The revised p_{jk} matrix based on the J scale in which G1 and R2 are reversed (called the J3 scale) is given in Table 43. The J scale values corresponding to Table 43 are as follows.

(J3 scale)

N	R1	R2	G1	R4	G2	G4	G6	R6	R8
0	.5044	.8549	.8975	1.0181	1.0432	1.4748	2.1619	2.6324	2.8857

TABLE 42

P_{jk} MATRIX FROM UNILATERAL JUDGMENTS (J2)

[illegible]**TABLE 43**

p_{jk} MATRIX FROM UNILATERAL JUDGMENTS (J3)

[illegible]

The average scale separation on this J scale is .32 and therefore the average separation between a subject and a unilateral stimulus pair is .64. While this separation is not as high as would be desired, this scale will be used in the remaining analysis of the combined set of stimuli.

Od analysis.-The histogram of subjects plotted on the J3 scale in Figure 11 does not appear to have a normal distribution. The histogram can, however, be fitted rather well with two normal distributions. The procedure employed in fitting two normal distributions to this histogram was as follows.

The mean of one normal curve, M_1 , was set at .9034, which is approximately in the center of the modal bar, the bar containing 13.47 subjects. Since there are 13 subjects to the left of M_1 , the first normal curve will consist of 26 subjects. This leaves 22 subjects for the second normal curve or 11 subjects to the right of M_2 . Using this fact, M_2 was located at 2.0288. The standard deviation of the first normal curve σ_1 was determined by the two extreme bars on the left in Figure 11. This was done by computing for each of the two bars a σ which would accurately predict the number of subjects in each interval and then averaging the two standard deviations. A similar procedure was used for the second normal curve, employing instead the two extreme intervals on the right. The values of the six parameters of the two normal distributions are summarized in Figure 11.

The χ^2 based on the values of the six parameters of the two normal curves is equal to .6661, which for four degrees of freedom has a probability level between .95 and .98. The fit of the two normal curves to the histogram is therefore reasonably good.

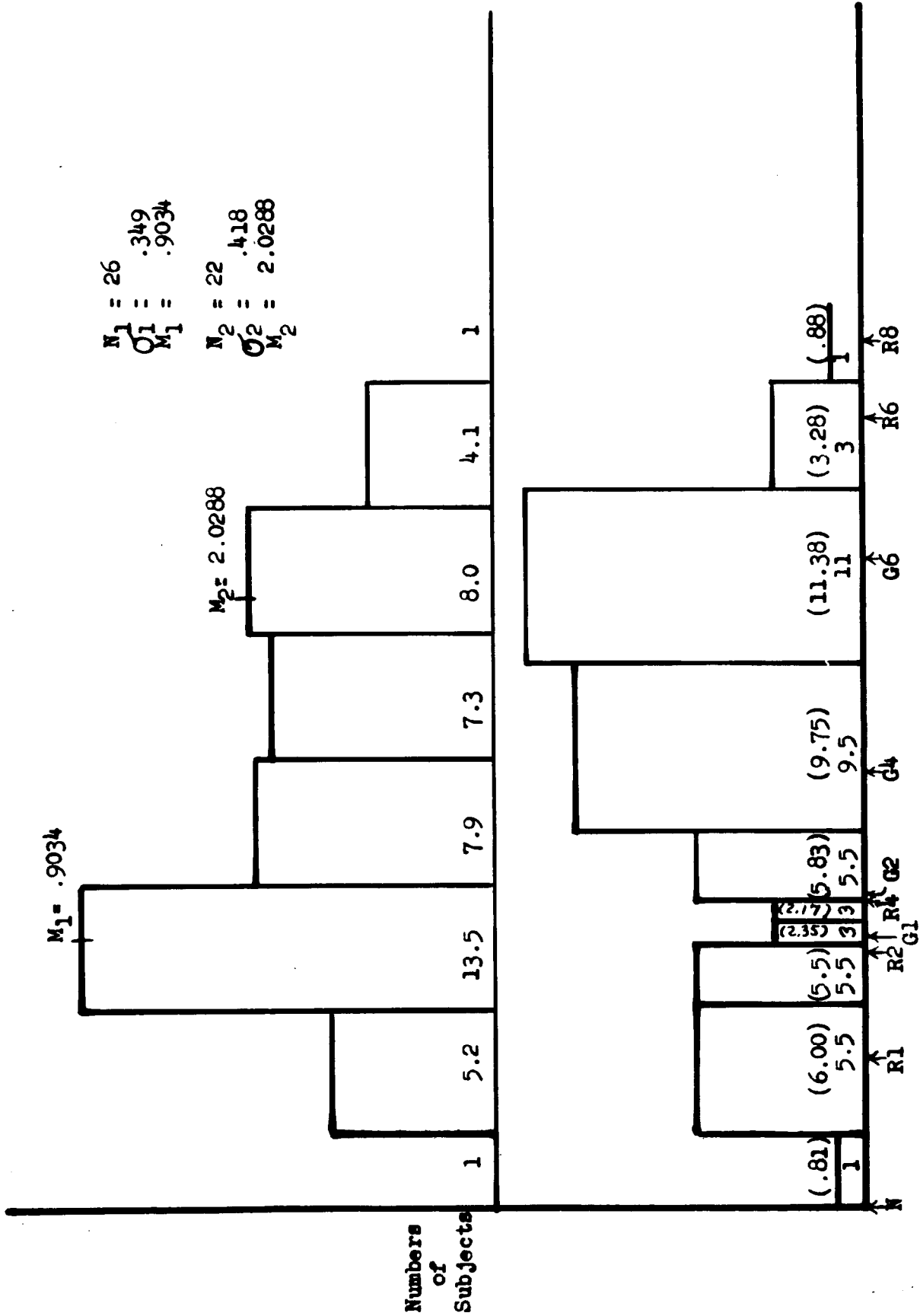


Fig. 11 - Histogram of subjects on the Qd (J3) scale. Upper histogram was obtained by equalizing the intervals in the lower histogram (assuming a uniform distribution of subjects within an interval). Numbers in parentheses are predicted values from the normal curve.

In Table 44 the p_{jk}^* values predicted from equation A using the values determined in Figure 11 are presented for each of the two normal curves. The combined p_{jk}^* values, obtained by a weighted average of the two p_{jk} values in Table 44, are presented in Table 45. The residual sum of squares (SS_D) derived from this table is equal to 1378.80. This value is notably larger than the SS_D value for the U1 analysis (363.407) but it is less than the SS_D of the U2 analysis (3313.03). In fact, in terms of the goodness of fit, it would appear that the Qd scale in the present analysis lies midway between the U1 and U2 scale, having an SS_D three times larger than the former scale and one third as large as the latter.

In comparing the residual sum of squares that are summarized in Table 46, it would appear that the Qd analysis of the combined set of stimuli is somewhat inferior to the Qb analysis of the R stimuli. This conclusion is based on the observation that the SS_D of the Qb analysis is more similar to the SS_D of the U3 analysis than is the SS_D of the Qd analysis to the SS_D of the U1 analysis. It should also be noted that the superiority of the Qd analysis to the U2 analysis cannot be directly attributed to the amount of data which each theory used since, in fact, the Qd analysis uses only about half as much data in determining the stimulus parameters.

