MATHEMATICAL MODELS APPLIED TO SUBJECTIVE APPRAISALS IN ENVIRONMENTAL ENGINEERING

by

A. G. STAINSBY

Thesis submitted for the degree of

Doctor of Philosophy

in

THE CITY UNIVERSITY

LONDON

Department of Electrical and Electronic Engineering

May 1975

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

List of Figures and Tables (Cont.)

ACKNOWLEDGEMENTS

The Author wishes to acknowledge the cooperation of Mr. Mainwaring and of other members of the Lighting Department in the work leading to the publication of Paper 1. He would also like to thank his many other colleagues in the G.E.C. Hirst Research Centre for cooperation and helpful discussions.

He wishes particularly to thank the co-authors of Papers 2, 3, 4 and 5, without whom the work would not have been possible. Discussions on the work were wide-ranging, always interesting and nothing if not frank.

He wishes also to thank the City University and the Department of Electrical and Electronic Engineering in which this work was written. He also acknowledges the use made of the University computing facilities and wishes to thank Mr. J. Snell of the Mathematics Department for writing the subroutine INTEG used in Gaussian quadrature in Program TRYA.

DECLARATION

In the work leading to the joint paper with Mr. Mainwaring, (Paper 1) the Author initiated the work and took part in its planning and carrying out, but made negligible contribution to the optical or mechanical design of the lanterns and mountings. The analysis, and that part of the written paper relating to it, is his own work; the conclusions were the result of discussions within the Lighting Department of Osram (G.E.C.) Ltd., of which the Author was Research Manager.

In the work leading to the papers published jointly with M. Halstead, D. Morley and D. Palmer (Papers 2, 3, 4 and 5), the Author carried out the bulk of the analysis and developed the mathematical model used in it. Most of the critical discussion of the C.I.E. method for colour rendering in Part 2 of Paper 3 was also his. Free discussion within the group was important in clarifying some of the Author's ideas and in bringing the work to a successful conclusion.

The ideas expressed in this thesis are the Author's own.

SUMMARY

Mathematical models appropriate to different types of psychophysical measurement are considered. It is stressed that the model should arise out of what is believed to be the type of observation made by the observer. This is in contrast to the type of model which is introduced simply because its equations can be made to fit (rather than explain) the data. In considering the type of model required, consideration must be given to the type of scale likely to arise. This consideration leads to a subdivision of S.S. Stevens' ratio scale into two classes, here called the geometric and arithmetic classes. It is shown that models based on these two scales will explain some of the differences noted by Stevens between his prothetic and metathetic continua.

The model used by the Author in analysing categorical observations of differences in a published paper, is discussed in greater detail than in the original paper and the method of analysis is extended to include variability of the observer in two dimensions i.e. on the colour surface. It is shown that average values of observations almost identical to those obtained for variability along only one axis are predicted, but the distribution of scores between categories is different. Comparison of the predictions with actual observations indicate

that the bivariate distribution is an appropriate model for some colours.

The model used for analysis of colour differences is put forward as one appropriate for analysis of observations near to the limit of observation, and a theory is proposed.

Five published papers are included in which models have been used in the solution of environmental engineering problems. One relates to street lighting, the others relate to colour rendering of lamps.

CLAIMS

1. It is believed that Paper 1 was the first paper to give quantitative relations between the subjective factors glare, brightness and uniformity and the physical characteristics of a street lighting distribution.
2. It is believed that the method of analysis of data used in Paper 3 was novel and the Author knows of no other instance of that model having been used.
3. The Author has found no other reference to an explanation for S.S. Stevens' 'prothetic' and 'metathetic' continua, nor to the necessity of dividing the ratio scale into two distinct classes. (Section 4)
4. The method of analysis of data arising from a bivariate distribution is novel. (Section 6.4)
5. The Author believes that the proposal made in Section 7 to use the model of Paper 3 for understanding the 'threshold' is novel and that it could prove fruitful.

1. INTRODUCTION.

The purpose of engineering is the satisfying of human needs. Some times these needs are very obvious for example in providing enough light to enable people to read. Success or failure is readily assessed by taking a book and trying for oneself. When that initial need is satisfied, future progress is not so easy. The provision of more light is easy enough at a cost, but finding out whether it is sufficient, too much or too little, and finally obtaining information adequate for a cost benefit study presents considerable difficulties. Such studies, however, are part and parcel of the work of a practising engineer. The assessment of the end product is an essential part of the development process. Too many engineers neglect this part of the development process with the well known result that follies and shortcomings are perpetuated from one development to the next.

Further development requires obtaining information directly from observers. This can be done either by questioning them directly or by observing the pattern of their behaviour under different conditions. In either case, the information will be characterized by its great diversity, and it is the work of the engineer to obtain from it information in the form of relations between the conditions imposed on the observers and their reactions. To do this, it is necessary to postulate a mathematical model of the behaviour of the ensemble of observers in order to interpret their observations. The choice of model is critical,and must be in accord with a physical process that is believed to occur. If on analysis the results are in accord with the model, confidence is gained to apply it in another situation: if not, it must be discarded and a new process postulated.

This thesis discusses the types of model that can be used and explains a particular model used by the Author to deal with threshold problems.

1.1 Street Lighting.

The work of the Author in this field of study started with a street lighting problem (Paper 1.). At the time of its writing, the company of which the Author was Research Manager had developed a new lamp, the high pressure sodium lamp, for street lighting purposes. It differed from street lighting lamps available at that time, viz. the low-pressure sodium lamp and the colour-corrected high-pressure mercury lamp, in several respects which were believed to be advantageous. These differences were (1) It had excellent colour rendering, particularly in the red region of the spectrum. (2) It had high efficacy, about 100 l/w compared with 50 l/w for high-pressure mercury lamps. (3) It was a compact source, being a discharge tube of only 6 mm in diameter and about 60 mm long. This compared with the large fluorescent envelope of the colour-corrected mercury lamp some 250 mm long and 150 mm in diameter.

This third difference enabled the light distribution from a street lighting lantern to be closely controlled and in particular, it enabled light distributions having a very rapid change of luminous intensity with angle to be achieved in a simple lantern. This fall off of intensity is loosely referred to as the 'cut-off'. It was the advantages or otherwise of this facility which were investigated.

From previous observations of street lighting installations, it was expected that the following trends would be seen. As the angle of the beam was lowered, the installation would appear less glaring, but at the same time the apparent brightness would decrease and the surface of the road would look increasingly patchy. It was also thought that the visibility of stationary cars, pedestrians on the pavement and other hazards might be affected. It was therefore decided to investigate these factors from the point of view (literally) of a car driver, by errecting a street lighting installation and observing glare, brightness, uniformity and visibility as the cut-off was adjusted.

A special street-lighting installation was errected in East Lane, Wembley. It was fitted with specially adjustable lanterns so that the vertical and horizontal angles between the beam and the longitudinal axis of the road could be separately controlled. Laboratory tests on these lanterns were made so that their physical performance would be accurately known. The lanterns were errected so that they could be adjusted from ground level; thus by installing an operator at each column, the effect of a steady change of beam angle could be seen.

To investigate the effects, groups of about 20 observers were called on to view the installation from the same level as that of a car driver and with an eye shield to simulate the effect of a car roof. The observers were asked to record their results in one of five verbal categories. These were arranged to provide two categories above and two below an essentially neutral one. For example, to record their opinion of the uniformity of road brightness, they could record (1) Very poor. (2) Poor. (3) Fair. (4) Good. (5) Excellent. Observations were taken on both dry and wet nights, for in the rain the reflection from a road surface becomes almost specular and the appearance of the installation is completely changed. The results were analysed by the Author on the assumption that the observers treated the categories as of equal width. This assumption was to be justified or otherwise by the distribution of the results. The results were analysed for inter-relationships between the factors observed. It was found that, within the limits of the experiment, there was a linear relationship between glare, brightness and uniformity, and that these combined together to give the overall impression of the quality of the installation. Furthermore, the magnitude of the glare was related logarithmically to the intensity received at the eye from the first lantern in the field of view.

The technique of analysis used was straightforward. The first procedure was to examine the data by plotting frequency distributions, to see whether normality was a reasonable assumption, then to take means and variances. The variances were first examined, and where these differed markedly from the average value, the original data was re-examined. This enabled some anomalous observers to be found and eliminated. For example, where an observer saw no changes in one of the factors for all variations of the installation, it was assumed that that observer was not a member of the normal population of observers, and his results were eliminated. It is sometimes argued that since such an observer existed, his results should be used. However, by examining the performance of other observers, the general distribution can be estimated and, since there are only say 20 observers, it is logical to eliminate any observer whose performance is clearly outside the one in twenty limits.

It should be emphasized that this examination of data was an important and essential part of the analytical process. The next stage was to plot graphs of each set of observations against the expected physical correlates and possibly some less expected ones. Unless the correlation appears obvious graphically, it is seldom worth carrying out a formal analysis. Also unless some physical reason is suspected for the correlation, it must be regarded with grave suspicion. Finally, a three factor analysis was carried out to find the relation of the physical parameters of the lantern distributions to the observations.

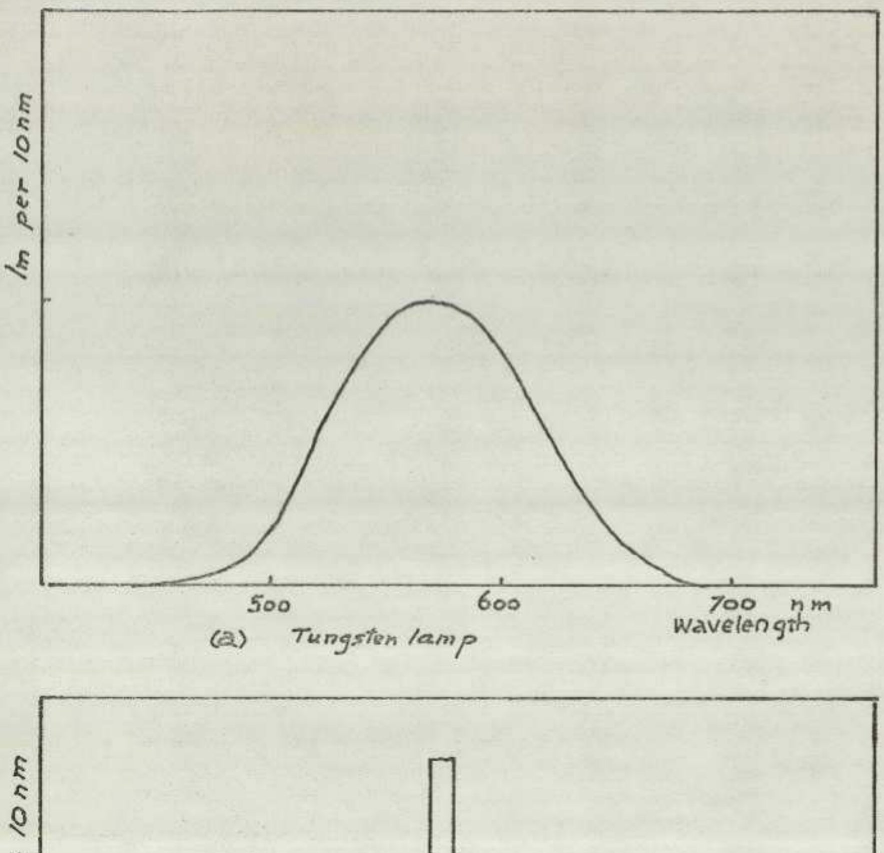
In the group of observers, about half were lighting engineers and the rest were not. No difference was found in the observations of the two groups. Problems occurred which were associated with the ends of the category scale. When an extreme situation was presented, e.g. one of extreme glare, all observers would tend to use the highest category. As a result the variance of the results was small. It is assumed that if there had been a higher category, many observers would have placed the situation in the higher category. The average found for such situations would thus be low. Apart from these cases, the results were found to be sufficiently normally distributed for analysis of variance and linear regression analysis to be applied. The fact that linear relationships were obtained with small residuals justified the assumption of equal category widths. Linear relationships were found between the factors over the range investigated and the relative importance placed on them by observers was discovered.

It is thus seen that an engineering problem had been solved by the process of asking the customer. There is no other way open to the engineer to justify his designs. It was satisfactory that the results validated general street lighting practice, which has been established largely by the qualitative observations of highly skilled lighting engineers. A quantitative relation had been established, for one type of lighting installation, between glare, brightness and uniformity. Further experiments on other installations and at other levels of illuminance would be necessary in order to assess the generality of the conclusions. Such further work was beyond the limits of expenditure that could be justified in the company at that time.

1.2 Fluorescent Lighting.

The second major problem involving the Author in psychophysical testing and analysis also arose from a lighting problem. This forms the main topic of this thesis.

Since the introduction of the fluorescent lamp for interior use in the early 1930 s, lighting engineers and colour experts have attempted first to control and later to define the colour rendering properties of lamps. Before the invention of the fluorescent lamp, the only lamps available had a continuous spectrum close to that of a full radiator. Although the generally used tungsten lamp introduced colour distortions, these were in a known direction so that users became used to designing colour schemes either for indoors or for out doors, and allowance could easily be made for the lack of the blue light



Luminous output as a function of wavelength.
(a) Tungsten lamp
(b) Fluorescent lamp

Fig 1.

in radiation from tungsten at a temperature of 3200 K. Furthermore, the general public was used to the difference, and scarcely appreciated that it was there.

With the introduction of the fluorescent lamp, all this was changed. A range of lamp colours could be obtained by altering the phosphor coating. The colours could have the appearance of 'white' light of a colour temperature ranging from that of incandescent tungsten up to that of north daylight. However, because of the line emission from the mercury discharge and also because of the desire of lamp makers to achieve the highest efficacy consistent with acceptable rendering of colours, the spectra of fluorescent lamps differed greatly from those of a full radiator at the corresponding colour temperature. In fig. 1., the spectra of two lamps having the same colour temperature are compared.

The first attempt at standardization was to define for each lamp type (a) the colour appearance and (b) the luminous output in each of eight bands in the visible spectrum. The colour appearance is the colour that a white surface would appear to have when illuminated by the lamp. As makers were in the main aiming at white light, the appearance could be specified in terms of the correlated colour temperature of the source, since it was assumed that a white surface would appear white under a full radiator. Putting this crudely, it should mean that a white specimen would look the same as when illuminated by a full radiator of the defined temperature. Thus for example a lamp to simulate north light would have a colour appearance corresponding to 6500 K. To define colour rendering, the first attempts were to specify the proportion of the luminous output which was emitted in each of eight spectral bands. This defined the lamp reasonably well so that two lamps having the same output in each band might be expected to be interchangeable and to render coloured samples in the same way. The method was agreed internationally and became the C.I.E. (Commission Internationale d'Eclairage) approved method of defining colour rendering in 1948.

This method was accepted because no other was available, but it was considered objectionable on two counts. Firstly, it did not define the colour rendering, but only controlled it. Secondly, by defining the spectrum of lamps that could be made at the time of publication and using the technology which was available at the time, the method could inhibit or even preclude future developments in phosphors and lamps. This was felt strongly in countries where the national specifications, which were usually based on internationally agreed standards, have the force of law. For this reason the international experts endeavoured to devise a method of measuring the colour rendering properties of sources for interior use. Two methods were considered. In this country, Crawford (5) attempted to extend the band method and justified his procedure by extensive subjective trials. On the continent, however, the method finally accepted by the C.I.E. (3) was developed mainly on the basis of paper studies of data on existing lamps and coloured samples. These methods are discussed in the general introduction and in part 2 of Paper 3.

In this method, eight specially selected test-colours are considered. It is assumed that these are viewed when illuminated by the source under test and then under a reference source. The reference source approximates to a full radiator or to daylight and has the same colour temperature as the source under test). If the colour of the source under test is not identical to that of the reference source, some colours will change their appearance. In C.I.E. publication 13 (3), formulae and tables are given to enable the colour shift of each test colour to be calculated on the uniform chromaticity scale. The colour rendering is then derived from the average of the shifts of the eight coloured samples. In a series of trials, the Author in collaboration with three other members of the British national experts committee on colour rendering, attempted to test the method of Publication 13 and to see whether it defined a unique figure with respect to a lamp, and whether the figure found was in agreement with the opinion of groups of observers viewing scenes illuminated by measured lamps. This work is reported in Papers 2,3,4,5. It should here be mentioned that the initial division of labour among the authors of these papers was to be: Miss Halstead, provision of lamps and all spectroradiometric data; Mrs Morley, arrangement of the test rooms and organization of observers and trials; Dr Palmer to act as consultant particularly on vision and the physiology of the eye. The present author was to deal with data analysis. All cooperated in the design of the experiments.

This approximate division of labour continued through Papers 2, 3 and 4, but after that paper, the boundaries of the work became more diffuse. Furthermore during the writing of Paper 5, alternative methods of analysis were advocated, in fact those explained in Chaps 9 and 10 of Torgerson's book(27), which had been rejected as inappropriate to this problem during the work leading up to Paper 3. These methods are criticized in section 3.2.

Paper 3 gives the background to the work and fully discusses and criticizes the C.I.E. method. It also contains the Author's theory on which the analysis was based. The conclusions of that paper are also the most extensive. Paper 2 was preliminary and, as will be seen, had its theoretical analysis based on the idea of random noise. This has since been found to be unnecessary and possibly misleading. Paper 4 was an attempt to include strong test samples and a more general scene in order to relate the special rendering indices of individual test samples to the general colour rendering index. Finally, Paper 5 was a study of high pressure lamps. This was a crucial experiment, since it had been the contention of the authors of the above papers that much of the apparent sense in the method arose from the choice of fluorescent lamps, rather than from its own merits. Unfortunately, data obtainable from the lamps available did not allow firm conclusions to be drawn.

In this thesis, the theoretical background to the analysis will be discussed.

2. PSYCHOPHYSICAL METHODS.

The aim of psychophysical methods is to link commonly used descriptions of human sensations with other such descriptions or with physical quantities. Many people would say that it is the sensations themselves that are under study, but that cannot be proved since verbal or similar communication is essential in order to specify the sensation under study. It is therefore impossible to say whether what one subject understands by a sensation, for example of coldness, is the same as that understood by another. Thus although we shall refer to correlation of responses with physical stimuli, it must be remembered that verbal communication is nearly always an essential link in the chain. The aim is to use the human as a measuring instrument and to establish correlates with sensations such as loudness, brightness and coldness. Closely related to these methods and frequently overlapping in the use of the same techniques, are psychological methods in which the aim is to find out qualities in the observer. Thus, an experimenter wishing to rate or order a series of observers in their sensitivity to sounds, light or heat, might use the same methods as an experimenter wishing to establish scales of loudness, brightness or coldness, but his method of experimental replication and analysis would differ. In this section, the aim is to study the use of the observer as a measuring instrument for finding relationships between sensations and stimuli. We shall therefore consider only psychophysical methods. To avoid continually having to assemble groups of observers to estimate the subjective sensations due to physical effects, an attempt is usually made to find a physical property that is related to the sensation to be evaluated. Then if a relationship can be established, the physical property can be used as if it were directly related to the subjective response. An extreme example of this is found in the definition of the candela which is one of the base units of the S.I. system. Although this is a purely physical unit relating to the directional radiation from a black body under specified conditions, the spectral output has to be modified by the luminosity curve of the normal eye in order to convert to the candela. The error is commonly made of assuming that observed sensations of brightness and light would relate directly to the candela. However, it is well known that the luminosity curve of the normal eye is only an average, and that the curve for one particular eye depends on the illuminance at it and the sensation of brightness depends also on the adaptation. It is thus directly related to the luminance only under carefully controlled conditions. Sensitivity varies from one person to another. It is therefore not to be expected that there will be an absolute relationship between a human response and a physical stimulus. However it is reasonable to expect that, if all other factors can be kept constant, the response and the stimulus will be related by a monotonically increasing function provided that the stimulus is not so excessive that saturation is reached.

It should be emphasized that the response to a stimulus such as a light signal is an extremely indirect one. It is not practicable to probe into the nerve cells at the back of the human eye in order to get a more direct signal. But even if this could easily be done, such a response would not be the one required. Firstly, it would not take into account the effect of the narrowing of the aperture of the eye when exposed to light, and secondly, it would not include the translation carried out in the brain. What we need to know is what the observer thinks he sees, as this is what will determine his sensation. However, he must be influenced as little as possible by extraneous effects, otherwise we shall measure other factors in addition to those of interest.

To make use of our observations, it is necessary to set up some form of scale. This is the essence of measurement. This has been described by S.S. Stevens, paraphrasing N.R. Campbell, "as the assignment of numerals to objects or events according to rules". (17)

2.1 Types of Scale.

S.S.Stevens (17,18) has distinguished four types of scale which are used in measurements as defined above. These are

(1) Nominal Scale; in which the numbers are used only as labels and give no indication of any property of the object or class so numbered. Few arithmetic operations can be done with such a scale. When it is applied to classes, it enables the most frequently occurring class (the mode) to be determined.

(2) Ordinal Scale; in which numbers are assigned in accordance with the occurrence of some property, for example the arrangement of students in the order of the results of an examination. Nothing more is implied by this type of scale than rank order. The statistics that apply to this scale in addition to the mode, are the median and percentiles. The scale may be transformed by the equation

$$x' = f(x)$$

where $f(x)$ is any monotonically increasing function.

(3) Interval Scale; in which numbers are assigned to points on a continuum such that there is a linear relationship between the resulting scale and the magnitude of some property of the class. This is a 'true' scale which may be transformed according to the

equation

$$x' = ax + b$$

where a and b are constants.

Thus equal intervals on the scale represent equal differences in the magnitude of the property. To this type of scale all the usual statistical processes such as calculation of arithmetic mean, standard deviation, t-test and F-test may be applied. Only processes involving a zero are excluded, since the zero is not an invariant under transformation. Examples of this type of scale are temperature in degrees Celsius and potential energy.

(4) Ratio Scale. This scale has all the properties of the interval scale and also a natural origin. Thus, when the point on the scale is zero, the property is zero. The only transformation permissible for this scale is

$$x' = ax$$

where a is a constant.

All statistics applicable to the other types of scale apply, but in addition, the geometric mean and the coefficient of variation apply also. Typical examples of this type of scale are those of length, mass and absolute temperature.

Stevens has summarized the properties in various papers (17, 22, 23). For the purposes of this thesis, they may be summarized as follows:-

Scale	Permissible Statistics	Measure of Location
Nominal	Number of cases. χ^2 test.	Mode.
Ordinal	Percentiles. Order correlation.	Median.
Interval	Standard deviation. t-test. F-test.	Arithmetic mean.
Ratio	Coefficient of variation.	Harmonic mean. Geometric mean.

Another scale, is the ordered metric. This has been used by Coombs (4) and his co-workers for psychological testing. In his method, he asks a subject to consider a number of factors or statements and to say to what extent he is in agreement. If it is assumed that the statements lie on a continuum and the attitude of the subject is somewhere in the middle, then clearly he will disagree with both the extreme statements. By questioning a number of subjects and assessing their agreement or disagreement with the statements it is possible to place the subjects and the statements in order. The resulting scale is called an ordered metric and the process by which the results are obtained is described as unfolding. The method is essentially one for rating people and will not be considered further; but it is of interest in that it employs 'unfolding' in order to deduce a true scale from a number of ambiguous

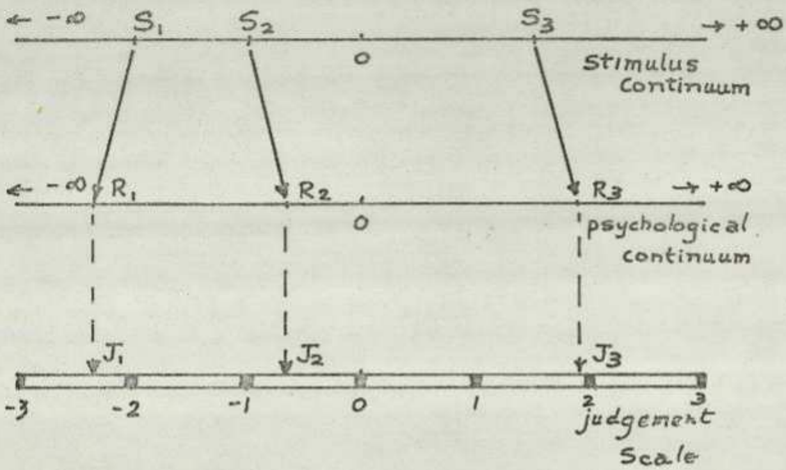


Fig 2

Stimuli S_1; S_2, S_3 give rise to discriminal processes R_1, R_2, R_3 on the psychological continuum. The observer says that these occur at J_1, J_2, J_3 on the judgement scale.

observations. Unfolding using a different method is the basis of the Author's analysis of colour differences discussed in Paper 3 and in Section 6.

2.2 The Psychological Continuum.

We cannot assume that the judgement given by an observer necessarily represents his true internal response to a stimulus. There must exist a translation process between them. To discuss this, let us assume that the set of all stimuli can be represented by a continuous straight line stretching from $-\infty$ to $+\infty$ (fig.2). (This represents the simplest case; multidimensional stimuli will be considered in a later section) There corresponds to this a psychological continuum on which the response R to any stimulus S occurs. Thus the stimuli S_1, S_2, S_3 on the stimulus continuum give rise to responses R_1, R_2, R_3 on the psychological continuum. These responses will be translated by the observer who will give judgements J_1, J_2, J_3. The judgement will not usually be on a continuum since the subdivision of the judgement scale is limited. There will thus be a quantizing error in making the judgement in addition to any errors due to rationalization on the part of the observer.

In designing a psychophysical experiment, every effort is made to deduce the value of the response R from the judgements J given by the observer. To achieve this, great care must be taken in framing questions to ensure that the observer understands unambiguously what he is being asked to judge. The questions must also be free from physical or emotional bias, but where this is impossible, the sequence of questions and replications should be planned to balance out the effect.

In his writings, L.L.Thurstone (26) uses the expression 'discriminal process' for the response on the psychological scale. He defines it as the process by which 'the organism identifies, distinguishes or reacts to stimuli'. The main advantage of the use of the term is to distinguish it from the overt 'judgement' which includes the processing and translation which is carried out in the brain. In order to reduce any rationalization process, J.C.Stevens et.al. (15) and others have used the method of cross-modality matching in which the subject reacts to one stimulus by some non-verbal act such as squeezing a hand dynamometer. The pressure on the dynamometer is then taken as a measure of his reaction to the stimulus. This then takes the place of a judgement.

The problem remains to evaluate the positions of the reactions or discriminal processes, given the set of judgements by the observer or observers.

This page has been removed for copyright reasons

2.3 Psychophysical Theories.

Some of the earliest speculations on psychophysics related to games of chance. In the 17th and 18th centuries, the foundations of statistical theory were laid in the work of Pascal, De Moivre and Bernoulli, who were interested in the probabilities of success at the card table. In 1728, Gabriel Cramer (6) speculated further on why people will chance their monies even though the odds are against them. This led him to suggest that the subjective value of money fell as the owner possessed more, and further that it fell according to a square root law. Thus, suppose V is the value put on the money by the owner and Q is the quantity he owns, Cramer suggested that the added value ΔV for a small increment ΔQ was proportional to ΔQ but inversely proportional to the subjective value V. Hence

$$\Delta V = \frac{k\Delta Q}{V}$$

and thus by integration

$$V^2 = 2kQ$$

since it is assumed that zero money has zero value. Daniel Bernoulli (2) in 1738 published a treatise "Exposition of a New Theory of the Measurement of Risk" in which he quoted Cramer's proposal, but himself proposed that the subjective value was inversely proportional to the actual quantity of money. The equation thus becomes

$$\Delta V = \frac{k\Delta Q}{Q}$$

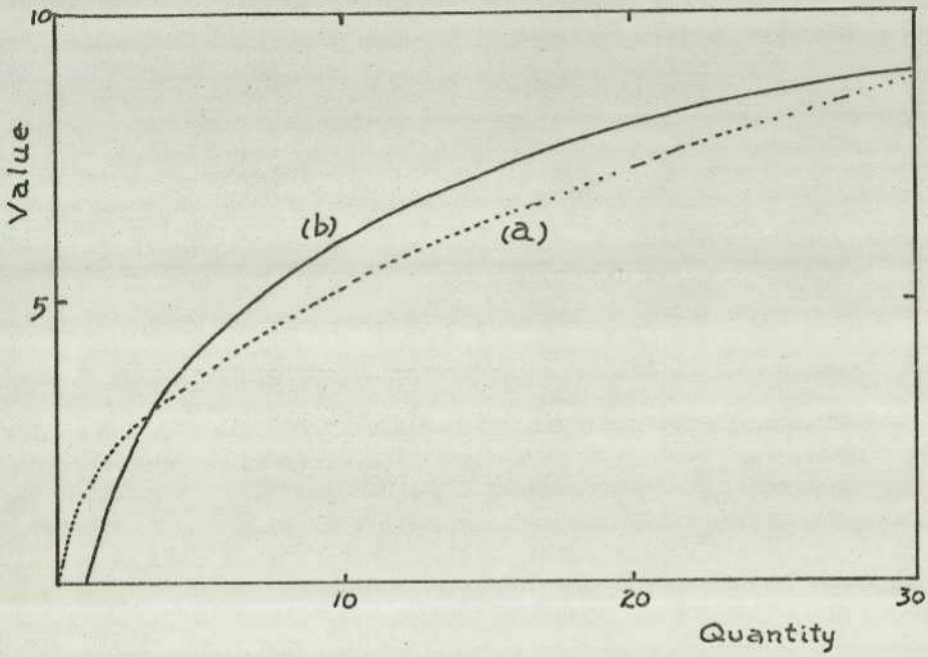


Fig 3.

Value as a function of Quantity.

(a) Cramer function $V = k_c Q^{0.5}$

(b) Bernoulli function $V = k_b \log Q$

(Note that with Bernoulli, value is zero when Quantity is finite.)

and thus

$$V = k \log Q$$

Clearly there is no zero in this equation.

There were thus two theoretical proposals, both leading to curves which fell off with increasing quantity, but having very different shapes. (Fig. 3). No experiments were reported to have been made to distinguish these two curves and the logarithmic equation of Bernoulli became accepted by economic theorists.

It seems curious that Bernoulli should have considered his equation to be an improvement. For according to Bernoulli, the observer must know his quantity of money Q in real terms in order to obtain his subjective estimate of increase in value. If he knows the quantity, it is difficult to see why he should not also know the quantity increment as well. Alternatively, if he only has a subjective impression of increment of value, how has he obtained more than a subjective impression of the total value?

Cramer's theory on the other hand seems wholly logical, since the subject's sensation of increase of wealth depends only on the wealth that he senses that he has, and not on the absolute quantity which he cannot comprehend.

There is a further objection to the Bernoulli hypothesis. Since the curve does not go through zero, the observer will think that he has no value when he still posesses a finite quantity.

In the early 19th century, Hebart introduced the concept of an absolute threshold or lower limit of sensation. Following this, Weber proposed his concept of the just noticeable difference and his law that the just noticeable increment in a stimulus is proportional to the stimulus. Weber's law is thus the law enunciated by Bernoulli but applied to physical stimuli. Weber's law states that

$$\frac{\Delta S}{S} = \text{const}$$

It was left for Fechner to carry out the integration and to state

$$R = k \log S + \text{const}$$

and hence when $R = 0$, $S = S_0$ and thus

$$R = k \log \frac{S}{S_0}$$

where R is the response, S is the stimulus and S_0 is the absolute threshold of the stimulus.

2.4 Methods of Scaling.

2.4.1 Methods employing successive just-noticeable differences.

Fechner's modification of Weber's law was more than a simple integration. For, whereas Weber said simply that the size of a just-noticeable difference (jnd) in a stimulus S is directly proportional to the size of the stimulus, Fechner said that each increase in stimulus by one jnd caused an increase in response of constant size. Fechner carried out a major experimental programme in which the Fechner law was used to measure responses. Thus, after setting a reference stimulus at some value S_1, a second stimulus S_2 could be increased from the value of S_1 until the observer could detect a difference. S_2 could then be decreased from a large difference until the observer detected no difference. The average of a number of such runs in each direction would establish the just-noticeable difference. If S_1 was now increased to $(S_1 + \text{jnd})$, a new series of tests could be carried out. Using this technique, the separation of two widely spaced stimuli could be bridged and hence, using the Fechner law, their ratio determined. The technique has been developed under the general title of Method of Minimal Changes. Details are described by Guilford (9) Chap. 5.