Comparison of discrimination scales

In Figure 12 the discrimination scales, J2 and J3, are plotted along with the Munsell scale. Since the R and G stimuli are interwoven on the J scales while kept separate on the Munsell scale, it is obvious that the J scales and Munsell scales are not compatible. It appears that if the Munsell scale were folded onto itself at the neutral gray

TABLE 44

PREDICTED p_{jk} VALUES FROM NORMAL CURVES (1) AND (2) IN THE Qd ANALYSIS

	G4	G6	G2	R2	R1	R4	G1	R6	R8	N
G4 (1)		.671	.449	.448	.475	.448	.447	.793	.847	.560
G4 (2)		.455	.597	.647	.680	.604	.635	.507	.551	.851
G6 (1)			.300	.306	.341	.300	.304	.661	.742	.427
G6 (2)			.625	.665	.738	.631	.656	.555	.605	.826
G2 (1)				.495	.476	.497	.493	.802	.846	.611
G2 (2)				.556	.665	.508	.543	.431	.476	.808
R2 (1)					.528	.504	.497	.791	.834	.618
R2 (2)					.615	.451	.486	.393	.438	.773
R1 (1)						.475	.472	.752	.795	.600
R1 (2)						.340	.373	.323	.367	.679
R4 (1)							.495	.801	.845	.613
R4 (2)							.536	.426	.471	.804
G1 (1)								.794	.837	.618
G1 (2)								.402	.447	.782
R6 (1)									.595	.331
R6 (2)									.557	.767
R8 (1)										.283
R8 (2)										.727

Note: Upper entry in each cell denotes prediction from curve (1); lower number from curve (2).

TABLE 45

P_{jk}^* MATRIX FROM THE COMBINED PREDICTIONS
OF CURVES (1) AND (2) IN THE Qd ANALYSIS

		G4	G6	G2	R2	R1	R4	G1	R6	R8	N
G4	P_{jk}^*										
	p	.572 .548	.517 .578	.539 .619	.569 .626	.519 .653	.533 .641	.662 .714	.712 .769	.693 .728	
G6	P_{jk}^*										
	p		.449 .517	.471 .599	.523 .619	.451 .692	.465 .578	.612 .755	.679 .694	.610 .664	
G2	P_{jk}^*										
	p			.523 .578	.563 .619	.502 .646	.516 .726	.632 .626	.677 .714	.702 .721	
R2	P_{jk}^*										
	p				.568 .592	.480 .456	.492 .599	.609 .565	.652 .435	.689 .728	
R1	P_{jk}^*										
	p					.413 .476	.427 .500	.556 .544	.599 .605	.636 .707	
R4	P_{jk}^*										
	p						.514 .503	.629 .531	.674 .578	.701 .644	
G1	P_{jk}^*										
	p							.615 .503	.658 .558	.693 .538	
R6	P_{jk}^*										
	p								.577 .571	.531 .578	
R8	P_{jk}^*										
	p									.487 .524	

TABLE 46

COMPARISON OF RESIDUAL SUM OF SQUARES FOR U, V, AND Q ANALYSES

U	V	Q
$U_1 = 363.41$	$V_1 = 354.12$	} $Q_d = 1378.80$
$U_2 = 3313.03$	$V_2 = 3247.43$	
$U_3 = 157.40$	$V_3 = 157.49$	} $Q_b = 163.79$
$U_4 = 494.23$	$V_4 = 495.12$	
		} $Q_c = 196.44$

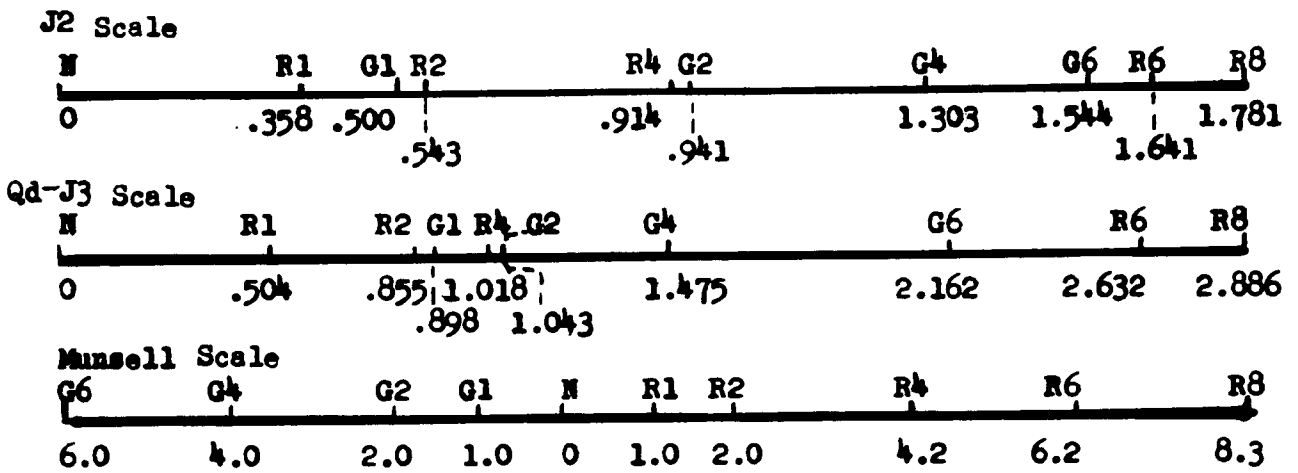


Fig. 12 - Comparison of J scales with Munsell scale.