With the development of the use of statistical methods at the turn of the century, improved techniques were developed. In the so called Constant methods for measuring the jnd, a number of fixed stimuli close to the reference stimulus are successively presented to the observers (or in the case of a single observer they must be presented many times). The observer has to state whether he perceives a difference or not. Making use of the repetitions, an estimate of the probability that an observer will perceive a difference is obtained for each stimulus. The probabilities can be plotted against the value of the stimulus. These should form

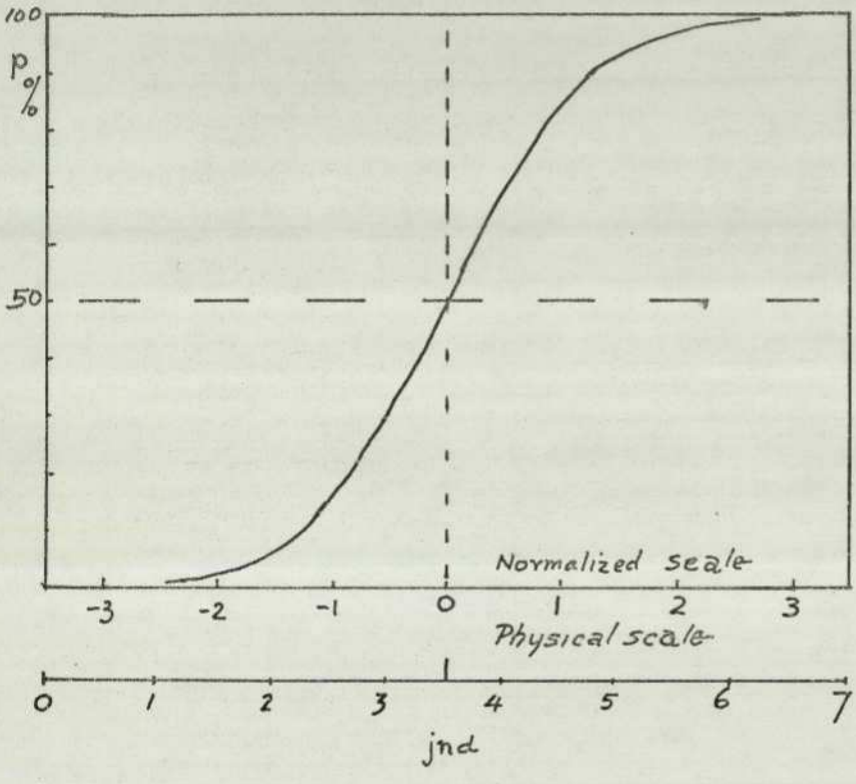


Fig 4.

Proportion of positive responses as a function of the magnitude of the physical stimulus.

some sort of ogive curve. In the crudest type of analysis, the point where the best curve through the observations cuts the line corresponding to the probability = 0.5 is taken to be the value of the jnd. There are many refinements of this approach in which various assumptions are made of the shape of the curve. For example, if the ogive is taken to be the error function, values of the corresponding standard normal deviates can immediately be derived for each stimulus (fig. 4). This enables a linear plot of stimulus as a function of normal deviate to be obtained. Hence interpolation for the 0.5 probability point can be much more accurate; see Güilford (9) Chapter 6. Care is taken in the design of the experiment not to give the observer clues as to which stimulus he is observing at any time, as such information could introduce bias.

2.4.2 Judgement Methods.

For problems in measurement well away from the threshold, a number of allied techniques have been used which rely on an observer's ability to judge when one stimulus is a fraction or a multiple of another. Merkel (11) in 1888 asked his observer to adjust a stimulus until it was double a reference stimulus. Plateau (13) in 1872 asked eight painters to paint grey so that the interval from black to grey was equal to that between grey and white. The process was developed in which observers were asked further to subdivide and ultimately in 1953

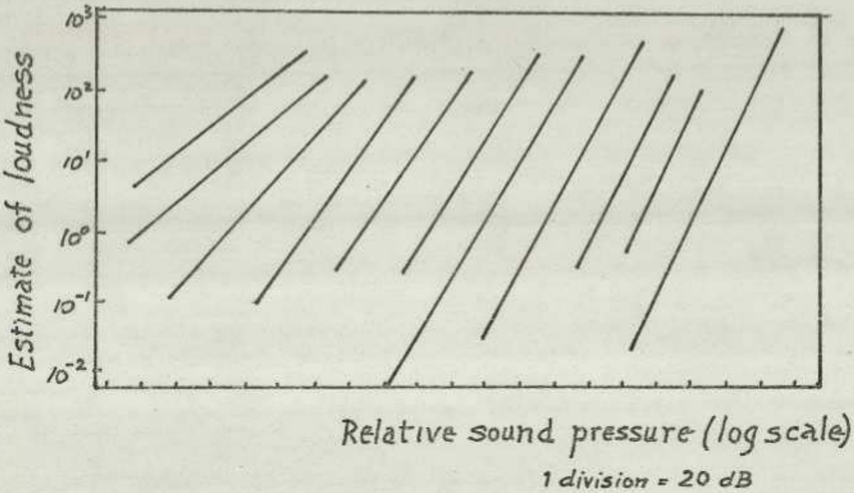


Fig 5.

Magnitude Estimation.

Estimates of loudness for individual observers as a function of the logarithm of sound pressure. Each line represents the observations of a single observer. (Stevens and Guirao. 1964)

S.S.Stevens (19) introduced the method of magnitude estimation. In this method the observer is presented with a series of stimuli in irregular order. He is then asked to give any number to the first one, and then to assign numbers to subsequent stimuli in accordance with his subjective impression. Thus if he thinks that the second stimulus is only one fifth as great as the first, he assigns to it a number equal to one fifth of the first value and so on. Whole numbers, fractions or decimals may be used at will. It is usual for the first stimulus to be near to the middle of the range of stimuli to be presented. Some twenty to thirty stimuli are presented to each observer, some of which are replications. It is found that the relationship between estimate and stimulus is not logarithmic as would be expected from the Fechner-Weber law, since a straight line is not obtained when response is plotted against the logarithm of the stimulus. On the contrary, a linear relation is found between the logarithm of the response and the logarithm of the stimulus. For example, a linear relation between the logarithm of subjective loudness and logarithm of sound pressure was found by one observer over four decades. (J.C.Stevens and Guirao (16)). This work is illustrated in figure 5, where the results for many individual observers are plotted on the same graph. Individual observers found similar slopes, indicating conclusively the power law

$$R = kS^n$$

In this case the exponent was close to 0.8. Such results as these, which were repeated for a large range of stimuli such as hand grip, electric shock and brightness led in every case to a power-law relationship.

Thus a serious disagreement appeared between the results from the new method and the results which had been accumulated over the previous hundred years, based on successive just-noticeable differences.

2.4.3 Categorical Methods.

A method intermediate between the two general classes described above is the method involving categorical judgement. The observer is asked to place each stimulus in one of several categories. The number employed usually ranges from four to nine. The categories are described to the observer either verbally, such as cold, cool, normal, warm, hot, or as numbers such as -2, -1, 0, 1, 2 . Alternatively the observer may be asked to mark along a line the position which corresponds most closely to his judgement of the sensation. (figure 6). The category method has the advantage of being very easy to administer and is widely used in market research as well as in psychophysical and psychological testing. In the analysis, the arithmetic mean of the scores is taken, each category having

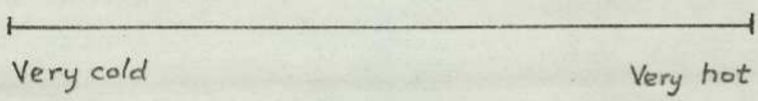


Fig 6.

The observer marks the line according to his sensation.

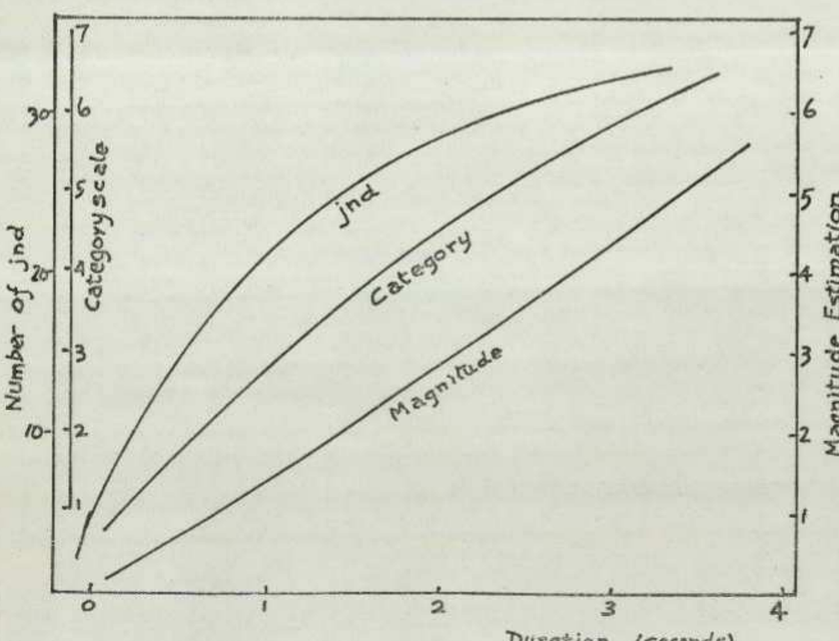


Fig 7.

Comparison of estimates of duration of a tone.

been allocated a number. The method yields results intermediate between those obtained from the other two methods. Figure 7 (Stevens (21)) shows the type of results obtained.

The method has difficulties in that it is impossible to arrange that the observer makes all his categories of the same width. Also observers have a tendency to avoid the extreme categories, possibly because they are not sure whether more extreme stimuli will subsequently be presented. However, the method is very widely used and yields results which are just as meaningful as the others. Various techniques exist for analysing the results, providing certain assumptions are made. These will be discussed in section 3.2

3. THEORETICAL MODELS AND ASSUMPTIONS

Any of the techniques described in the last section will yield a relation between the stimulus and the response. However, as seen, the relations derived by two methods will not necessarily be the same, but will usually be related to one another by an equation of the type

$$x' = f(x)$$

where f is a monotonic function. This means that only an ordinal scale has been found, if it is assumed that both equations are true. A scale of this type is of little value. We need to establish at least an interval scale and preferably a ratio one. Furthermore, we cannot be satisfied unless all methods yield the same results or alternatively we find good reason for accepting one method and rejecting another.

In order to do this, we should proceed along the lines of any physical theory i.e. we should postulate a theoretical model which fits known facts and we should use it to predict further experimentally verifiable facts. So long as success in prediction is achieved, we retain the model; but if the predictions of the model do not accord with all facts, it must be rejected or modified. The model basis for one method may rely on one or two simple assumptions, another may rely on many. However, all must be relevant and essential, and every effort must be made to devise critical tests to compare the performance of one model with that of another.

Any method of analysis implies some sort of model, although this is often not explicitly stated. On the model will depend the type of statistics that should be used in the analysis. An attempt will now be made to investigate the limitations of the different methods.

3.1 Jnd and Category Methods.

The models on which these methods are based have much in common as also do the methods of analysis. They will therefore be discussed together. Two possibilities are open to the investigator. He may seek a direct relation between the continuum of stimuli and the set of observations or he may attempt to relate the stimulus continuum to the psychological continuum whose scale may be related only indirectly to the observations. Most of the earlier work employed the direct approach.

3.1.1. Methods employing just-noticeable differences.

The direct approach has been extensively used not only for scaling stimuli producing sensations in the normal working range of observers, but also for those producing sensations near to the absolute threshold of detection. The methods appropriate to investigations near the absolute threshold will be considered in Section 5; we shall here be concerned with sensations in the normal working range of an observer.

The model assumes Weber's law

$$\frac{\Delta S}{S} = k$$

so that the stimulus continuum may be divided into a series of just noticeable differences. As has been seen, Weber's law is not based on any particular biological model, and many experiments have been made showing that k is far from constant when the full range of observer sensitivity is considered. Even in the working range, it may still vary by 2:1 for 20 dB variation in stimulus.

If Fechner's law is assumed, then there exists a scale between the stimulus and the response.

In the earliest methods such as the method of average error and the method of minimal changes, no mathematical model was assumed. It was believed that for reasons unspecified, the least difference between two stimuli which the observer can notice varies from time to time. It varies depending on whether the stimulus is rising or falling. If therefore these effects are carefully balanced in the experiment, the final average should represent the difference limen.

In the Constant stimulus methods, a series of steady stimuli are presented relative to a standard.

The proportion of 'yes' and 'no' scores is calculated for each stimulus. The difference limen is defined as that stimulus at which the proportion of 'yes' scores is 50%.

Different models are used to represent the relation of stimulus to response. These are required for finding the 50% point. For example, it is most frequently assumed that the curve is a normal probability integral. This therefore assumes that observations are distributed normally on the stimulus scale. If however the probability is related to the logarithm of the stimulus, a log-normal distribution is assumed. This model is consistent with Fechner's law, whereas the normal distribution is only an approximation true for small values of dispersion. The quantal method assumes an abrupt threshold, and thus applies only to a single observer under carefully controlled conditions. For each method a different model is appropriate, and it is important that the conditions of the experiment should be carefully planned to ensure that the model is appropriate.

Thurstone(26) introduced a judgement model which was applicable to his Law of Comparative Judgement, and this enabled him to relate the responses to scale values on the psychological continuum. The model assumes that every time a stimulus S_1 is presented to the observer, it gives rise to a discriminal process R on the psychological

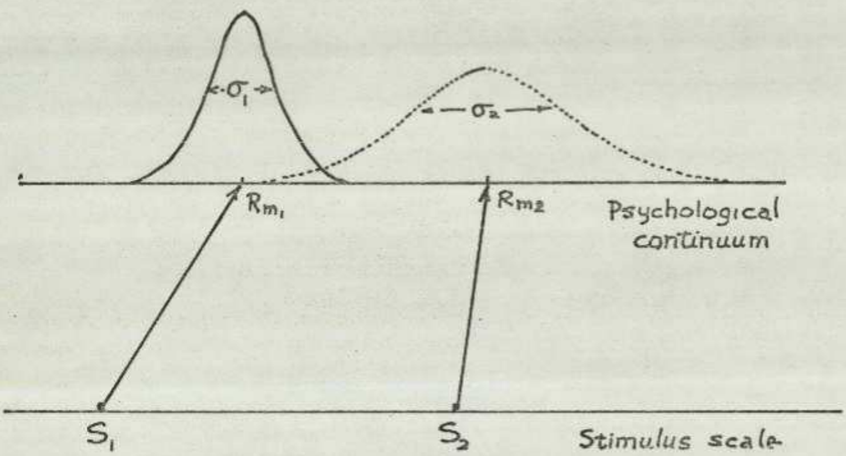


Fig 8.

Stimuli S_1 and S_2 give rise to discriminal processes R_1 and R_2 on the psychological continuum.

continuum (figure 8). However, because of fluctuations in the organism, the responses R are not always in the same position on the psychological continuum, but rather form a frequency distribution. These responses will cluster around a modal value called the modal discriminal process, R_m. R_m is taken as the scale value of the stimulus S_1. Associated with the distribution is a standard deviation and this is known as the discriminal dispersion for that stimulus. For another stimulus, it may be greater or less. The model assumes that the discriminal processes form a normal distribution. This is a major assumption.

This model can be used in interpreting the constant stimulus type of experiment. If we wished for example to find the scale relation for luminance of surfaces and to measure the jnd, the experiment might proceed as follows. We should present our observer with an unknown surface U_1 and the reference surface K. The observer would be asked to say whether U_1 was brighter than K or whether K was brighter than U_1. Other surfaces U_2, U_3 etc would be presented with K. One of the U s could then be used as the standard and the experiment repeated until all surfaces had been seen many times with every other surface. Now suppose that stimulus U_1 gave rise to a modal value R_{u1} on the psychological continuum (figure 9) and the distribution had a dispersion σ_u; also that K gave rise to a modal value

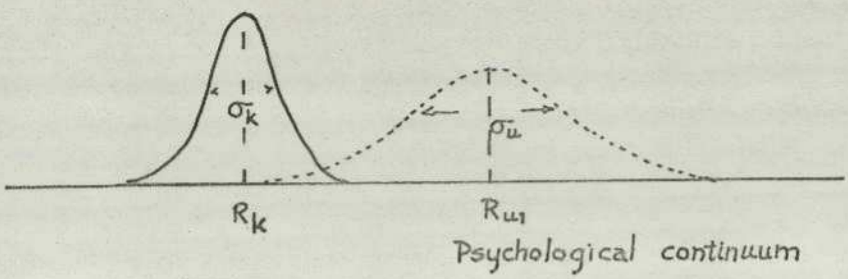


Fig 9.

Stimulus U_1 gives rise to response R_{u1} having dispersion σ_{u1}. The reference stimulus K gives rise to response R_k having dispersion σ_k.

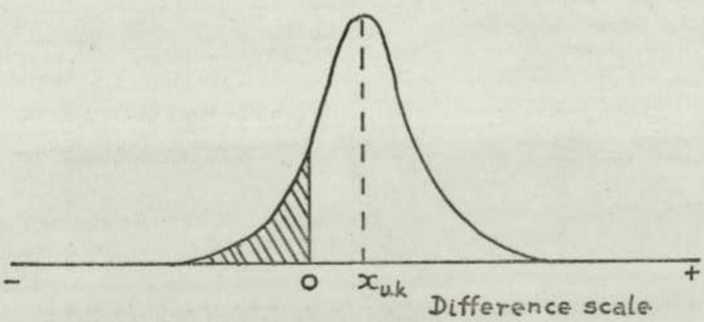


Fig 10.

The unshaded area gives the probability of a positive response when the modal difference $R_u - R_k$ is equal to x_{uk}.