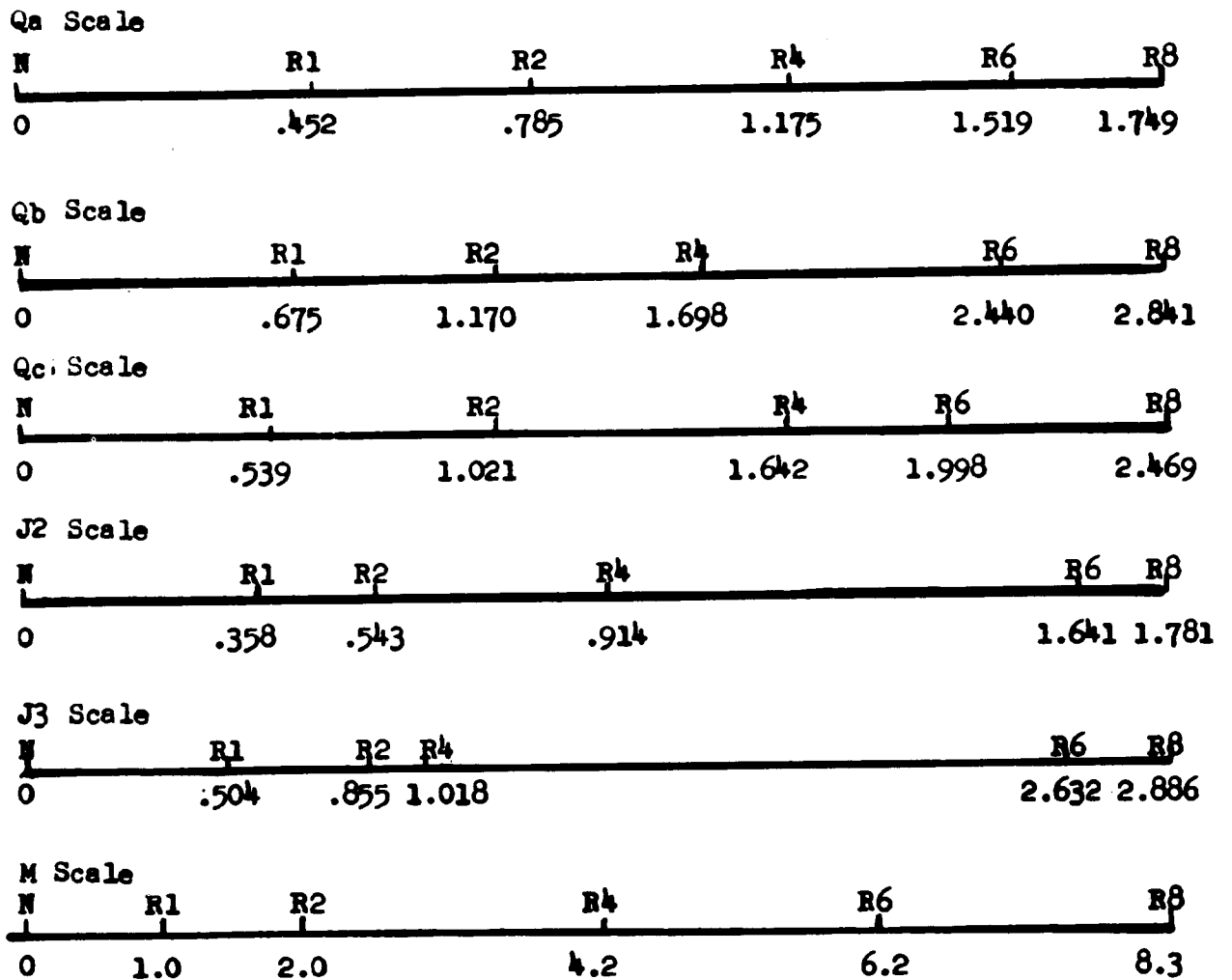


Fig. 13 - Comparison of J scales with Munsell scale (R stimuli)

stimulus, N, the Munsell scale would agree more closely with the J scales. This folded Munsell order is presented below.

N, (R1 G1) (R2 G2) G4, R4, G6, R8

It will be seen that the folded Munsell scale does not quite agree with the two J scales. In J2 the rank of R4 differs by two and, in addition, in J3 the pair G1, R2 are reversed. However, the conclusion that the folded Munsell scale is incompatible with the discrimination scale obtained with preference judgments must be viewed as highly tentative, considering the fact that the Qd analysis based on the J3 scale did not fit the data particularly well. Furthermore, the following analysis of the R stimuli suggests the reverse conclusion.

In Figure 13 the five J scales of the R stimuli are plotted; three of these scales are based on the analysis of the R stimuli only and two are derived from the J scale of the combined set of stimuli. The squared correlations between each of these scales and the Munsell scale of the R stimuli are as follows.

$$r^2_{J2,M} = .9716$$

$$r^2_{J3,M} = .9346$$

$$r^2_{Qa,M} = .9496$$

$$r^2_{Qb,M} = .9685$$

$$r^2_{Qc,M} = .9623$$

These correlations are strikingly related to the predictability of each scale. It will be recalled that the goodness of fit of the Qb analysis was superior to the goodness of fit of the Qc analysis—the SS_D of the former being less than the SS_D of the latter—and correspondingly the correlation of the Qb scale with the Munsell scale is higher than

the correlation of the Qc scale with the Munsell scale. Also it will be recalled that the Qb analysis appeared to be superior to the Qd analysis, since the analysis of the former was more similar to the corresponding U theory analysis employing the reproduction method. Correspondingly, the correlation of the Qb scale to the Munsell scale is higher than the correlation of the Qd (J3) scale to the Munsell scale.

The implication of the results based on the R stimuli confirms the validity of the Munsell scale as a description of the discrimination scale obtained from discriminative judgments as well as the discrimination scale obtained from preferential judgments. In other words, for a single hue these results suggest that the independence assumption is tenable.

Although some of the J scales correlate better with the Munsell scale than others, there are nevertheless some consistent discrepancies in the metric relations between the Munsell and the J scales. The interval between stimulus N and stimulus R1, for example, is one of the smallest on the Munsell scale, while it is one of the largest adjacent intervals on the J scale. The actual rank of the N, R1 distance on each of the J scales is as follows. (Rank 1, denotes the largest distance, rank 5 the smallest.)

Scale	Rank
J2	3
J3	2
Qa	2
Qb	2
Qc	2

The magnitude of the N, R1 interval on the J scales can be explained as the result of a perceived qualitative difference between these two stimuli. While logically the N stimulus is the natural limit of the R stimuli as the stimuli decrease in saturation, most subjects reacted to the N stimulus as a neutral gray, not as a desaturated red. A subject,

for example, who preferred desaturated reds tended to avoid the neutral gray.

There is also some discrepancy in the magnitude of the interval R6, R8 on the J scales and this interval on the Munsell scale. The R6, R8 interval on the Munsell scale is one of the largest intervals while on the J scales it is one of the smallest. The rank of the R6, R8 interval on each of the J scales is as follows.

Scale	Rank
J2	5
J3	4
Qa	5
Qb	5
Qc	4

Apparently subjects do not have strong preferences between R6 and R8, both of which are seen as bright red colors.

The discrepancy between the Munsell and the J scales can also be explained by the rapid exposure of the stimuli. The rapid exposures of the stimuli might have distorted the usual perception of the colors. Some slight confirmation of this conclusion is afforded by comparing the preferences obtained in the slow and rapid exposures. Of the eleven judgments made under the slow presentations, on the average 1.45 disagreed with the judgments made under the rapid exposure. A disagreement between the judgments for a given stimulus pair x, y presented under slow and rapid exposures was defined as follows: if the number of stochastic choices for x under the rapid presentations exceeded the number of choices for y by more than one, and if y was preferred to x in the slow presentations, then the judgments between the slow and rapid exposures were said to disagree. Although the average discrepancy of 1.45 appears relatively small, it is high enough to make one suspect that some distortion of the perceived character of the colors may have resulted

from the rapid exposure.

There is, however, another explanation for the discrepancies between slow and rapid presentations. The discrepancy could be explained by the confounding of the exposure time with inter-stimulus presentation time. This confounding could have resulted from the fact that the panel was moved mechanically by the experimenter. For rapid exposures the panel was moved continuously and swiftly across the display board, so that the two stimuli appeared successively with essentially no pause between presentations.

For the slow presentations the experimenter moved the panel first to a position in which one stimulus was visible then, after a few seconds he moved the panel slowly to the second position, to display the second stimulus. In this case the time between stimulus presentations was quite perceptible. The subject's task under the slow presentation may therefore have been more difficult than was originally supposed, difficult enough perhaps to produce inconsistent judgments. If this were the case, the discrepancy between slow and fast presentations would not necessarily result from a distortion factor but from the inconsistency or errors of judgments which occurred under both sets of conditions.

CHAPTER VIII

DISCUSSION

The probabilistic Q theory was applied to the experiment reported in Chapters VI and VII in order to test the adequacy of the theory and the validity of the independence assumption, the assumption that the discrimination scale obtained from preferential judgments is equal to the discrimination scale obtained from discriminative judgments. In the present chapter the results and further implications of this experiment will be discussed.