R_k with dispersion σ_k, then the difference in modal values is $(R_{u1} - R_k)$ and the dispersion of this difference is

$$\sigma_{(u-k)} = (\sigma_u^2 + \sigma_k^2 + 2 r_{uk} \sigma_u \sigma_k)^{1/2}$$

where r_{uk} is the correlation coefficient between the momentary values of the discriminal processes. This is clearly an important factor since U and K are presented simultaneously. Thus the observer would always say that U was brighter than K when the discriminal processes were spaced so that the curves did not overlap. However in the region where the curves overlapped the observer might say that K was greater than U (i.e. the difference was negative) or that U was greater than K (i.e. the difference was positive). We can draw the frequency curve for the difference $(R_u - R_k)$, figure 10. Thus when the momentary difference was positive, the value would lie in the unshaded part; when it was negative, it would lie in the shaded part. The zero point is the boundary between the two regions, The modal value of the distribution x_{uk} is the scale value of the difference in brightness. The unshaded area is proportional to the probability of a 'correct' judgement and the shaded area to the probability of a 'wrong' decision given the stimulus U. It should be noticed that the scale is on the psychological continuum and not the physical scale of the stimulus. These scales need not be linearly related. The unit is related to the scale through the discriminal dispersion of the difference, $\sigma_{(u-k)}$. Thus if x_{uk} is the normalized deviate representing the separation of U and K, then

$$R_u - R_k = x_{uk}\sigma_{(u-k)}$$

or

$$R_u - R_k = x_{uk}(\sigma_u^2 + \sigma_k^2 - 2r_{uk}\sigma_u\sigma_k)^{\frac{1}{2}}$$

This is a statement of Thurstone's law of comparative judgement. (Guilford(9), Torgerson(27)). This equation is not soluble, since there are too many unknowns. By presenting all available stimuli in pairs, however, we can obtain sufficient equations to obtain a solution if some simplifying assumptions are made. The essential is that the correlation term, r_{uk}, is the same for all stimuli. It is then assumed that the dispersions are nearly equal. The equation then reduces to

$$R_u - R_k = x_{uk}(2(1 - r))^{\frac{1}{2}}(\sigma_u + \sigma_k)$$

where $\sigma_u \doteqdot \sigma_k$ and x_{uk} is the normalized deviate. Experimentally, the values of dispersion will depend on whether the experiment is carried out by one observer with many replications, by many observers each making one set of observations or by many observers each making several sets of observations. In practice each stimulus, i, is compared with all other other stimuli, j, and an estimate of the probability that i is greater than j is obtained. Then since it is assumed that all discriminal processes are normally distributed, the value of the normalized deviate x_{ij} can be read off from tables of the normal distribution function. Clearly the x_{ii} terms i.e. one stimulus compared with itself should all be zero and $x_{ij} = -x_{ji}$. The square matrix of all x values will thus be skew symmetric.

Since the discriminal processes are assumed to be normally distributed, the model implicitly assumes that the psychological continuum stretches from $-\infty$ to $+\infty$. In practice it could be limited so long as its ends are at least three times the discriminal dispersion beyond the extreme discriminal processes. If at least an interval scale is expected, this implies that the continuum of stimuli could notionally extend indefinitely beyond the region of investigation in both directions. The method is therefore unsuited to investigations close to the absolute threshold.

Attempts are frequently made to overcome this difficulty by relating the psychological continuum to the logarithm of the stimulus. But difficulties remain since there is then no logical reason for a threshold.

3.1.2. Category Methods.

The analysis of categorical data employs the same judgement model. For whereas in the method of paired comparisons described in the last section each stimulus was compared with another stimulus, in the method of categorical judgement each stimulus is compared with a category boundary. When an observer places a stimulus in a certain category, it is assumed that he has set up a category scale on his psychological continuum so that t_1 is the boundary between category 1 and category 2, (e.g. between 'visible' and 'bright'), t_2 is the boundary between categories 2 and 3 (between 'bright' and 'very bright') and so on. When a stimulus is presented to the observer, he compares it with the successive category boundaries; he finds the boundary next above the discriminal process corresponding to the stimulus and then places the stimulus in that category.

It is assumed that various random disturbances cause each category boundary to vary in position on the psychological continuum so that its position relative to its mean value is described by a normal distribution. The observer judges the stimulus to be below a given category boundary when the discriminal process is below the position of the boundary at the moment of comparison. It is thus seen that the category boundaries behave exactly like the second stimulus in the method of paired comparisons.

The same equations apply except that R_k is replaced by B_g, B_g being the position of the gth category boundary. This has a dispersion σ_g. Thus

$$B_g - R_u = x_{ug}(\sigma_u^2 + \sigma_g^2 - 2r_{ug}\sigma_u\sigma_g)^{\frac{1}{2}}$$

This is the law of categorical judgement, and can only be solved if the same simplifying assumptions are made as before. The two most important are that the dispersion σ_g is the same for all category boundaries and that the correlation term vanishes. Then

$$B_g - R_u = x_{ug}\sigma_u$$

where σ_u is the dispersion of the discriminal process. If in addition the dispersion of the discriminal process is assumed to be constant, the equation reduces to

$$B_g - R_u = x_{ug}c$$

where $c = (\sigma_u^2 + \sigma_g^2)^{\frac{1}{2}}$

Clearly these assumptions simplify the analysis and they may be fully justified if the quantity of data is limited and the variability is high.

Guilford states that there must be perfect correlation between the judgements given and the momentary response on the psychological continuum. This seems unnecessarily strict. It is only necessary for the judgement to be unbiassed. For example, if there is a third variance σ_j^2 representing an error in the actual judgement of the category boundaries, this will simply add to the value of σ_{ug} and will be undetectable so long as it is not correlated with either σ_u or σ_g.

It is seen that analysis will show the position of the category boundaries on the psychological continuum and the positions of the stimuli. The scale used takes the standard deviation as unity, and hence a relation can be found between the category scale and the jnd scale, if an appropriate probability, such as 50%, is taken to represent the jnd boundary. This jnd is the value that would be found in the absence of a standard. It is likely to be much larger than that found by the constant method of the previous section, because the fluctuations in the category boundaries are likely to be greater than the fluctuations in the standard.

3.2 Methods of Analysis of Data.

The methods advocated range from essentially graphical processes to fairly elaborate least squares procedures. For finding a single limen by the constant stimulus method, Torgerson(27) lists the following procedures.

(1) Connect the points with straight lines; linear interpolation method.

(2) Make plot on normal probability paper (or convert to normal deviates and plot on ordinary paper) then join up the points with straight lines;

(3) As 2, but fit best straight line; normal graphic method.

(4) Fit integral of the normal curve to the points using least squares procedure.

(5) As 4, but weight the proportions appropriately. (Urban and Kelly's method)

(6) Substitute the logarithm of the stimulus in the first five procedures.

(7) Take pencil in hand and fit a smooth ogive to the points.

The inclusion of (7) suggests that Torgerson recognized that the result is no better than the data from which it is derived, and that no sophistication in analysis can improve the accuracy if that of the original data is limited. It should be noted that these methods are not interchangeable, but rely on different models. For example, the substitution of the log of the stimulus for its normal value can only be justified if the model justifies it. The first essential is therefore to decide on the model which will be assumed to represent the decision processes. Once the model is decided, tests must be included in the analytic procedure to verify whether or not it truly represents the process within the limits of experimentsl error. If this cannot be done, it may be possible to rework the analysis using an alternative model to see which fits the data better. The model is then chosen on the basis of the best fit.

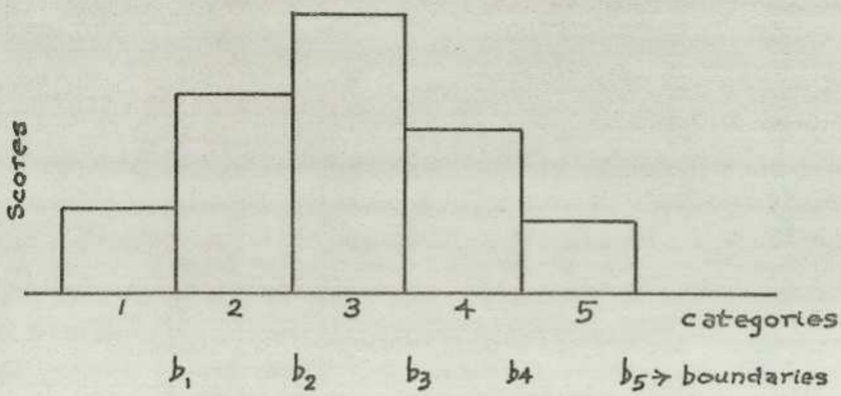


Fig 11.

Scores in the categories.

The score attributed to category 1 is the sum of the responses occurring below boundary b_1, and so on.

Stimuli \ categories	1	2	3	4	5
1	p_{11}	p_{12}	p_{13}	p_{14}	$p_{15}=1$
2	p_{21}	p_{22}	p_{23}	p_{24}	1
3	p_{31}	p_{32}	p_{33}	p_{34}	1
4	p_{41}	p_{42}	p_{43}	p_{44}	1

Fig 12.

The p - matrix.

It is unfortunate that methods of testing models do not seem to be so well advanced as the methods of scaling.

The methods of analysis of data are substantially similar for comparative and for categorical data. The latter will therefore be discussed. The appropriate equation is

$$B_g - R_u = x_{ug} a_{ug}$$

where B_g is the position of the boundary of category g, R_u is the position of the response to stimulus u, both measured on the psychological continuum, x_{ug} is the corresponding normal deviate and a_{ug} is the standard deviation of response u relative to the boundary of category g, i.e.

$$a_{ug} = (\sigma_u^2 + \sigma_g^2 - 2r_{ug}\sigma_u\sigma_g)^{\frac{1}{2}}$$

The method of analysis is as follows. The data consist of a number of scores either from one or from many observers who rate the stimuli into their different categories. The scores for one stimulus may be as shown in figure 11. The number of scores in category 1 divided by the total number of scores gives the probability of a score occurring below the upper boundary of category 1. Similarly the sum of scores in categories 1 and 2 divided by the total gives the probability of a score being less than the upper boundary of category 2 and so on. It is therefore possible to build up a probability matrix

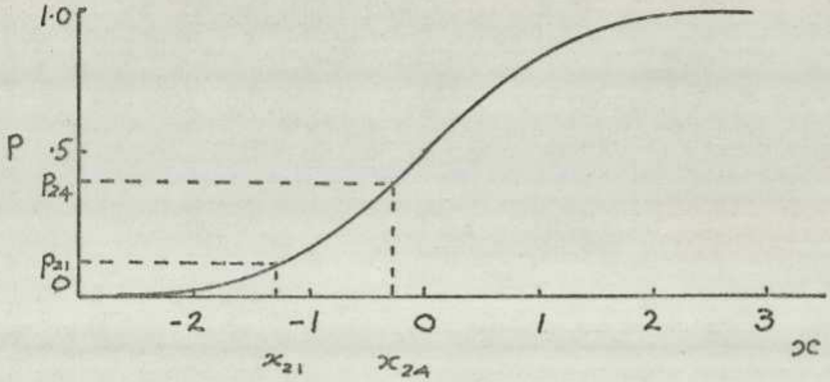


Fig 13.

The normalized scale value x is derived from the probability p using the normal distribution function.

categories / stimuli	1	2	3	4
1	x_{11}	x_{12}	x_{13}	x_{14}
2	x_{21}	x_{22}	x_{23}	x_{24}
3	x_{31}	x_{32}	x_{33}	x_{34}
4	x_{41}	x_{42}	x_{45}	x_{44}

Fig 14.

The x - matrix.

(p-matrix) in which the rows represent the different stimuli and the columns the upper boundaries of each category(figure 12). The last category has no upper bound, thus the probability is always unity.

Assuming a normal distribution and hence a normal probability integral, each of the probabilities can be converted into a corresponding normal deviate x using the normal distribution function; (figure 13) Thus we could evaluate the scale value corresponding to this probability p_{ij} if we knew the discriminal dispersion a_{ij} of the stimulus i relative to category boundary j. The scale value would be

$$X_{ij} = x_{ij}a_{ij}$$

Thus initially the unit of measurement is the discriminal dispersion. The x-matrix is as shown in figure 14. It should be noted that the result of this process is to reduce the data by one column. The purpose of analysis is to find the position of category boundaries and stimuli on the psychological continuum and to find the values of the dispersions a_{ij}. There are not enough equations to find all these constants but making the restrictions referred to in the last section i.e. that the correlation coefficient is zero and that the dispersion of the category boundaries is const nt, σ_u can be left free to vary from one stimulus to another. A further restriction that σ_u is constant for all stimuli simplifies the analysis still further, but it is doubtful whether it is a satisfactory a priori assumption. If the less stringent of these restrictions is applied, then the equation becomes

$$B_g - R_u = x_{ug} a_u$$

a, the standard deviation, is a function of u but not of g.

The analysis can then be carried out somewhat empirically or by a least squares process. In the former apprach, the category widths are first found by subtracting the (j - 1)th column from the jth column to give the width of the jth category. Then by cumulating these from the left, a scale of categories with an arbitrary zero is obtained. The distance of the x values for each stimulus then yields values for the average scale position of each stimulus. The standard deviations can be deduced from the interquartile ranges. The problem is thus solved. Now by using the values so found, a theoretical probability matrix can be deduced. These probabilities can be compared with the original values using the χ^2 test of goodness of fit. This however seems to be the most unsatisfactory part of the procedure since better agreement is frequently obtained than can be justified. It has been suggested that this is due to the cumulating procedure. This procedure is used in all methods of analysis, therefore all are subject to this criticism. Also since the whole set of data have been used to derive the theoretically expected values, it is not clear how many degrees of freedom exist.

The more formal methods use a least squares analysis applied to the x matrix. The method of Gulliksen minimizes

$$\sum_{u=1}^{n}\sum_{g=1}^{m} (x_{ug}a_u - B_g - R_u)^2$$

subject to an arbitrary zero and a scale unit equal to the variance of the category boundaries. The method of Tucker minimizes

$$\sum_{u=1}^{n}\sum_{g=1}^{m} w_{ug}(x_{ug} - \frac{B_g}{a_u} - \frac{R_u}{a_u})^2 \cdot$$

He, it will be seen, included a weight w_{ug} which, for a normal distribution has the value y^2/pq, where y is the ordinate of the normal frequency curve, p is the a posteriori probability and $q = (1 - p)$. These weights are called the Müller-Urban weights. The first factor, y^2, allows for the fact that in converting probability values to normal deviates, the slope of the probability integral changes. Thus referring to figure 13, an error in x_{24} derived from p_{24} will be less than the error in x_{21} derived from p_{21}. The errors will be inversely proportional to the slope which in turn is proportional to the ordinate of the frequency curve. Thus in minimizing the square of the deviation, x^2 values must be weighted by y^2. The second factor, $1/pq$, arises from the reliability of an estimate based on proportions. The mean square error for a binomial distribution $\sigma^2 = pq$. Thus the weighting is proportional to $1/pq$. It seems a criticism of Gulliksen's method that he does not apply these weights. There seems little point in the relatively elaborate analysis if they are omitted. If the weights are not used, great emphasis is placed on the extreme results. To avoid this error, some workers have suggested throwing away the extreme data. However, it seems a mistake to do this, as it is often at the extremes that one distribution differs most from another one. The various methods of analysis are described in detail by Torgerson(27).

Torgerson compared the estimation of scale values for the categories and stimuli obtained on three different sssumptions being applied to the same data. The assumptions were

(i) σ_u is variable, σ_g is fixed.

(ii) σ_g is variable, σ_u is fixed.

(iii) Both σ_u and σ_g are fixed.

The data supplied was artificial and fitted assumption (i) perfectly. The results were then analysed using the second and the third assumptions. It was found that the results were then completely distorted. Nevertheless, Torgerson seemed satisfied that the three methods agreed. It would therefore seem that more work needs to be done on these methods in order to set up additional criteria for the validation of the model used.

It should also be noted that many data are required if a high level of significance is to be obtained. For it is usually stated that the χ^2 test cannot be applied unless there are a minimum of five data in each cell. If therefore the stimuli have an appreciable variation, it will be found that many cells, corresponding to the tail of the distribution will have less than five data in them. In this case, it is impossible to build up a complete matrix. If the matrix is not complete, the methods lose much of their elegance and the data must be handled more individually rather than directly by computer.

It seems likely that unless data are available from more than 200 observers, (or observer occasions) the analysis for a five category scale must be carried out graphically and the significance assessed individually based on the contents of the cells. For more than five categories, there must of course be more observers. This makes the methods increasingly academic.

The most useful simplifying assumption is that the categories which have an upper and a lower bound are of equal width. In other words the category boundaries are equally spaced on the psychological continuum. This is very often found to be the case within the limits of precision allowed by the data. The χ^2 test can then be used to test the hypothesis that the observations are normally distributed on this scale. If this is the case, there is justification for using the simplified model.