Test of the Q theory

The Q theory cannot easily be assessed because the method of solution employed does not satisfy a least squares criterion. To evaluate the Q theory, a procedure was described in Chapter VI which consisted of comparing the residual sum of squares of the Q theory to two different residual sum of squares of the U theory, one obtained by a method which satisfies a least squares criterion (the reproduction method) and one obtained by a method which does not (the prediction method). The rule adopted in Chapter VI was the following: If the residual sum of squares of the U theory based on the reproduction method exceeded the residual sum of squares of the Q theory, the Q theory would be considered superior to the U theory; if the reverse relation held between the Q theory analysis and the U theory analysis with the residual sum of squares of the U theory based on the prediction method, then the U theory would be considered superior.

The results of the previous chapter indicated that the magnitude of the residual term of the Q theory was between the magnitudes of the two residual terms of the U theory; the SS_D of the Q theory was greater than the SS_D of the U theory based on the reproduction method and less than the SS_D of the U theory based on the prediction method. In one case involving the Qb analysis of the R stimuli, the residual term of the Q theory was nearly equal to the SS_D of the reproduction method. It thus appears reasonable to conclude that, with respect to the R stimuli, the Q theory is essentially equivalent to the U (and V) theories in accounting for the data.

A consideration of the non-SST characteristics of the data did not lend further support for the Q theory. The Qd analysis, which best accounted for the data, did not account for the non-SST triples which existed in the data. If the Q theory is assumed to be an essentially correct representation of the data, then one has to conclude that the SST triples which were present in the data resulted from random or chance factors only. In other words, one cannot conclude from the Q theory analysis that preferential choices give rise to non-SST probabilities. However, since the Qd analysis and the U and V analyses as well, did not completely account for the data, as indicated by the Mosteller test, there exists the possibility that the obtained non-SST triples were significant.

The discrepancy between the predictions of the Q theory and the data has two possible sources. One arises from the fitting procedure and does not necessarily reflect a basic inadequacy of the Q theory. The second lies in the essential nature of the Q theory. These two sources will be considered separately.

Inadequacy of the fitting procedure. - The effectiveness of the fitting procedure was limited by the approximate nature of the unilateral equations. The analysis in Chapter III indicated that three digit accuracy could only be expected from the unilateral equations when unilateral subjects were at least two standard deviations from a given pair of stimuli. The separation between unilateral subjects and stimulus pairs in the experiment was always considerably less than this. It would seem, therefore, that the limitations of the Q theory analyses could be attributed to the inaccuracy of the unilateral equations. However, the results suggest otherwise. The Qb analysis predicted better than the Qd or Qc analyses despite the fact that the unilateral equations were less accurate in the Qb analysis. It appears, therefore, that while the unilateral equations limited the predictability of the Q theory, they were not the major cause of the discrepancy between the Q theory predictions and the data.

The discrepancy may also have arisen from the normal distribution assumption which was made in order to treat the group data. The assumption that C_1 is normally distributed over the group of subjects was made merely because the fitting procedure did not permit determining the parameters of the Q theory for each subject. In principle, however, it would have been possible to estimate C_1 for each subject in the experiment, despite the small number of replications which were obtained. The group predictions could then have been obtained by averaging the predictions for each subject, and in this manner, the normal distribution assumption could have been dispensed with. The normal distribution assumption, therefore, is not considered an essential part of the Q theory but only as an assumption necessary for the fitting

procedure described in Chapter III.

Although the normal curve which was fitted to each of the histograms in the Q analyses did not perfectly fit these histograms, the discrepancy between the Q theory predictions and the data cannot readily be attributed to their lack of normality. This conclusion is based on the finding that the Qb analysis, involving a histogram that deviated most from normality, gave the best predictions.

It is interesting to examine further the character of the histograms. Inspection of Figure 11 reveals that the bimodal aspects of the Qd histogram is the result of the difference between subjects whose most preferred stimulus was a G stimulus and those whose most preferred stimulus was an R stimulus. According to this figure, most subjects who preferred a G stimulus preferred a high-saturated G stimulus; those preferring an R stimulus selected a low-saturated stimulus. The bimodal character of this histogram arises, therefore, not merely from a distinction between subjects preferring high-saturated stimuli and those preferring low-saturated stimuli, as indicated in Eysenck's study (1941b), but from a difference between the hue preferences of the subjects.

According to Table 40, subjects who prefer highly saturated G stimuli tend to prefer highly saturated R stimuli. One would thus expect that the histogram of subjects on the J scale of R stimuli only would have a bimodal character to it also. One mode, occurring at the low-saturated end of the scale, would reflect those subjects whose most preferred stimulus was an R stimulus, while the second mode, occurring at the high-saturated end of the scale, would reflect those subjects whose most preferred stimulus was a G stimulus. Close

inspection of the histogram on the J scale of the R stimuli (Qb and Qc) reveals that, in fact, these distributions do have a slight bimodal character to them. This would explain, in part, the reason why these histograms did not perfectly fit a normal curve.

Another limitation of the fitting procedure stems from the small amount of data used. The greater the amount of data used in the Q analysis the better was the predictability of the Q theory. The percentage of data used by each Q analysis was as follows:

Qb	Qc	Qd
21.3%	12.8%	11.1%

According to this tabulation, Qb used the largest relative amount of data, and from the results in the previous chapter, Qb predicted best. It would seem, therefore, that any improvement in the amount of data which the fitting procedure employs would materially improve the predictability of the Q theory. Improving the fitting procedure, therefore, should be the object of further study.

It should be mentioned once again that while the prediction method of the U theory was based on more data than the corresponding Q theory analysis, the U theory in this case did not account for as much data as the corresponding Q theory analyses. (The U2 analysis used 20 percent of the data; U4 used 33.3 percent.)

Inadequacy of the Q theory. - Tables 39 and 40 taken together are essentially incompatible with the single dimensional Q theory. Table 39 indicates that the first and second choices of subjects tend to consist of stimuli that have the same hue, and this fact implies that the R and G stimuli are dichotomized and located at different ends

of the J scale. However, Table 40 indicates that the most preferred R stimulus tends to have the same saturation as the most preferred G stimulus. This finding suggests that the R and G stimuli are interwoven together on the J scale and not at opposite ends of the scale. The single dimensional Q theory used in this study cannot explain both of these results. A multidimensional version of the Q theory, however, may explain the results. For example, one could plot the R and G stimuli on separate orthogonal axes and locate the subjects at a point in the plane of these two axes. A subject located at a given point could base his preferences on the distances between the stimuli located on the axes and his ideal point located in the plane; the smaller the distance the greater the preference.

This model could explain the implications of Table 40 (the correlation of the different hues in terms of their saturation) by assuming that the ideal points in the plane were colinear. The implications of Table 39 could also be explained by this multidimensional model if a different metric were postulated on each of the two axes. This would permit locating a subject at a point that was closer to, let us say, R6 and R4 than to G6, with G6 and R6 being the ideal G and R stimuli, respectively. It could be worthwhile, therefore, to attempt to develop the equations for a multidimensional and probabilistic Q model, equations analogous to A or B of Chapter III.

The independence assumption

In general, the results tended to confirm the assumption that the Munsell scale represents the discrimination scale for both preferential and discriminative judgments involving a single hue. This conclusion is implied by the finding that the degree of correlation

between the J scale of the R stimuli and the Munsell scale was directly related to the predictability of the J scale. There were, however, certain consistent differences between the J scales and the Munsell scale which suggest that the Munsell scale is not a perfect representation of the discrimination scale for preferential judgments involving a single hue. This could be tested directly by using the Munsell scale to predict the preferential choices. One could fit the histogram of subjects plotted on the Munsell scale, instead of the J scales derived from the unilateral equations.

For preferential choices involving more than one hue there was evidence that the Munsell scale was not an adequate representation of the discrimination scale. However, it also appears that the Q theory did not adequately account for the data involving more than one hue. What is apparently required is some multidimensional version of the Q theory as indicated in the previous section.

Part of the discrepancy between the J scale and the Munsell scale may have been due to a distortion of the hues produced by the rapid exposure of the stimuli. Since there was some evidence for this distortion factor, as indicated in Chapter VII, it would be desirable to study choices of stimuli that were more indiscriminable than was employed in this study. This would permit slow presentations of the stimuli.

It would also be desirable to investigate the effects of context on the stimulus scale. In the present study the subjects were instructed to judge the stimuli with respect to their desirability as a color of paint for a living room. There is good evidence to indicate that preferential choices change drastically as the context changes from a living room, let us say, to a clothing article, or a car.

However, it would be interesting to determine whether the metric relations on the discrimination scale also changed as the context changed.

Concluding remarks

In general it would appear that the usefulness of the probabilistic Q theory as it is applied in the present study is confined to situations in which (1) the ideal stimuli of the subjects are known, (2) the order of the stimuli on the J scale is known, and (3) a minimal number of preferential judgments is desired to construct a numerical J scale. It is clear certainly that under these conditions the Q theory parameters could be determined from a relatively small number of preferential choices. In fact, since the order of the stimuli and the ideal stimuli are assumed to be known in advance of the scaling procedure, no preferential judgment need be asked of a subject which is not to be used in the analysis. The predictability of the Q scale based on the small number of judgments should be considerably better than the predictability of the U or V theories based on the same number of judgments.

CHAPTER IX

SUMMARY

Unlike most theories of preferential choice, in the Coombs Quadrant I (or Q) theory the stimuli can be scaled either in terms of their preferability or in terms of their discriminability. However, the Q theory has two severe limitations. It is difficult or impossible to obtain a solution if there is inconsistency or error in the data and at best it can only yield a relatively low level of measurement. This study developed and tested a probabilistic Q theory designed to overcome these two difficulties.

In the first part of the study, the equations for the probabilistic theory were developed and an approximate method of solution was indicated. This approximate method estimates the parameters of the theory from the preference probabilities based on judgments satisfying certain conditions. In general, not all of the judgments in a given set of data would satisfy these conditions, so that only part of the data would be used in arriving at a solution. To test the adequacy of the theory in predicting the preference probabilities based on all judgments present in a given set of data, numerical methods for evaluating the integrals appearing in the equations of the Q theory were described.

The transitivity implications of the probabilistic Q theory were also discussed. It was shown that the Q theory does not necessarily imply strong stochastic transitivity (SST), which is in

contrast with the Thurstone and Bradley-Luce theories, both of which imply SST.

In the experimental part of the study, subjects were required to indicate their preferences for Munsell colored chips presented pairwise. The ten stimuli ranged over two hues (red and blue-green) that formed a single dimension in the Munsell space. The data were analyzed by the probabilistic Q theory, the Thurstone theory and the Bradley-Luce theory.

The results indicated that for the stimuli employed in the experiment the Bradley-Luce and Thurstone theories were essentially equivalent. For a single hue, the Q theory was approximately equal to these theories, but when two hues were involved the Bradley-Luce and Thurstone theories better accounted for the data. However, in all cases when the parameters of the Bradley - Luce and Thurstone theories were evaluated from only part of the data, the Q theory better accounted for the entire data. There was some evidence that a two dimensional extension of the Q theory was required for two hues.

The results also had implications for the Munsell scale. When a single hue was involved the Munsell scale represented, to a first approximation, a discrimination scale valid for both discriminative and preferential judgments. This, however, was not the case when more than one hue was involved.

One of the major factors limiting the degree to which the Q theory can reproduce data is the relatively small amount of data which can be incorporated into the method of solution. In particular, it was shown that the reproducibility of the Q theory was directly related to the amount of data used in the analysis.

APPENDIX 1

NUMERICAL INTEGRATION

Although the approximate method employing the unilateral equations are used to obtain a solution in the Q theory, it is necessary to employ numerical integration techniques. This is necessary for two reasons. It is desirable, first of all, to ascertain how accurate the unilateral equations are for given values of C_i , Q_j and Q_k . Secondly, in order to determine how good the fit of the model is to the data, it is necessary to reproduce or postdict each p_{jk} value---not merely the p_{jk} value of certain unilateral persons---from the parameters of the model. In each case evaluation of the integrals in equation A or B are necessary and these integrals can only be evaluated by numerical methods.

One of the simpler numerical methods is called Simpson's method, and considering its simplicity, it is probably one of the most accurate. It is the method used in this study. For a single integral,

$$y = \int_{x_0}^{x_n} f(x) dx$$

the method essentially consists of (1) evaluating $f(x)$ for discrete values, $x_0, x_1, x_2 \dots x_n$, evenly spaced from x_0 to x_n ,

- (2) fitting to each triple, x_i, x_{i+1}, x_{i+2} a quadratic curve, and
 (3) then determining the area under each quadratic curve. All this is embodied in the rather simple equation

$$(1) \quad y = \frac{|x_n - x_0|}{3n} \left[y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \dots + 4y_{n-1} + y \right]$$

for n even, and

$$y_i = f(x_i).$$

The accuracy of this equation for fixed limits of integration depends only on n , the number of subdivisions in the $|x_n - x_0|$ interval; the larger the n , the better the accuracy. A simple estimate of the error, E , in Simpson's method, given by Scarborough (1958, p.178) involves computing y for two different values of n , n_1 and n_2 , where $n_2 = 2n_1$. The error in the calculation involving n_2 is

$$(2) \quad E_2 = \frac{R_2 - R_1}{15},$$

where R_1 and R_2 are the estimates of y based on n_1 and n_2 , respectively.

Simpson's method can be generalized to the double integral

$$z = \int_{y_0}^{y_m} \int_{x_0}^{x_n} f(x, y) dx dy$$

as follows. Let

$$z_{ij} = f(x_i, x_j) \quad \text{for } i = 0, 1, \dots, n \\ j = 0, 1, \dots, m$$

For fixed y_j , (1) can be applied to the $\underline{x}$ variable to obtain what will be called A_j ; that is,

$$(3) \quad A_j = \frac{|x_n - x_0|}{3n} (z_{j0} + 4z_{j1} + 2z_{j2} + 4z_{j3} + \dots z_{jn})$$

If (47) is applied once again, this time to A_j , the approximation to $\underline{z}$ is obtained. Thus

$$(4) \quad z \approx \frac{|y_m - y_0|}{3m} (A_0 + 4A_1 + 2A_2 + 4A_3 + \dots A_m).$$

The evaluation of $\underline{z}$, therefore, will require evaluating the integral $(n+1)(m+1)$ times, or for $n = m$, $(n+1)^2$ times.

For $n = m = 10$, the integrand will have to be evaluated 121 times, and if the integrand is a bivariate normal distribution, it becomes necessary to use a high speed digital computer.

An estimate of the error term for a double integral does not seem to be readily available in the mathematical literature. The method employed in this study was to check the numerical integration against certain known integrals before proceeding to the unknown integrals. This is done in Appendix 2.

APPENDIX 2

EVALUATION OF EQUATION A

Equation A rather than A' was evaluated despite the fact that the latter involves a single integral (and one double integral). Equation A consists of two double integrals, F_1 and F_2 , but because of the similarity between F_1 and F_2 ---they have the same integrand---programming the IBM 650 is simpler for equation A.