3.3 Magnitude Methods.

These methods have been referred to in Section 2.4.2. The main assumption is that an observer can estimate the ratios of stimuli. This is what he is directed to do. If it is accepted that he can do this, (and it seems well verified by many sets of data) then no complicated analysis is required, since the scaling is direct on to the stimulus scale and does not require the psychological continuum in between.

For each observer the logarithm of the magnitude of estimate is related to the logarithm of a physical stimulus and the best straight line is either drawn by eye, or is calculated by least squares procedure. It is necessary that the stimuli should be distributed evenly over a logarithmic scale covering the range of interest, in order to obtain the best relation. The slope is found for each observer. Figure 5 indicates the type of results that are obtained. The average of the slopes is taken as giving the appropriate relationship. No attempt is made to find absolute relationships as the numbers are purely arbitrary.

The model assumes that errors are distributed normally on a logarithmic scale, i.e. they are distributed log-normally on a stimulus scale. This is consistent with the assumption that the observer judges ratios. No attempt is made to find a scale on the psychological continuum.

The hypothesis that errors are distributed log-normally rather than normally conflicts with the models assumed in the first five methods of analysis of constant stimulus data referred to on p.39. Clearly the two models cannot be simultaneously valid. The criteria which should be applied in choosing the appropriate model will be considered in the next section.

<u>4. CONDITIONS OF VALIDITY OF METHODS.</u>

In this section we shall consider why the methods described above give different answers and how it should be decided which of the methods is applicable in any particular set of circumstances. This problem seems to have worried S.S.Stevens who wrote in 1971(23) "When three different scales have been created of the same continuum by three separate sets of operations - - jkds, category estimation and magnitude matching - - on what grounds do we presume to ascribe to one of those scales a superior position?" Surely to decide this we must consider the type of continuum that is being scaled and see whether all methods are appropriate to all types of continuum. Stevens speaks of two types of continuum, the prothetic and the metathetic continua. The dictionary does not help in understanding his meaning, but he states (24) that prothetic continua are those concerned with intensity, degree or amount, whilst metathetic continua are qualitative. He says that a power function will describe the former, but usually not the latter. On the other hand he says that a partition scale such as a category scale or a jnd scale constitutes a straight linear segment of a magnitude scale on a metathetic continuum.

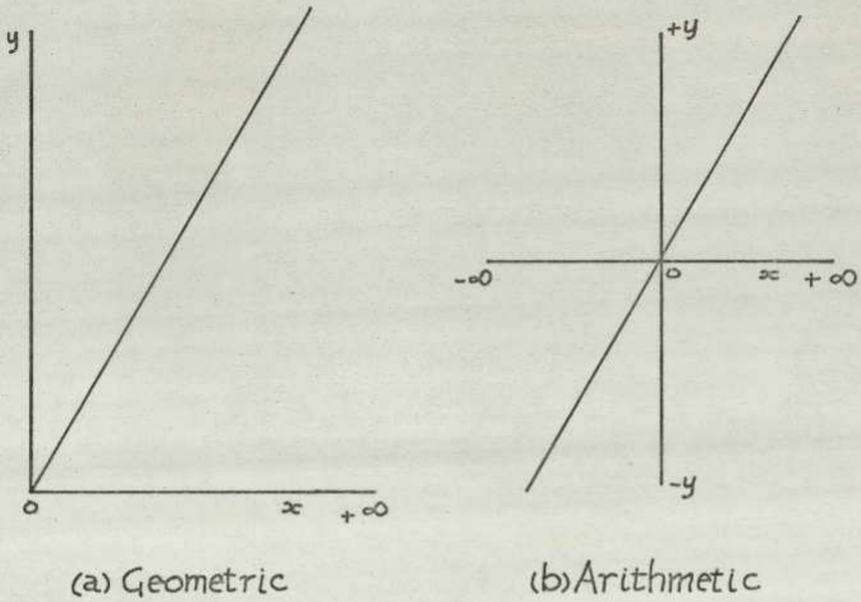


Fig 15.

Two types of ratio scale.

4.1. Two types of Ratio Scale.

Let us consider what types of scale these may be. Both may be ratio scales, but ratio scales can be of two types. This possibility of two types of ratio scale does not seem to have been noticed in the literature.

A ratio scale, it will be recalled must be invariant under the transformation

$$x' = ax$$

However, there are two classes of such a scale; in the first x lies in the range zero to infinity whilst in the second, x lies in the range minus infinity to plus infinity. These are illustrated in figure 15 (a) and (b). Let us call these two classes the geometric and the arithmetic classes.

4.1.1. Geometric Class.

Continua that fall into this class include physical quantities such as length, mass, force and absolute temperature. All of these are limited to positive quantities for, whereas it is easy enough to talk of negative length as a mathematical abstraction, no one can show an observer such a quantity for him to estimate. Consider the process of estimating a length of say 300 mm. The observer must consider what he remembers the length of a metre to look like and estimate how many times his unknown would go into it, or he may consider how many times a remembered centimetre would go into it. He will estimate it by ratio. If he could do this a large number of times (forgetting his previous estimate between each estimation), the estimates would be distributed about the geometric rather than the arithmetic mean, since all measurements were made on the basis of a ratio. So although the measurements are on a linear scale, the errors are distributed according to a log-normal distribution. Errors on such a scale are usually considered to be normally distributed, but the conditions for a normal distribution are clearly not satisfied, since the continuum does not extend between minus infinity and plus infinity. It is of course a perfectly valid assumption so long as the standard deviation is small compared with the magnitude of the response. This is so in conventional physical measurement, but it is not true in most psychophysical experimentation. The coefficient of variation in such work is usually of the same order of magnitude as the quantity which is being estimated, and the log-normal distribution must therefore be used.

Since the measurements are made on a ratio basis, the method of magnitude scaling is entirely appropriate. The jnd and category methods on the contrary rely on a normal distribution being obtainable on the psychological continuum or even on the stimulus continuum itself. Thus a continuum stretching to minus infinity is required. Any attempt to use these methods is thus likely to lead to curvature, since one is attempting to map a semi-infinite continuum on an infinite one. This is in fact what is found.

The characteristics required of the continuum are first that the whole of it can be mapped on a semi-infinite line and further that the zero is accessible to the appropriate senses of the observer. To illustrate this second requirement, consider observations of temperature on the Kelvin scale. The Kelvin scale is a ratio scale of the geometric type, but to an observer unaided by instruments, any temperature below about 260 K is entirely outside the range of his senses. He will therefore not be able to apply magnitude scaling in this case. He would in fact use his knowledge of the scale and his knowledge of ice and boiling water to estimate the temperature. The zero is far from either of these points, thus normal statistics would here be applicable.

4.1.2. Arithmetic Class.

This type of continuum, in which there is extension on both sides of zero, will occur in a typical physical measuring situation. For example in measuring the length of an object using a graduated scale. The unknown length is placed next to the scale and the distance to the nearest graduation is estimated. The graduation on either side of the end point may be used. Then because the distance to the nearest graduation is small compared with the length being estimated, the errors may be considered to be normally distributed with a constant standard deviation. The same might be considered to apply in using a visual photometer where the illuminance due to an unknown source is compared with that due to a standard. The following technique might be used. The photometer is set close to balance, then a number of observers are asked to say which side is brighter. From their observations the probability that A is greater than B can be estimated. The photometer is then moved and the process is repeated. The conditions would change from one in which A was brighter than B, through a 50% probability point to the condition where B is found to be brighter than A. It would be reasonable to assume that the frequencies were normally distributed on this continuum. This type of continuum will be obtained in any test in which the observers are asked to answer questions which lead to answers which are balanced about a central value. For example, " Is a note of higher or lower pitch than a standard?" or " Is a street lighting installation too bright, satisfactory or too dim?" Again we expect the answers to be normally distributed on a ratio scale. In order to test whether a continuum is of this type or not, it is necessary to ask whether there is a standard, and if so whether it is large compared with the deviations from it. Alternatively we may ask whether the zero of the absolute magnitude under investigation may be treated as infinitely distant. In the cases of the length or illuminance measurements, we have a physical standard; in the case of the streetlighting we have an implied standard, i.e. the level of brightness that is considered to be comfortable by the observer. It will be noted that it is unnecessary for the basic scale of the property to be a ratio scale or even an interval scale, provided that it is a smooth monotonic function of the stimulus scale. Since only a small portion of the scale is observed, it is likely to appear sensibly linear over this range, or at least sufficiently linear for the variations to be considered as normally distributed. It is also unnecessary for the observer to know the location of the zero.

4.2. Comparison with Stevens' Continua.

An attempt has been made to define two classes of continua both having ratio scales. This has been done purely in terms of the type of measurement which is applicable. As noted above, Stevens(23) distinguishes two continua in terms of the type of curves that they yield when different systems of measurement are applied to them, regardless of whether the system of measurement is appropriate or not. He says that the prothetic and metathetic continua can be distinguished by four criteria. These are

(a) Size of jnd.
(b) Form of category rating scales.
(c) Time-order error.
(d) Hysteresis.

We will now consider how each of these criteria might be affected by the choice of the geometric or the arithmetic continua.

4.2.1. Size of jnd.

Stevens states that the size of the jnd grows as we go up the scale for prothetic and remains constant for metathetic continua. It has been seen that for the geometric class described above, the estimation error increases proportional to the

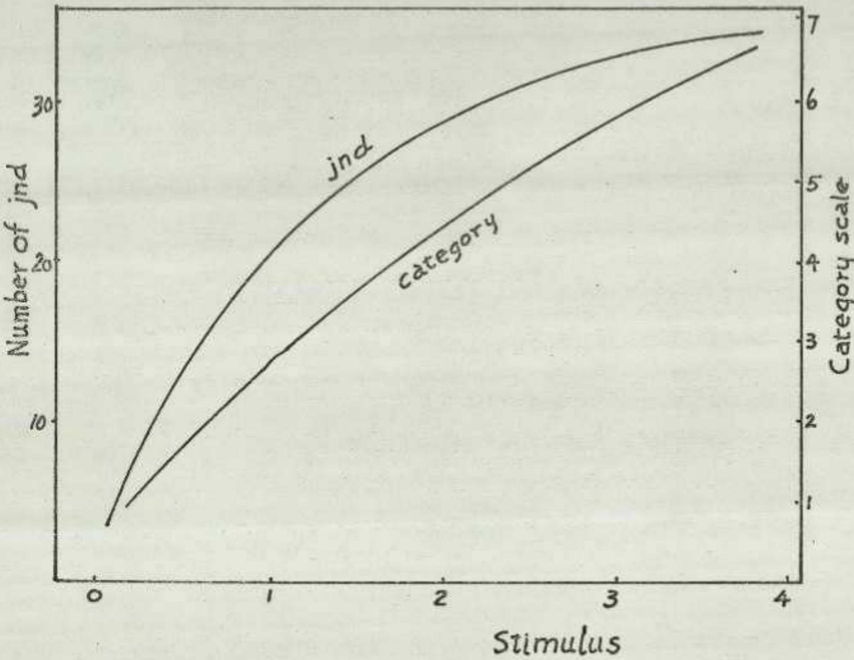


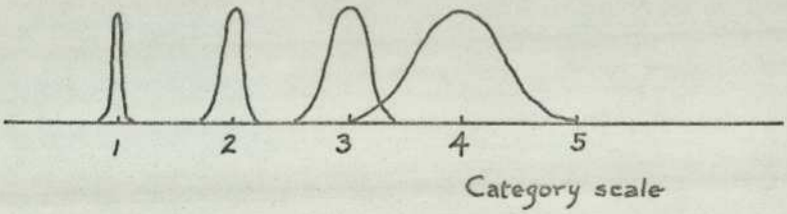
Fig 16.

Relation of jnd and category scales to ratio scale.
(S. S. Stevens 1957)

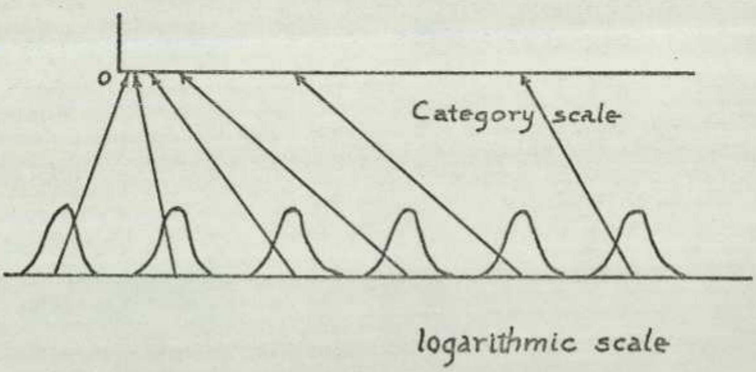
estimate of the magnitude of the property being measured. This is obvious for such continua as length and mass, and one would expect it to be general over the range of any property that is readily accessible to the senses. For example, one might expect it to hold for a length between 0.5 mm and say 50 m, a range of 10^5. For distances greater or less than this range, the eye needs some aids. The method of estimating is thus changed. This could give rise to different constants in the equation. We therefore expect the jnd to be proportional to the magnitude on the psychological scale. Errors are distributed log-normally. The magnitudes themselves are measured on a uniform ratio scale which starts from zero in a positive sense.

On the other hand, the arithmetic scale leads to constant variances of estimation and therfore to constant values of the jnd. This is the characteristic of Stevens' metathetic continuum.

We thus see that in respect of the size of the jnd, the two classes of scale agree with Stevens' two continua.



(a)



(b)

Fig 17.

Two ways in which an observer may categorize a geometric scale

4.2.2. Form of Category Rating Scales.

Stevens states that measurements of a prothetic continuum using a jnd or category scale yield a curve which is concave downwards (figure 16) with the category curve always below the jnd curve. A metathetic continuum on the other hand yields a linear relationship on the category scale. This latter is consistent with associating the metathetic continuum with the arithmetic class. Since with the arithmetic class, the frequencies are normally distributed one would expect this linear relationship.

Conversely since on a geometric class of ratio scale the dispersion is proportional to the magnitude of the stimulus, we would expect curvature in a categorical experiment, but the degree of curvature to be expected would depend on how the observer operated. There seem to be two possible modes of operation.

He may set up his category boundaries so that they are equally spaced on a magnitude scale,i.e. they are a uniform distance apart. These boundaries would each be subject to fluctuation, but the dispersion would be distributed normally on a logarithmic scale, i.e. the dispersions would be proportional to the magnitude. Since the dispersion controls the precision with which he can assign stimuli to categories, he will be much more precise for the low categories than for the high ones (figure 17 a).

There will thus be more jnd s per category for small stimuli than for large ones. This is as shown in figure 16. The exact relation will depend on the number of categories used. This will control the degree of averaging. One can thus state that this explanation would be consistent with Stevens' findings.

An alternative possibility would assume that the observer set up his categories so that their widths were proportional to the magnitude of the stimulus. i.e. the category widths are distributed log-normally as also are the dispersions (figure 17b). One might then expect the jnd and category scales to be similar since each category would contain the same number of jnds. This explanation is not consistent with Stevens' findings and furthermore seems to be a more complicated process than the first. It therefore seems likely that the first explanation is the true one.

4.2.3. Time-Order Error.

When two stimuli are presented one after an other, there is a tendency for the observer to overestimate the second relative to the first. Thus if the standard is repeated, the observer will usually say that the repeated stimulus is the larger one. This is referred to as a negative time-order error. Various theories have been put forward to explain this phenomenon, though none seems

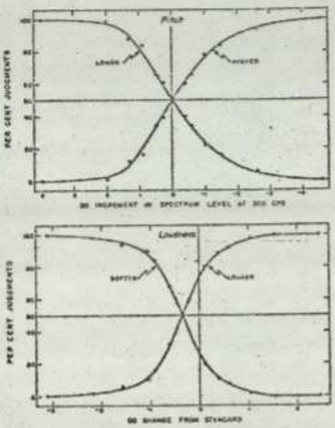


FIG. 2. Psychometric functions from 50 subjects obtained with recorded tests for ability to discriminate the pitch and loudness of bands of noise. Each test contained 110 test items consisting of a standard noise (2 seconds) followed immediately by a comparison noise (2 seconds). The band of noise used in the pitch test was shifted up or down the frequency scale. Note that there is a "time error" for loudness but none for pitch.

Fig. 18.

Steven's experimental results for pitch and loudness.

satisfactory. Stevens(23) suggests that it is a function of the continuum itself, and that the prothetic continua show a negative time-order error whilst the metathetic ones do not.

Stevens does not analyse the reasons for this in detail, but he shows figure 18 in which the limen for pitch and loudness have both been estimated by the constant method. It is seen that for pitch, the crossing point of the curves corresponds to the point where the stimulus and the reference have the same value. Pitch he says is a metathetic continuum. For loudness, the crossing point does not coincide with zero difference or, put in another way, when the stimulus and the reference have identical values, 80% of observations would indicate that the stimulus which is heard second was louder than an identical stimulus which was heard first. He states "What the Fechnerian time-order error really seems to depend on is the same basic process that makes the category scales on Class I (prothetic) continua turn out to be concave downward, namely the asymmetry of sensitivity - the fact that discrimination is better toward the low than toward the high end of the range."

This idea of the sensitivity being greater toward the low end is what may be expected in a geometric type of continuum. Applying our model to this problem, we assume that the distribution of sensation of loudness is log-normal on the ratio scale. Thus the observer will experience more sensations in the high direction than in the low direction relative to the mode of the distribution. If we assume that in making his decision the observer subjectively takes the mean of his distribution of sensations, we see at once that the median, i.e. the 50% point at which the loudness curves of figure 18b cross, is on the lower side. It is noticeable that the ogive type curve is approximately anti-symmetrical about the 50% point. One may thus infer that the distribution is normal on a logarithmic scale. Since the relation of sound intensity to loudness is given by

$$L = kI^{0.3}$$

(as derived by Stevens from magnitude estimation) then

$$\log L = 0.3 \log I + \text{const}$$

Hence a symmetrical distribution on the intensity scale indicated a symmetrical distribution on the log-loudness scale. If we assume this to be a normal distribution, then the distribution on the loudness scale will be log-normal.

From the percentage points on the log-intensity scale, it is found that the standard deviation σ is equal to 0.069 on the log scale to the base 10, or 0.16 on the natural logarithm scale. On a linear scale the ratio of the mean to the median

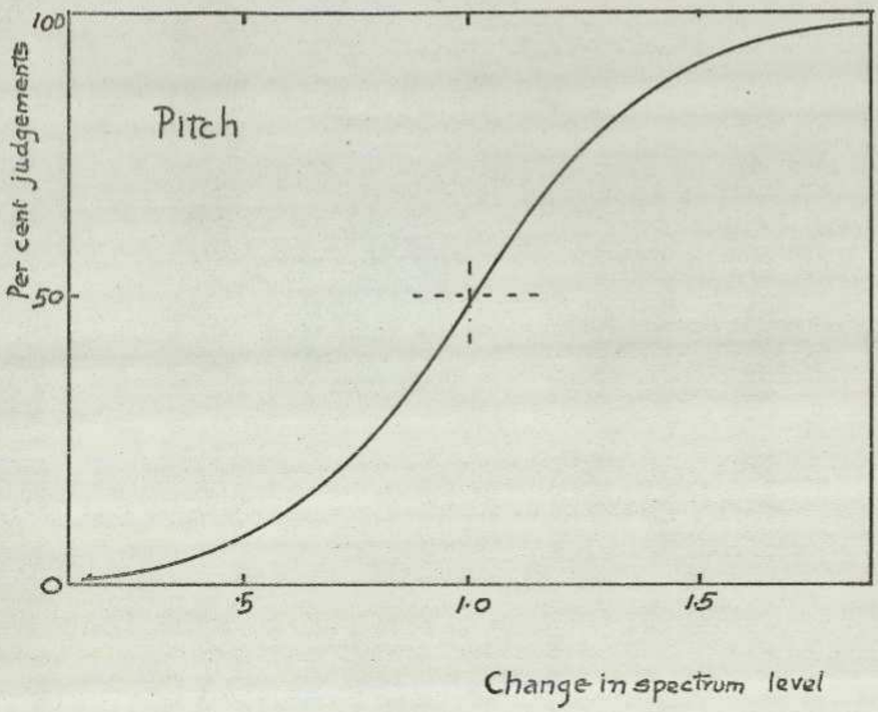


Fig 19.

Stevens' pitch curves of Fig 18 redrawn on a linear scale.

for a log-normal distribution is $e^{\frac{1}{2}\sigma^2}$ (1) where σ is the standard deviation on the log scale. Hence we see that this ratio for the distribution under consideration is 1.013, corresponding to 0.056 dB. Thus , on the model proposed, the displacement of the crossover (50% point from the reference value should be only 0.006 dB instead of a value of 0.3 as shown in Stevens' graph. Thus, although it gives an answer in the right direction, the above explanation cannot be considered to be a complete one for the time-order error.

As an aside, it should also be noted that the ogee curve for pitch is not antisymmetrical but has much greater curvature on the high frequency side than on the low frequency side. The reason for this is that the logarithmic scale is the wrong scale for this quantity. There is no magnitude sensation felt by an observer when he distinguishes two sounds of differing pitch. They differ qualitatively. The linear scale should therefore be used. When this is done the curves appear substantially normal. (figure 19). This result is somewhat surprising when it is remembered that all music is based on ratios, but it must be noted that Stevens' results concerned only small sections of a noise spectrum. This once more illustrates the importance of using the correct model.

Stevens states that another manifestation of the time-order error is the way in which measurements

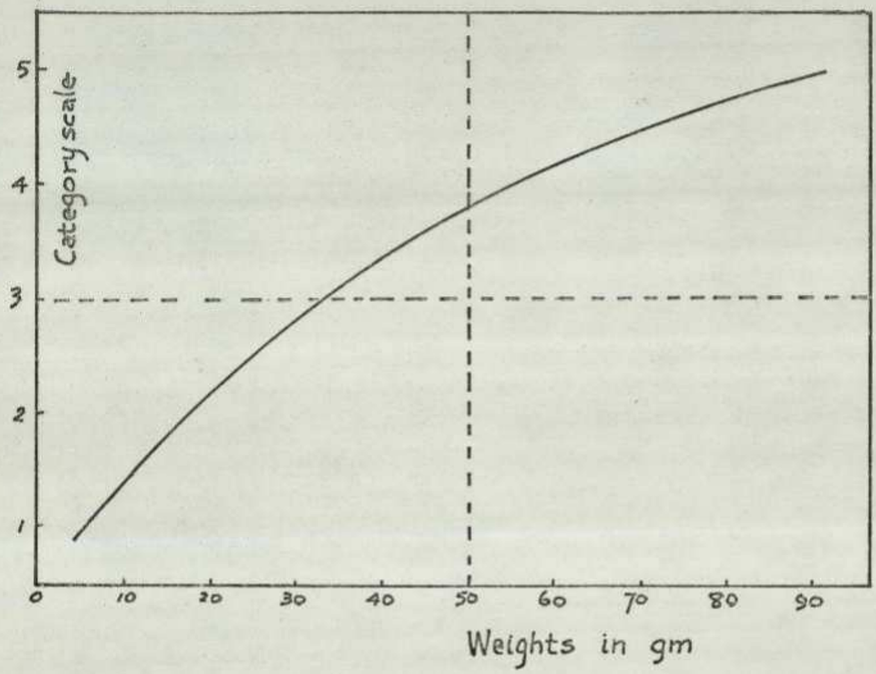


Fig 20.

In category methods, the mid-value of the stimulus does not always coincide with the middle category.

of prothetic stimuli on a category scale result in a curve which does not pass through the origin. This is illustrated in figure 20 where the sensation of weight is measured on a category scale over a relatively wide range of weights. It is seen that the reference weight, 50 gm, is not rated to have a scale value in the middle of the category scale. An effect of this type would be expected on the assumption that the sensations were log-normally distributed, for the same reason that the curve is concave downwards as discussed in section 4.2.2. The size to be expected for this effect, however, cannot be estimated unless more is known of the dispersion.

4.2.4. Hysteresis Effect.

In experiments on bisection, in which the observer is given two different stimuli and then is asked to adjust a third stimulus until it is midway between the first two, it is found that the variable stimulus has to be higher when the signal is increasing than it does when the signal is decreasing. There is thus an overshoot or hysteresis effect. Results of the type illustrated in figure 21 are obtained. Two surfaces are first presented: the brightnesses of these are given the values 0 and 100. The brightness is then increased from the low value and the observer is asked to say when the

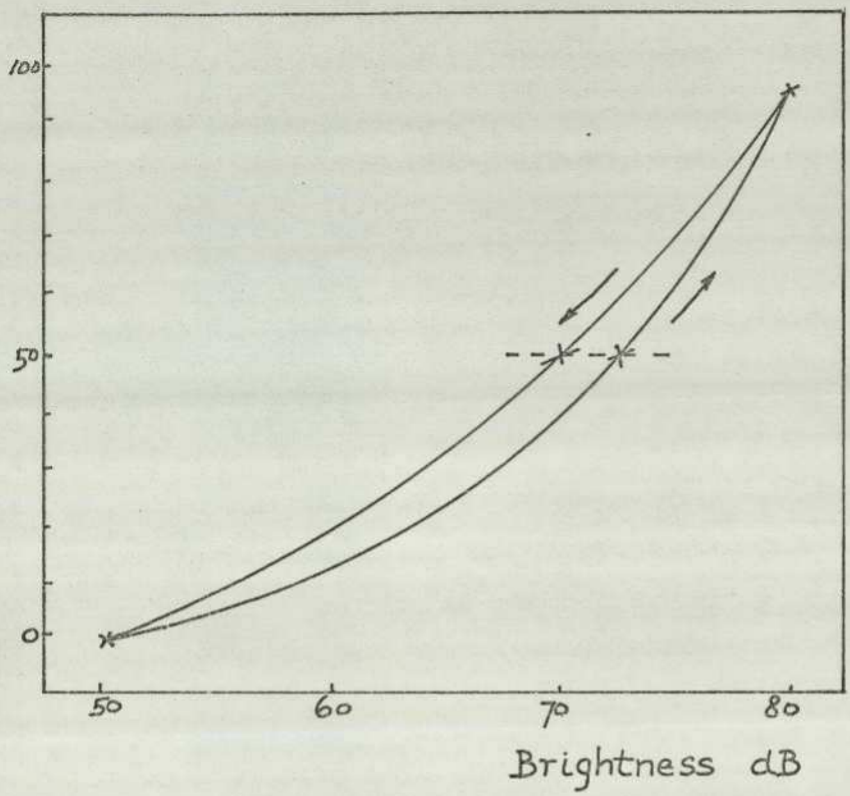


Fig 21.
Hysteresis.

The observer requires a higher brightness when the level is increasing than when it is decreasing.

brightness is halfway between the two reference levels. The experiment is repeated, but this time the level is reduced from the high reference level and again the observer is asked to say when the brightness is halfway down. It is usually found that there is an overshoot as shown for prothetic continua, and Stevens draws attention to this. He does not however say that a different result is obtained for metathetic ones.

Because of this uncertainty, this effect has not been related to the two classes of scale. However, it might be expected that a lopsided sensitivity curve such as is obtained with the geometric type of scale would give a different result from that obtained with a symmetrical distribution. This needs further investigation.

4.3. Conclusions on Validity of Models.

It is thus seen that Stevens' two classes, the prothetic and the metathetic, which he has distinguished on a phenomenological basis, correspond to the two classes of ratio scale, the geometric and arithmetic classes, which have here been distinguished mathematically. It is here suggested that no confusion need arise as to the type of experimental approach to be applied for measurement of physical and psychophysical phenomena. If the stimulus has a zero, and cannot be negative, such as a sound or light signal, then the only appropriate experimental technique is magnitude scaling such as has been extensively and successfully exploited by Stevens and his co-workers. If however the stimulus has a zero, but may be positive or negative relative to that zero, such as measuring a length relative to a standard and available measure, then Thurstone's model applies and the techniques employing the laws comparative and categorical judgement are appropriate.

If the dispersion of observations is small compared with the scale value, then the assumption of an arithmetic continuum and a normal distribution is a permissible approximation, even though the stimulus has a known zero.

5. CONDITIONS NEAR THE ABSOLUTE THRESHOLD.

In the above discussion, it was assumed that the signal to be measured was well above the threshold of observability. This condition implied that it was large compared with the ultimate discriminal dispersion which would be obtained as the signal was allowed to fall to zero. In all measuring systems, there are various sources of uncertainty. Some of these are of little significance over most of the measuring range, but become dominant near the limit of resolution of the equipment.

As a concrete example, consider a low noise amplifier preceded by an attenuator. Suppose that a signal is being amplified and its magnitude is displayed on an oscilloscope. The value of the signal can be measured by the size of the trace on the oscilloscope if the gain of the amplifier is known. It is assumed that the whole signal is displayed and that there is no backing-off voltage to improve the precision of measurement. The precision will be determined entirely by the observers ability to estimate the height of the trace. If the attenuation preceding the amplifier is increased, the gain of the amplifier will have to be increased to keep the trace the same size, but provided that the attenuator and amplifier are both correctly calibrated, the precision of the measurement will be unchanged, being limited entirely by the precision with which the trace can be measured on the oscilloscope. This situation is analogous to those which we have considered in the previous sections. The gain of the amplifier may be analogous to the adaptation mechanisms in an observer such as the iris of an eye, the pores of the skin and others of a chemical or electrical nature in the sensory and nervous systems.

If the attenuation before the amplifier is further increased, the state will be reached where the second source of uncertainty, the noise in the first stages of the amplifier will become imporatnt. The precision of measurement will now fall, because the noise due to the amplifier as well as the signal will be displayed on the oscilloscope. Further increase of attenuation and increase of gain in the amplifier will lead to the displayed output being due only to the inherent fluctuations in the measuring system, and at this stage no further information can be obtained about the signal.

In the same way, we should expect that the magnitude estimation methods and also the method of just noticeable differences would indicate a change of law close to the absolute threshold. In the case of the former method, this is not important, because both signal and response are plotted on logarithmic basis, thus relatively small portions of the curve are close to the threshold. Using the quantal method, Miller(12) has investigated the

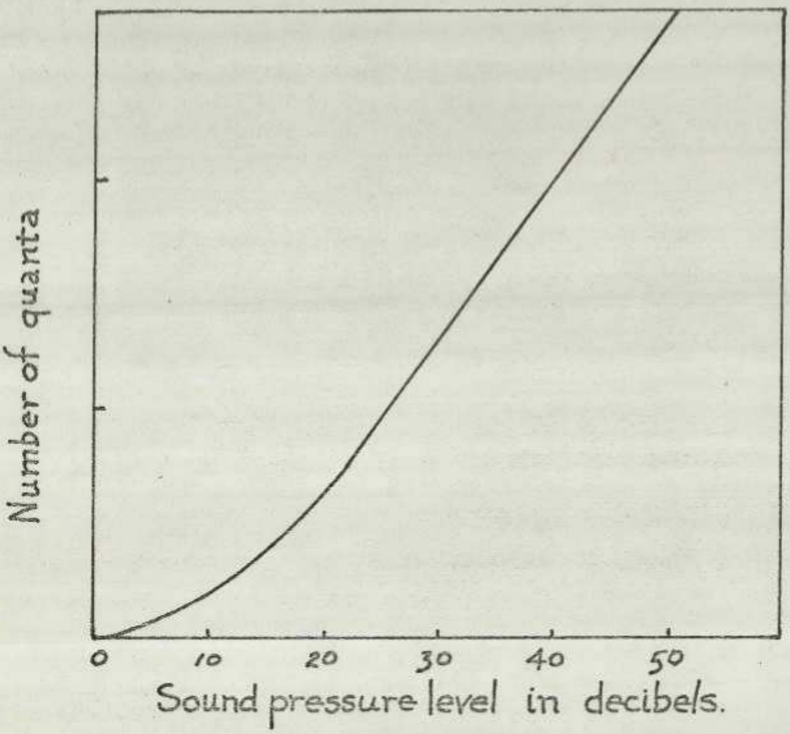


Fig 22.

Curvature occurs in the relation of jnd's to sound pressure level at low signal levels.

audibility of sound signals into the threshold region. In figure 22, is plotted the number of quanta (just noticeable differences) against the logarithm of the sound pressure level. It is seen that a linear relationship is obtained down to about 20 dB above the ambient noise level, when curvature sets in. This region which is commonly called the threshold region will now be discussed.

5.1. Threshold theory.

Two types of theory have been designed especially for explaining phenomena occurring at the limit of sensitivity of the observer. One type is based on the classical Fechner hypothesis of a threshold, the second is based on signal detection theory.

In the threshold theories, it is assumed that there is a threshold or absolute limen below which the observer cannot detect a signal. Above the threshold he can. This should lead to a well defined level of signal below which the observer would always say that there was no signal and above it he would say that there was. This is clearly not the case. Experiments of the constant stimulus type were designed to detect this threshold. A number of signals close in magnitude to the threshold level were shown to the observer (or observers), who had to say whether the signals were present or not. For each signal level a probability of detection was obtained and the level at which the signal was detected was derived using the methods discussed in section 3.2. This is inconsistent with the concept of a threshold. It therefore has to be assumed that there are fluctuations occurring either before or after the detection process takes place. The threshold is then defined as the value of the stimulus which arouses a response for 50% of the occasions tested. It is found that the curve of probability of a positive response against stimulus value is frequently not an anti-symmetrical curve, but shows a finite probability for zero signal. In some methods this is taken into account by assuming that for part of the time the observer answers purely at random. By such methods the theory may be made to fit the experimental data, but it should be noted that at no time does the theory suggest why there should be a natural threshold or what mechanism lies behind its existence.