Simpson's method is more easily used when the limits of integration are fixed. This can be accomplished for both F_1 and F_2 by setting

$$x_j = \frac{t_j}{t_k} \quad \text{and} \quad x_k = t_k.$$

Equation A can then be written as

$$(1) \quad p_{jk} = \int_0^\infty \int_{-1}^1 x_k f(x_j x_k, x_k) dx_j dx_k - \int_{-\infty}^0 \int_{-1}^1 x_k f(x_j x_k, x_k) dx_j dx_k.$$

Before (3) and (4) in Appendix 1 can be applied to obtain p_{jk} , values of $\underline{n}$ and $\underline{m}$ must be chosen, (see Appendix 1) as well as a finite number (denoted by $\underline{a}$) to replace the infinite limits of integration appearing in F_1 and F_2 . These are two related problems. Writing (1) as

$$(2) \quad p_{jk} = \int_0^a \int_{-1}^1 x_k f(x_j x_k, x_k) dx_j dx_k - \int_{-a}^0 \int_{-1}^1 x_j f(x_j x_k, x_k) dx_j dx_k$$

it is clear that as a approaches positive infinity (2) will approach (1). However, for fixed n , the larger the value of a the less accurate is the numerical integration. This is due to the fact that the accuracy of the numerical integration is directly related to the size of the subdivisions a/n (denoted by h); and for fixed n , h increases as a increases. It is possible, of course, to increase n to compensate for a large value of a , but increases in n are paid for by increases in the time of calculation.

The criterion used to determine a is based on the knowledge that the variables in the integrand of F_1 and F_2 have marginal distributions which are normal. The probability that t_k will exceed a value of $3\sigma_{t_k}$ is therefore equal to .00135. However, in general not all of the region lying beyond 3σ in the t_k distribution will have $|t_k| > |t_j|$; the error, therefore, in setting $a = 3\sigma_{t_k}$ should not exceed .00135 and generally should be less. This criterion for determining a can be illustrated with the following example: If $Q_j = .2$, $Q_k = -.2$, and $C = .8$, then $T_j = .6$ and $T_k = 1.0$ and the limits of integration in equation A would be

$$p_{jk} = \int_0^4 \int_{-1}^1 f(x_j x_k, x_k) dx_j dx_k - \int_{-4}^0 \int_{-1}^1 f(x_j x_k, x_k) dx_j dx_k$$

In short, the criterion for determining a specifies that

$$(3) \quad a = 3 + \text{the larger of } (T_j, T_k) \text{ for } \sigma_{t_j} = \sigma_{t_k} = 1.$$

Since it is assumed that t_j and t_k have equal variances, $\underline{n}$ was set equal to $\underline{m}$ in the evaluations of equation A, and the value of $\underline{n}$ was selected which gave approximately three digit accuracy. Initial computations indicated that values of $\underline{h}$ not exceeding $\sigma_{t_j}/2$ would give three digit accuracy. Since $\underline{a}$ did not usually exceed $5\sigma_{t_j}$ in the evaluation of A, $\underline{n}$ was set equal to 10.

The accuracy of the numerical integration method obtained from the above values of $\underline{a}$ and $\underline{n}$ can be determined from Table 6. Table 6 contains three series of evaluations of p_{jk} , each series consisting of fixed Q_j and Q_k and varying values of C_1 . The initial p_{jk} value of each series refers to a condition in which C_1 is exactly midway between Q_j and Q_k , and considerations of symmetry require these p_{jk} values to equal one half. The three initial values of p_{jk} for each series are given in Table 4. These values of p_{jk} suggest that the numerical integration method underestimates p_{jk} , but that the results are approximately accurate to three digits.

The final sets of values of p_{jk} in each series may be used to estimate the accuracy of the computation for relatively large values of $\underline{a}$. If, for example, $Q_j = 1$, $Q_k = -1$, and $C_1 = 3$, then $\underline{a}$ will equal 7, and for $n = 10$, $h = .7$. It might be expected that for values of $\underline{h}$ this large, the accuracy of the p_{jk} estimates would be reduced considerably. The accuracy of these p_{jk} values can be assessed by comparing the asymptotic values of p_{jk} obtained from the unilateral equations.

The maximum value of p_{jk} in each series and the corresponding

asymptotic p_{jk} value are given in Table 5.

Inspection of Table 5 indicates that three digit accuracy is approximately maintained for $h = .65$ (or $a = 6.5$). This conclusion may be open to some criticism, however. If the numerical evaluations of p_{jk} are overestimating p_{jk} instead of underestimating this term, then the asymptotic value for each series would merely be reached at lower values of C_1 . Since no estimate exceeds the asymptotic values, then one could argue that the series was terminated too soon, that large enough values of C_1 were not taken. However, all indications are that the evaluations of p_{jk} are underestimates, particularly since large values of C_1 in series I produce smaller values of p_{jk} (see Table 6). Perhaps, a conservative conclusion would limit h to at most 0.6 for three digit accuracy.

APPENDIX 3

SOLUTION TO THE SCALE VALUES IN THE V THEORY USING THE PREDICTION METHOD

When the number of p_{jk} terms which are used in determining the V scale values exceeds $n-1$, the iteration scheme described in Chapter II must be used. In the prediction method employed in this study, $n-1$ p_{jk} terms are used to determine the scale values, and in this case the scale values in the V theory can be solved directly by a recursion method. The procedure is as follows. The p_{jk} terms are arranged in the order

$$p_{12}, p_{23}, p_{34}, \dots, p_{j-1,j}, \dots, p_{n-1,n},$$

and correspondingly, the scale values in the order

$$v_1, v_2, \dots, v_{j-1}, v_j, \dots, v_{n-1}, v_n.$$

The value of v_1 is arbitrarily set equal to some number, say 100, and the scale value of v_2 is then determined by

$$v_2 = \frac{v_1(m_{12}-n_1)}{n_1},$$

where n_1 denotes the number of times stimulus 1 was preferred and m_{12} , the number of times the pair (1,2) was presented. The remaining scale values are obtained from

$$v_j = \frac{m_{j-1,j}(v_{j-2} + v_{j-1})}{n_{j-1}\left(\frac{v_{j-2}}{v_{j-1}}\right) + n_{j-1} - m_{j-2,j-1}} - v_{j-1}.$$

Appendix 4

GLOSSARY OF SYMBOLS

- A - usually used in connection with equation A, a basic equation in the probabilistic Q theory. It also denotes $t_j - t_k/2$. (Chap. III, p. 48)
- B - usually used in connection with equation B, a basic equation in the probabilistic Q theory. It also denotes $t_k - t_j/2$. (Chap. III, p. 48)
- c_i - the location of the ideal point or stimulus for individual i at each moment.
- c - the momentary value of an ideal point without regard to a specific individual.
- C_i - the mean location of the ideal point of individual i.
- C' - the expected value of C_i over individuals.
- F_1, F_2, F_3 - terms in equations A and A'. (Chap. III, p. 47)
- I scale - the order of the s_{ij} values of the stimuli for individual i. The stimuli are ordered on this scale according to their preferability to the individual.
- J scale - a joint scale of stimuli and individuals in the Q theory. The stimulus values on this scale are denoted by Q_j and the individuals' locations by C_i . J scale occasionally refers to the stimulus values only on this scale.
- J1 scale - a J scale of the combined set of stimuli derived from certain initial considerations but subsequently rejected. (Chap. VII, p.161)
- J2 scale - a J scale of the combined set of stimuli; the numerical values of the stimuli were obtained by (1) deleting three stimuli from all unilateral judgments, and (2) using the direct and indirect estimates of each $Q_j - Q_k$ interval.
- J3 scale - the scale used in the Qd analysis.
- ICJ scale - a scale derived from Thurstone's Law of Comparative Judgment usually with the Case 5 assumptions; the Law of Comparative

Judgment is also referred to as the ICJ theory. (Chap. II, p. 25)

n_{jk} - the number of replications of stimuli j and k .

n - when n_{jk} is equal for all j and k , the subscripts are dropped. n usually denotes number of replications of the stimuli for a single individual, although occasionally it denotes the total number of replications within and between subjects.

n - the number of stimuli. In Appendices 2 and 3 it refers to the number of subdivisions used in the numerical integration.