5.2 Signal Detection Theory.

The signal detection theory as elaborated in a series of papers collected by Swets (25), differs completely in its foundations from those of the threshold theory. Instead of assuming a threshold, it assumes the existence of noise arising in the organism. The signal detection theory aims at measuring observers rather than measuring physical quantities or stimuli, It would thus be called a psychological rather than a psychophysical theory. The aim is to develop the receiver operating characteristic (R.O.C.) of the observer. In a simplified form, however, it has been used to measure sensations of warmth, touch and electrical stimulation.

The theory assumes that observation takes place in two stages. In the first stage, it is assumed that the stimulus or combination of stimuli give rise to a neural process which results in an attribute or state of the nervous system. It could be the number of nerve impulses in a given time. The level of this state is assumed to be monotonically related to the stimulus. This abstraction is called an 'observation'. The observation may be multi-dimensional.

In the second stage, the question is asked - "Given an observation, what will be the response of the observer?" To answer this in the case of a 'yes-no' experiment, it is assumed that when no

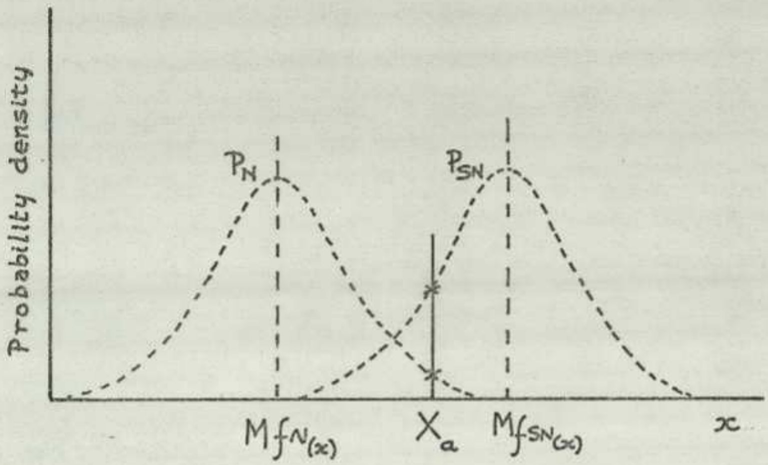


Fig 23.
Observations with noise.

$MfN_{(x)}$ is the modal value of noise alone, $MfSN_{(x)}$ is that of the signal plus noise.

signal is present, there is a noise level. For example, if the observer were detecting the presence of a spot of light against an illuminated background, the background on its own would give rise to an observation which fluctuated with time. This is called the noise. This noise will have a probability density which may or may not be of a Gaussian nature. Thus the probability that an observation lying between x and $x + \delta x$ will be made at any time is the ordinate of the probability curve for the noise alone. (figure 23)

When a signal is also present, it is assumed that a second curve will be obtained displaced from the first. For a one dimensional stimulus, the two probability density functions might be as shown in figure 23; one curve represents the noise and the other the signal plus noise. It is assumed that the observer carries both these distributions in his head. Then if at some time he makes observation X_a he will assess the probability that X_a arises from the signal plus noise by reading off the ordinate P_{SN}. He will assess the probability that X_a arises from noise alone from the ordinate P_N. He will take the ratio of these, P_{SN}/P_N, which is the liklihood ratio. He is now ready to proceed to the decision process. He is told the value of a correct decision and the costs of an incorrect one. He will then choose a value for the liklihood ratio to act as a criterion above which he will accept the hypothesis that the signal is present and below it the hypothesis that it is not.

It is seen that the liklihood ratio is a one dimensional variable even if the observation itself has many dimensions. The decision process is thus simplified. The subject has only to set up a criterion on the liklihood scale. If the liklihood is above the criterion he chooses one hypothesis, and if it is below it he chooses the other. His choice will depend on his rating of the value of success and the cost of failure. For example, if he is told that he must he careful not to overestimate the occurrence of the signal, he will tend to score low and so on.

A parameter, d', which is an index of detectability for a given observer, can be found if both distributions are one dimensional and have equal standard deviations. d' is defined as follows

$$d' = \frac{M_{fSN(x)} - M_{fN(x)}}{\sigma_{fN}}$$

where $M_{fN(x)}$ is the mean of the probability distribution of noise alone on the x scale, $M_{fSN(x)}$ is the mean of the probability distribution of signal plus noise on the x scale and σ_{fN} is the standard deviation of both distributions. d' is thus the difference between the observations normalized in terms of the standard deviation of the noise distribution.

Using this as a parameter, the receiver

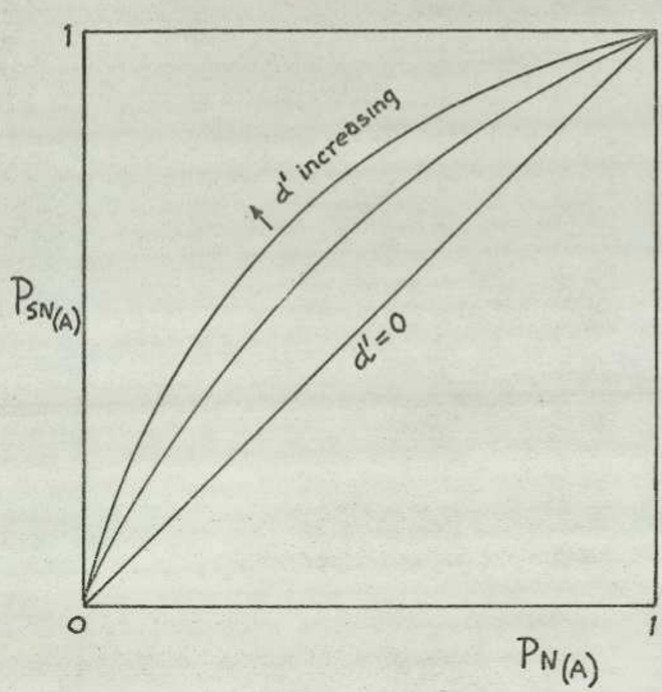


Fig 24.

Receiver operating curve.

$P_{SN(A)}$ is the probability of a correct answer given signal plus noise. $P_{N(A)}$ is the probability of an incorrect answer given noise only.

operating characteristic can be calculated on various assumptions. The value of the conditional probability that the observer will say "Signal plus noise" given signal plus noise is the area under the P_{SN} curve to the right of some criterion X_c. The conditional probability that the observer will say "Signal plus noise" when only noise is present is the area to the right of the criterion X_c under the curve P_N. These can be plotted to form a receiver operating characteristic given the conditions (A) (figure 24).

Clearly if the signal is indefinitely small, the two probability density curves will coincide and we have the diagonal $d' = 0$. As the signal increases d' increases so that the value of P_{SN} is greater than P_N for all values of X_c. When X_c is set very low, the observer will find the probability P_{SN} and P_N are both the same and are equal to unity. Similarly if X_c is set very high, he will again find the probabilities equal, both will be zero. Between the two, the probability of a correct answer "Yes, signal is present" is always greater than that of an incorrect answer "No, signal is not present" when the signal is displayed. If the value of the signal is increased, d' increases and the deviation from the diagonal on the ROC is greater.

Now, how will the criterion X_c be set up? It is assumed that the observer is given an a priori set of values and costs. From these he will try to maximize the expected value of his decision.

Suppose that he takes the ratio β as the liklihood ratio

$$\beta = \frac{f_{SN(x)}}{f_{N(x)}} = \frac{P_{(\bar{s})}(V_{\bar{s}\bar{S}} + K_{\bar{s}S})}{P_{(s)}(V_{sS} + K_{s\bar{S}})}$$

where V and K represent respectively the values of a correct decision and the cost of a wrong one. S represents a positive response, $\bar{S}$ represents a negative response. s represents the signal being present and $\bar{s}$ represents the signal not being present. Thus V_{sS} is the value of the correct response when the signal is present. It is shown that if the pay-off matrix of values and costs as shown below is assumed,

	Stimulus	
Response	s	$\bar{s}$ (= not s)
S	V_{sS}	$K_{\bar{s}S}$
$\bar{S}$ (= not S)	$K_{s\bar{S}}$	$V_{\bar{s}\bar{S}}$

then the criterion for the liklihood ratio to maximize the expected value is

$$p_s(S) - p_{\bar{s}}(S) \text{ is a maximum}$$

where $p_s(S)$ is the conditional probability that, if the signal were present, the observer would report that it was present. $p_{\bar{s}}(S)$ is the conditional probability that, if the signal were not present, the observer would say that it was. $P_{(\bar{s})}$ and $P_{(s)}$ are the a priori probabilities of 'no signal' and 'signal' respectively.

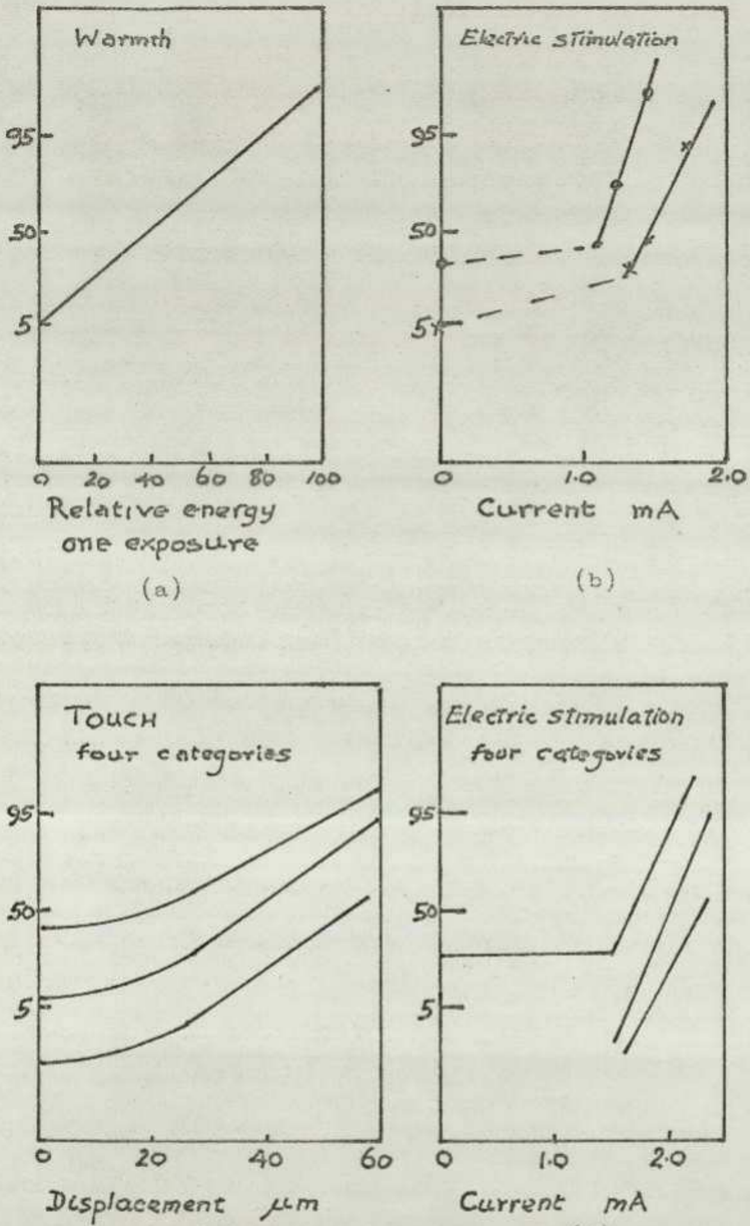
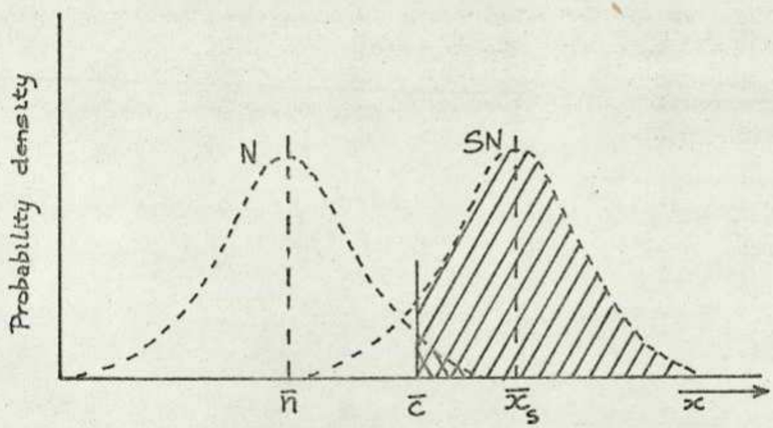


Fig 26

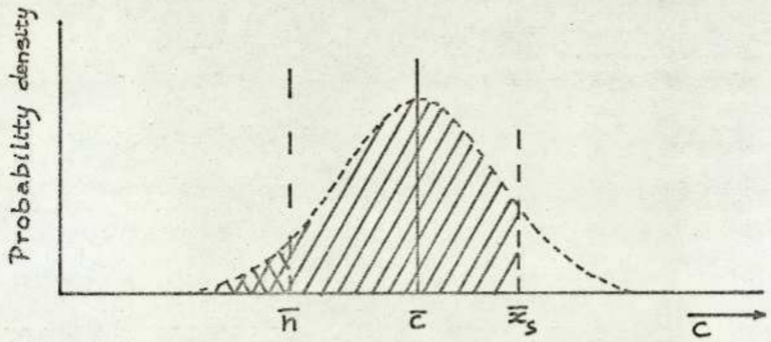
Probability of detection as a function of signal strength:

(a) Warmth
(b) Electric Stimulation
(c) Touch (four categories)
(d) Electric Stimulation (four categories)

(Eijkman and Vendrik)



(a)



(b)

Fig 25.

Probability density measured on the x - scale (a) and on the c - scale (b).

To maximize the expected value, we require the line on the receiver operating characteristic to be touched by the line from the origin whose slope is β.

If the pay-off matrix is altered by altering the costs and values, the slope of the tangent line and hence the value of the critical liklihood ratio will be altered.

Having said all this, we find that since the criterion is a liklihood ratio, it is completely defined. Thus although there is noise in the system, the observer will always make the same decision when a certain signal is presented. The theory does not seem to predict the variation from one observation to another that in fact occurs. It is therefore necessary to postulate a variability in the criterion.

The following seem to be the processes through which the observer must go in order to make his decision.

(1) Enter the a priori probabilities into his memory and develop his pay-off matrix.

(2) From the pay-off matrix establish his criterion.

(3) Make the observation, which may be multi-dimensional, and convert it into a point on the probability density curve. At the same time he must transform his multi-dimensional scale into a one-dimensional liklihood scale.

(4) Consider the probability density curve for the system in the absence of signal, again on the liklihood scale.

(5) Take the ratio of ordinates at a value of liklihood corresponding to his criterion.

(6) If the liklihood ratio is greater than his criterion, say "Yes, there is a signal"; if it is less, say "No, there is no signal"

This seems to make so great demands on the observer, that it is hard to imagine that he could accomplish all these operations automatically. However Eijkman and Vendrik simplified the theory in their investigations of warmth, touch and electrical stimuli. They used signal detection theory to determine relationships between the sensations above and the magnitudes of the stimuli. To carry out the analysis, they made the following assumptions. There is noise in the organism of the type already discussed, giving rise to a mean value $\bar{n}$ and having a variance σ_n^2. When a signal is present, no additional noise is introduced. Thus the density curve is shifted to a new mean value $\bar{x}_s$ on the observation scale x. A criterion c is set up such that if the observation appears to be above the criterion, the observer will say that a signal is present, if it is below, he will say that it is not. The criterion also may not be constant. If this has a variance σ_c^2 and if all distributions are Gaussian, it will be equivalent to having noise with variance

$$\sigma^2 = \sigma_n^2 + \sigma_c^2$$

and a constant criterion. (figure 25(a)).

This model can be shown to be equivalent to a

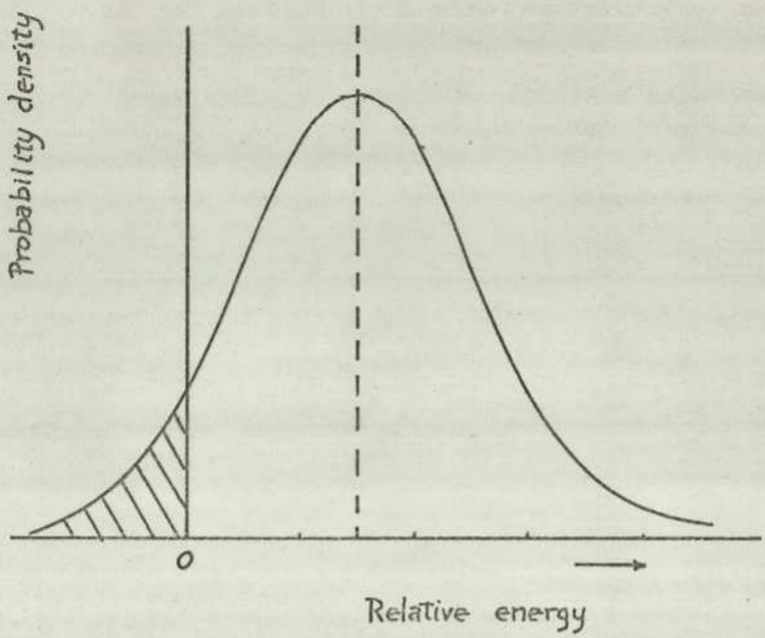


Fig. 27

The curve of Fig 26 (a) implies a distribution of
the type shown above.