N - the number of subjects

N^* - the size of the population from which the subjects are drawn.

n_{jk} - the total number of times stimulus j is preferred to stimulus k . This term usually includes all the judgments of all subjects.

n_j - the number of times stimulus j is preferred over all other stimuli.

$N(C_1, \sigma)$ - denotes a frequency function of a normal distribution having a mean of C_1 and a standard deviation of σ .

p_{jk} - the probability that stimulus j will be preferred to stimulus k . This probability is termed a preference probability. It also denotes the proportion of times j was preferred to k in a given set of data.

p_{jk}^* - the preference probability which is predicted or postdicted by a theory.

$p_{jk,1}$ - when it is necessary to distinguish the p_{jk} values of different individuals the i subscript is added. In this case, p_{jk} would represent the expected value of $p_{jk,1}$.

Q theory - refers to the Coombs Quadrant I theory. (Chap. I, p. 11) Usually the Q theory in this study refers to the probabilistic version of the theory.

Q_j - the mean value of stimulus j in the Q theory. This also refers to the scale value of the stimulus on the J scale.

q_j - the momentary value of stimulus j in the Q theory

Q_a - denotes a Q theory analysis or scale of the R stimuli in which (1) one stimulus was deleted from all unilateral judgments, and (2) the indirect and direct estimate of each $Q_j - Q_k$ interval were used.

Q_b - denotes a Q theory analysis or scale of the R stimuli in which (1) one stimulus was deleted from all unilateral judgments, and

(2) only the direct estimate of each $Q_j - Q_k$ interval was used.

Q_c - denotes a Q theory analysis or scale of the R stimuli in which (1) three stimuli were deleted from all unilateral judgments, and (2) only the direct estimate of each $Q_j - Q_k$ interval was used.

Q_d - a Q theory analysis or scale of the combined set of stimuli in which (1) three stimuli were deleted from all unilateral judgments, and (2) only the direct estimate of each $Q_j - Q_k$ interval was used.

SST - strong stochastic transitivity. This is a constraint imposed on the preference probabilities associated with three stimuli. (Chap. I, p. 14)

S_j - stimulus j

$$s_{1j} = |c_1 - q_j|$$

$$S_{1j} = |C_1 - Q_j|$$

$$SS_D = \sum_{j>k} (\theta_{jk} - \theta_{jk}^*)^2; \text{ called the residual sum of squares.}$$

$$t_{1j} = c_1 - q_j = t_j$$

$$t_j' = c - q_j; \text{ In Chapter III, p.45 it denotes deviations about the mean value of } t_j.$$

$$T_{1j} = C_1 - Q_j = T_j$$

$$T = T_j^2 - T_j T_k + T_k^2$$

$$T_j' = C' - Q_j$$

U theory - a theory implied by the quadruple condition. The ICJ theory is used interchangeably with the U theory. (Chap. II, p. 21)

u_j - the momentary values of stimulus j in the U or ICJ theory.

$u_{j,i}$ - when the mean stimulus value U_j depends on the individual, the subscript i is added to u_j and U_j .

U_j - the mean value of u_j

$U_{j,i}$ - when the value of U_j depends on the individual, the i subscript is added.

U_1 - a U theory analysis or scale of the combined set of stimuli based on the reproduction method.

U_2 - a U theory analysis or scale of the combined set of stimuli based on the prediction method.

- U3 - a U theory analysis or scale of the R stimuli based on the reproduction method.
- U4 - a U theory analysis or scale of the R stimuli based on the prediction method.
- V theory - a theory implied by the triple condition. A scale derived from this theory is called a V scale. (Chap. II, p. 23)
- v_j - the scale value of stimulus j on the V scale.
- V1 - a V theory analysis or scale of the combined set of stimuli based on the reproduction method.
- V2 - a V theory analysis or scale of the combined set of stimuli based on the prediction method.
- V3 - a V theory analysis or scale of the R stimuli based on the reproduction method.
- V4 - a V theory analysis or scale of the R stimuli based on the prediction method.
- WST - weak stochastic transitivity. This refers to a constraint imposed on the preference probabilities associated with three stimuli. (Chap. II, p. 20)
- $x_j = t_j/t_k$
- X_{jk} = the normal deviate corresponding to p_{jk} .
- $\theta_{ij} = \arcsin \sqrt{p_{ij}}$, measured in degrees.
- $\succ$ = is preferred to

BIBLIOGRAPHY

- Abelson, R. & Bradley, R. A. A 2×2 factorial with paired comparisons. Biometrika, 1954, 10, 487-502.
- Atkinson, J. W. Motivational determinants of risk-taking behavior. Psychol. Rev., 1957, 64, 359-372.
- Bernoulli, D. Specimen theoriae novae de mensura sortis. Commentarii Academiae Scientiarum Imperiales Petropolitanae, 1738, 5, 175-192. (Trans. by L. Sommer in Econometrica, 1954, 22, 23-36.)
- Bliss, C. I., Greenwood, M. C., & White, E. S. A rankit analysis of paired comparisons for measuring the effect of sprays on flavor. Biometrics, 1956, 12, 381-403.
- Bock, D. R. Remarks on the test of significance for the method of paired comparisons. Psychometrika, 1958, 23, 323-334.
- Bradley, R. A. & Terry, M. E. Rank analysis of incomplete block designs: I the method of paired comparisons. Biometrika, 1952, 39, 324-345.
- Bradley, R. A. Some statistical methods in taste testing and quality evaluation. Biometrics, 1953, 9, 22-35.
- Bradley, R. A. Incomplete block rank analysis: on the appropriateness of the model for a method of paired comparisons. Biometrics, 1954, 10, 375-390. (a)
- Bradley, R. A. The rank analysis of incomplete block designs: II additional tables for the method of paired comparisons. Biometrika, 1954, 41, 502-537. (b)
- Bradley, R. A. Rank analysis of incomplete block designs: III some large-sample results on estimation and power for a method of paired comparisons. Biometrika, 1955, 42, 450-470.
- Bush, R. R. & Mosteller, F. Stochastic models for learning. New York: John Wiley, 1955.
- Cartwright, D. & Festinger, L. A quantitative theory of decision. Psychol. Rev., 1943, 50, 595-621.

- Chandler, A. R. Beauty and human nature. New York: Appleton-Century-Crofts, 1934.
- Cochran, W. G. Sampling techniques. New York: John Wiley, 1953.
- Coombs, C. H. Psychological scaling without a unit of measurement. Psychol. Rev., 1950, 57, 145-158.
- Coombs, C. H. A theory of psychological scaling. Engineering Res. Bull., May, 1952, 34. KRI, University of Michigan.
- Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. J. exp. Psychol., 1958, 55, 1-7.
- Coombs, C. H., Greenberg, M., & Zinnes, J. A double law of comparative judgment for the measurement of attitudes and preferences. Unpublished manuscript, 1959.
- Cramer, H. Mathematical methods of statistics. Princeton: Princeton Univer. Press, 1951
- Davidson, D. Stochastic modifications of decision theory. Unpublished manuscript, Univ. of Michigan, Seminar on the Application of Mathematics to the Social Sciences, Dec. 20, 1956.
- Davidson, D. & Suppes, P. Decision making, an experimental approach. Stanford: Stanford Univer. Press, 1957.
- Davidson, D. & Marschak, J. Experimental tests of a stochastic decision theory. Technical Report 17, Stanford Univer., Applied Mathematics and Statistical Laboratory, July 25, 1958.
- Davidson, D. Learning transitivity of choice in paired comparisons. Unpublished manuscript, Univer. of Michigan, Seminar on the Applications of Mathematics to the Social Sciences, April 7, 1959.
- Debreu, G. Stochastic choice and cardinal utility. Econometrika, 1958, 26, 440-444.
- Dykstra, O. A note on the rank analysis of incomplete block designs--- applications beyond the scope of existing tables. Biometrics, 1956, 12, 301-306.
- Edwards, W. Probability-preferences in gambling. Amer. J. Psychol., 1953, 66, 349-364.

- Edwards, W. The theory of decision making. Psychol. Bull., 1954, 51, 380-417.
- Estes, W. K. Toward a statistical theory of learning. Psychol. Rev., 1950, 57, 94-107.
- Estes, W. K. A model for choice behavior. Unpublished manuscript, Univer. of Michigan, Seminar on the Application of Mathematics to the Social Sciences, Oct. 23, 1958.
- Eysenck, H. J. A critical and experimental study of colour preferences. Amer. J. Psychol., 1941, 54, 385-394. (a)
- Eysenck, H. J. "Type"-factors in aesthetic judgments. Brit. J. Psychol., 1941, 31, 262-270. (b)
- Farnsworth, D. Manual: The Farnsworth Munsell 100 hue test. Baltimore: Munsell Color Company, 1957.
- Ford, L. R., Jr. Solution of a ranking problem from binary comparisons. Amer. Math. Mon. Herbert Ellsworth Slaughter Memorial Papers, 1957, 28-33.
- Georgescu-Roegen, N. The pure theory of consumer's behavior. Quart. J. Econ., 1936, 50, 545-593.
- Granger, G. W. An experimental study of colour preferences. J. gen. Psychol., 1955, 52, 3-20.
- Guilford, J. P. The affective value of color as a function of hue, tint, and chroma. J. exp. Psychol., 1934, 17, 342-370.
- Guilford, J. P. Psychometric methods. (1st Ed.) New York: McGraw-Hill, 1936.
- Guilford, J. P. A study in psychodynamics. Psychometrika, 1939, 4, 1-21.
- Guilford, J. P. There is system in color preferences. J. opt. Soc. Amer., 1940, 30, 455-459.
- Gulliksen, H. A generalization of Thurstone's learning function. Psychometrika, 1953, 18, 297-307.
- Gulliksen, H., & Tukey, J. Reliability for the law of comparative judgment. Psychometrika, 1958, 23, 95-110.
- Guttman, L. An approach for quantifying paired comparisons and rank order. Ann. Math. Statist., 1946, 17, 144-163.

- Hicks, J. R., & Allen, R. G. D. A reconsideration of the theory of value. Economica, 1934, 14, 52-76, 196-219.
- Hull, C. L. Principles of behavior. New York: Appleton-Century-Crofts, 1943.
- Jackson, J. E., & Fleckenstein, Mary. An evaluation of some statistical techniques used in the analysis of paired comparison data. Biometrics, 1957, 13, 51-64.
- Kendall, M. G. Further contributions to the theory of paired comparisons. Biometrics, 1955, 11, 43-62.
- Landau, H. G. On dominance relations and the structure of animal societies. Bull. Math. Biophysics, 1951, 13, 245-262.
- Lewin, K. Field theory in social science. New York: Harper, 1951.
- Luce, R. D., & Raiffa, H. Games and decisions. New York: John Wiley, 1957.
- Luce, R. D. A probabilistic theory of utility. Econometrica, 1958, 26, 193-224.
- Luce, R. D. Individual choice behavior: a theoretical analysis. New York: John Wiley, 1959.
- Mosteller, F. Remarks on the method of paired comparisons: III a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. Psychometrika, 1951, 16, 207-218.
- Mosteller, F. The mystery of the missing corpus. Psychometrika, 1958, 23, 279-289.
- Munsell book of color. (Abridged ed.) Baltimore: Munsell Color Company, 1945.
- Newhall, S. M., Nickerson, D., & Judd, D. B. Final report of the O.S.A. subcommittee on the spacing of the Munsell colors. J. opt. Soc. Amer., 1943, 33, 385-418.
- Norman, R. D. & Scott, A. Color and affect: a review and semantic evaluation. J. gen. Psychol., 1952, 46, 185-225.
- Papandreou, A. G., Sauerlander, O. H., Brownlee, O. H., Hurwicz, L., & Franklin, W. A test of a stochastic theory of choice. Univer. of Calif. Publications in Economics, 1957, 16, 1-18.
- Pearson, E. S., & Hartley, H. O. Biometrika tables for statisticians. V. I. Cambridge: Cambridge Univer. Press, 1954.

- Pressey, S. L. The influence of color upon mental and motor efficiency. Amer. J. Psychol., 1921, 32, 326-356.
- Ramsey, F. P. Truth and probability. In F. P. Ramsey, The foundations of mathematics and other logical essays. New York: Harcourt Brace, 1931.
- Rotter, J. B. Social learning and clinical psychology. New York: Prentice Hall, 1954.
- Scarborough, J. B. Numerical mathematical analysis. (4th ed.) Baltimore: Johns Hopkins Press, 1958.
- Shepard, R. N. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. J. exp. Psychol., 1958, 55, 509-523.
- Terry, M. E., Bradley, R. A., & Davis, L. L. New designs and techniques for organoleptic testing. Food Technology, 1952, 6, 250-254.
- Thompson, G. L. Lectures on game theory, Markov chains, and related topics. Sandria Corporation Monograph, March, 1958.
- Thurstone, L. L. A law of comparative judgment. Psychol. Rev., 1927, 34, 273-286.
- Thurstone, L. L. The measurement of opinion. J. abnorm. & Soc. Psychol., 1928, 22, 415-430.
- Thurstone, L. L. The learning function. J. gen. Psychol., 1930, 3, 469-493.
- Tinbergen, N. The study of instincts. London: Oxford Univer. Press, 1951.
- Tolman, E. C. Principles of performance. Psychol. Rev., 1955, 62, 315-326.
- Torgerson, W. S. Theory and methods of scaling. New York: John Wiley, 1958.
- Vail, S. V. A stochastic model of utilities. Unpublished manuscript Univer. of Michigan, Seminar on the Application of Mathematics to the Social Sciences, April 23, 1953.
- Von Neumann, J., & Morgenstern, O. Theories of games and economic behavior. (1st Ed.) Princeton: Princeton Univer. Press, 1944.