Facing p 70

single Gaussian distribution for the criterion having a variance σ^2 and no noise on the signal or receiving system. The new scale unit c is equal to $(-x + \bar{c} + \bar{x})$. Comparison of the shaded areas of figure 25(b) shows this to be the case.

Eijkman and Vendrik carried out experiments in which five stimuli were presented in random order, one of which had zero amplitude. The stimuli were presented in a given time interval. The observer had to respond either Yes or No. No cost or values were suggested for a correct answer. The proportion of positive responses were plotted against the signal strength on arithmetic probability paper. Results showed, in the case of warmth, a linear relation intercepting the zero ordinate at about 6% (figure 26(a)). This the authors state is satisfactory agreement with the theory that the distribution is Gaussian. They say " No significant deviation from a Gaussian distribution was found." No comment is made about the intercept on the zero axis, indicating that the probability density function is truncated by the zero stimulus ordinate as shown in figure 27, and suggesting a probability of a positive response for negative signals. It will be seen that the theory has led us to the classical constant stimulus method of measuring the absolute limen. All objections that apply to that method apply to this one.

Almost identical results were obtained for

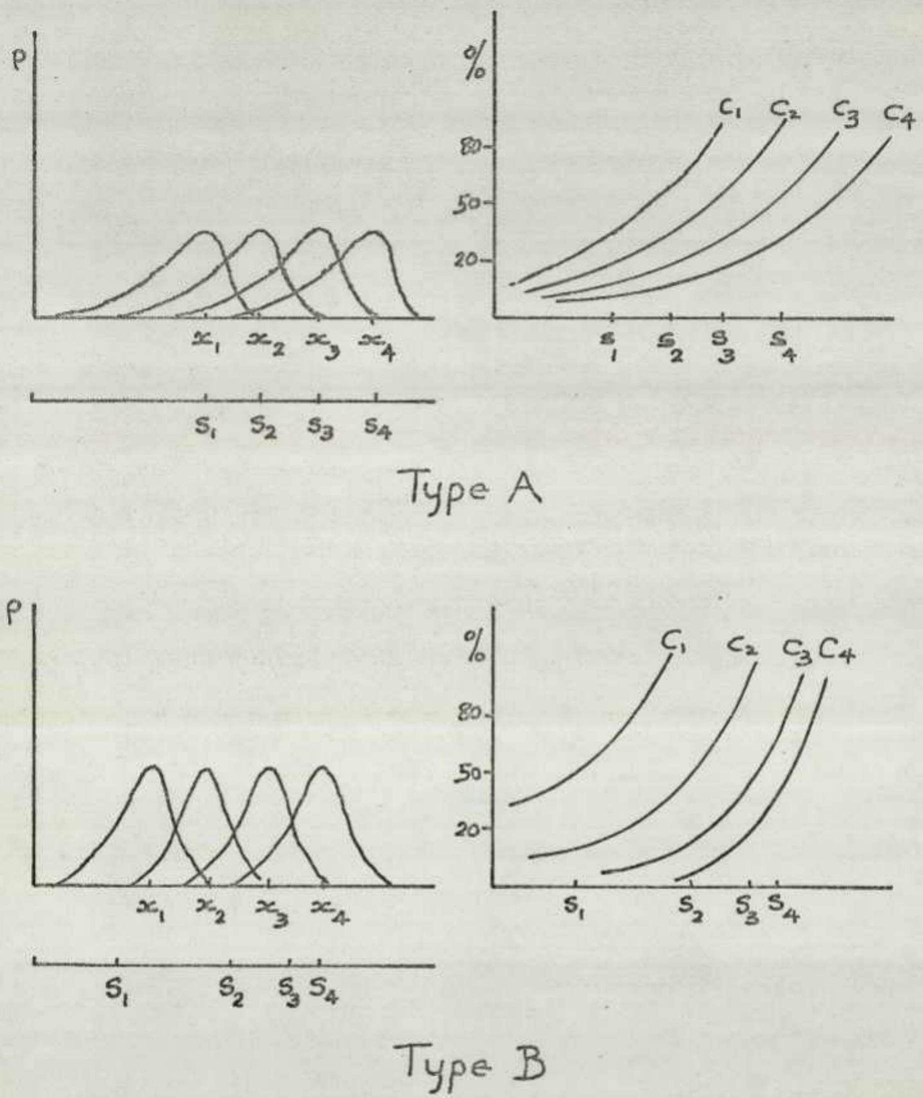


Fig.28.

Distributions and their corresponding probability curves.

Type (A) Skew distributions linearly related to signal

Type (B) Normal distributions non-linearly related to signal.

touch, but for electrical stimulation a different pattern was obtained in which it appeared that a Gaussian distribution was obtained for large signals, but for zero signals the probability of a positive response could be as high as 30% (figure 26(b)).

The authors suggested that this could be due to skewness in the distribution of the noise on the x axis or due to non-linearity in the relation of x to signal. To distinguish these two hypotheses, they asked the subjects to say which of the five stimuli was being presented. They thus obtained four monotonically related criteria. If the distributions were skewed, but the relation of signal to x was linear, they expected distributions of type A in figure 28; whereas if the distributions were symmetrical but the relation of signal to x was non-linear, they expected curves of type B. (N.B. The vertical scale is that used in probability paper.) Experiments were performed on the sense of touch and of electrical stimulation, and the results appeared to be of type B. These are shown in Figure 26 c and d.

In considering these results, it is interesting to note that the theory had had to be simplified to include only Gaussian distributions. In this last discussion non-Gaussian distributions were considered but the theory was not changed. It is not clear what type of non-Gaussian distributions would allow translation to be made from the x axis to the c axis. It would appear that skew distributions are excluded, and yet the authors have assumed that such a translation is permissible.

Also the very simplified model used in this paper, in which Gaussian noise and random fluctuations of the criterion are assumed, now reduces exactly to Thurstone's model. So that statistical decision theory has only led us back to the original starting point. This does not seem to have been discussed by the authors.

It now becomes clear that this theory in no way explains the mechanism of the limiting sensitivity of a sensory system, in that it derives from a normal distribution on a semi-infinite linear continuum. One is left with the awkward question of what is meant by the finite probability of response with zero signal, and what happens to the distribution to the left of the zero stimulus ordinate. (figure 27).

The question of integration time does not seem to have been discussed by the protagonists of signal detection theory. Fitzhugh(8) gives an account of experiments on the pulses from the ganglion cell of the eye of a decerebrate cat. He found that the impulses occur apparently randomly at an average of 0.5 per 10 ms for a dark adapted eye, but the rate rises after exposure to a flash of light of 5 ms duration to up to 3 per 10 ms. The time for the cell to recover its normal background count after a flash is of the order of 200 ms.

Thus if the human eye behaves in about the same way, the observer might average his measurement over 100 ms and should obtain a value of average noise with a high degree of certainty i.e. with little spread in value. Similarly in this same time he would obtain a fairly precise value of the rate of pulsing due to the signal. Thus when he calculates the liklihood ratio, it will be quite precise; and if the criterion is precise there will be no variation in his answers. One must therefore assume that the fluctuations in the criterion are slow compared with the decision time. If this were not so, then the observer would be able to integrate over a number of fluctuations of the criterion and again would always produce the same answer.

This assumption means that when one decision is made the criterion will have one value on the c axis, and when another decision is made on the same data the criterion will have another value. Some long term fluctuation seems to be essential to any theory of decision processes.

6. FLUORESCENT LIGHTING EXPERIMENTS.

In papers 2 - 5 a series of psychophysical experiments aimed at measuring colour differences are described. Paper 3 is the key one in the set. It describes the background and purpose of the trials in the Introduction. The experimental arrangement and initial analysis is described in Part 1. In Part 2, the Author put forward a new model for the mode of discrimination of an observer, based on Thurstone's judgement model and derived the form to be expected of observational data. Finally in Part 3, data from two trials in the observation of colour differences are analysed using the new model. It was found to work well in predicting the form of the data actually obtained. The data for Paper 3 was the most extensive of all data obtained in this series and for this reason it has been subjected to further analysis in this thesis.

Two experiments were made using different groups of observers. The questions put to the observers were similar except that in the second set they were asked to assign the number 0 to the category of "No difference" instead of category 1 as in the first set. This was felt to be more logical. So far as the presentation was concerned, the opportunity was taken to introduce some familiar objects, notably a red scarf, an orange and a vase of artificial flowers. Questions were asked about

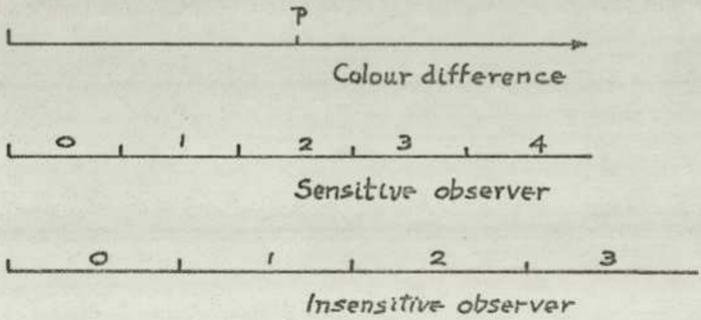


Fig 29.

Effect of observer sensitivity on category discrimination.

these objects as well as about the test colours, but no satisfactory use was made of the answers.

6.1. Study of Observers.

In a preliminary experiment in which the room was tested, it had been shown by analysis of variance that the two sides of the room were balanced, i.e. observers did not see larger differences on one side than on the other, nor was one side biased towards one hue. It was also shown that, for the group of observers used and for the colour differences involved, the observers all came from the same population. This came as a surprise to the industrial organization which employed them, since some observers studied colour as part of their job whilst the rest were not professionally involved. It was therefore decided to examine the performance of the observers in the main experiments.

In the preliminary experiment an analysis of variance had been used. This could be criticised in that it assumed that scores were distributed normally on a continuum measured by the category boundaries. That this was not true was evident from the distributions obtained which were markedly skew. However a variation in the sensitivity of observers would expand or contract the category scale on which measurements were made.

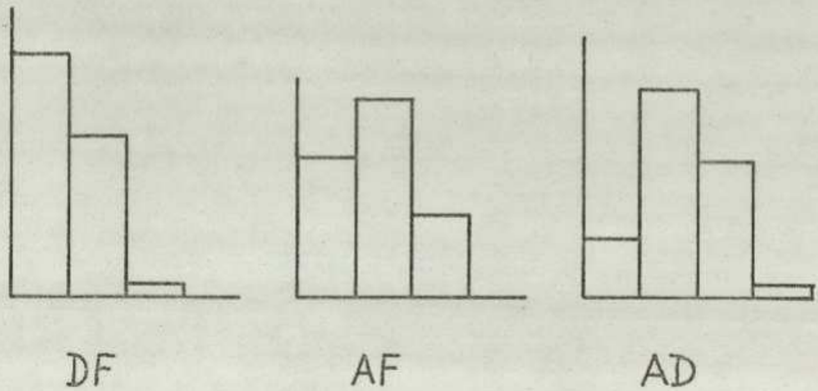
As seen in figure 29, if P represents a point on a physical colour difference scale, then a sensitive observer would place this in a high category and an insensitive one would place it in a low one. The observer sensitivity would thus impose a log-normal variability on the scores.

To examine this, the results of experiment 4 were used. In that experiment every observer saw the test colours under four different lamp combinations. e.g. observers in set 1 saw lamp combinations AD, AF, DE, EF, those in set 2 saw AE, AF, EE, EF, and so on.

The sensitivity of each observer was assumed to be proportional to his total corrected score within a set of 12 people. Each score was corrected by adding 0.5 to it, so that it scored as if it were in the middle of its category. Thus category 0 scored 0.5, category 1 scored 1.5 and so on.

It was assumed that within each group of 12 observers there was a normal spread of population. This was reasonable, since care had been taken in the design of the experiment to balance the number of experienced and naive observers in each set. If then the sum of the corrected scores for each observer arising from four observations was divided by one twelfth of the sums scored for the 12 observers in the set, a sensitivity factor could be obtained for each test colour.

Since the average of 12 was a rather unreliable normalization factor, the variability between the scores in the sets was investigated. It was found however, that there were significant differences between the lamp groups and thus it was not possible to normalize with a single factor for all observers. Normalization was therefore carried out within each set of 12 observers and for each test colour. Hence for every observer a sensitivity factor was obtained for every test colour. Analysis of variance within groups of 12 observers showed that the ratio of the variance having the highest value (12 degrees of freedom) to the mean variance (155 degrees of freedom) was 1.5. This ratio was not significant at the 0.05 level. It was therefore concluded that there was no significant difference in the variability of observer sensitivity. Likewise, comparing the variance of average sensitivity with the average variance of sensitivity for sets of 12 observers, it was concluded that there was no significant difference in the sensitivity of the observers. That this was so was demonstrated by dividing all original scores (duly corrected) by the sensitivity factor and comparing the adjusted scores with that of the original scores . This was done for a few colours and lamp combinations. The value of the variance of the original scores and of the adjusted scores are tabulated below.



Test colour 5

Fig 30.

Distributions of scores for three lamp combinations.

Variance of Scores

	White					Colour 8	
	AD	AE	AF	DE	DF	AD	DE
Variance of unadjusted scores	.48	.45	.76	.55	.54	.75	.77
Variance of adjusted scores	.56	.60	.65	.50	.52	.48	.61

It is seen that there is no systematic reduction of variance when adjustment is made for observer sensitivity. However, in the case of colour 8 and for lamps as different as A and D, there is a suggestion that the variance may be improved.

It was therefore concluded that there was no reason to believe that one observer was more sensitive than another when making observations close to the limit of resolution. However when definite colour differences should be visible, observer sensitivity could become important. For the purposes of the analysis of results, all observers were treated as being equally sensitive.

6.2. Analysis of Colour Difference Estimates.

The author rejected the methods of analysis described in previous sections firstly because they could not be reconciled with a reasonable model of the observers decision processes and secondly because all distributions were markedly skew. Typical results are shown in figure 30. The reasoning which

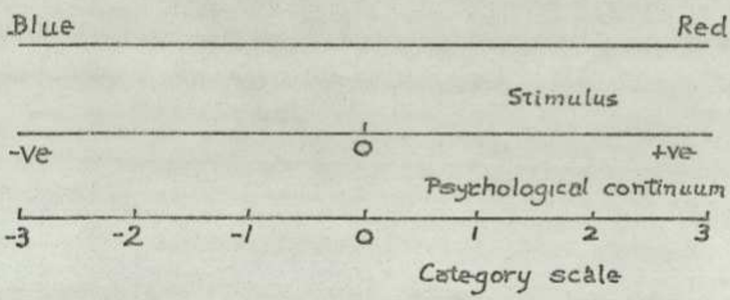


Fig 31.

Three continua in colour scaling.

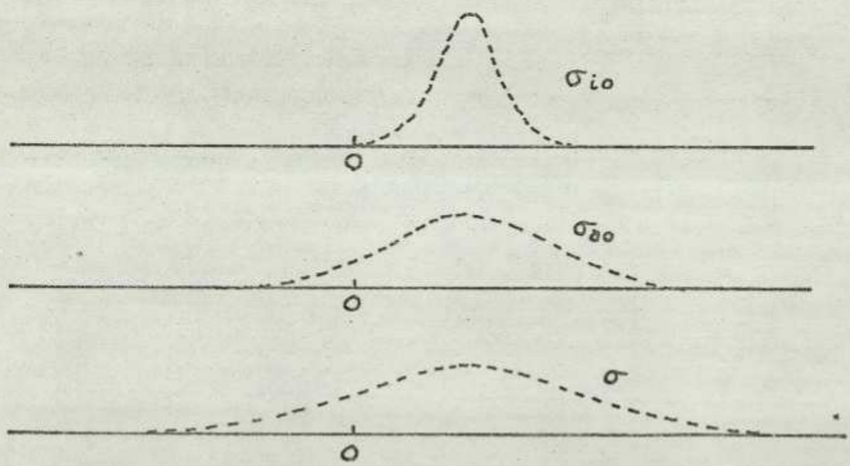


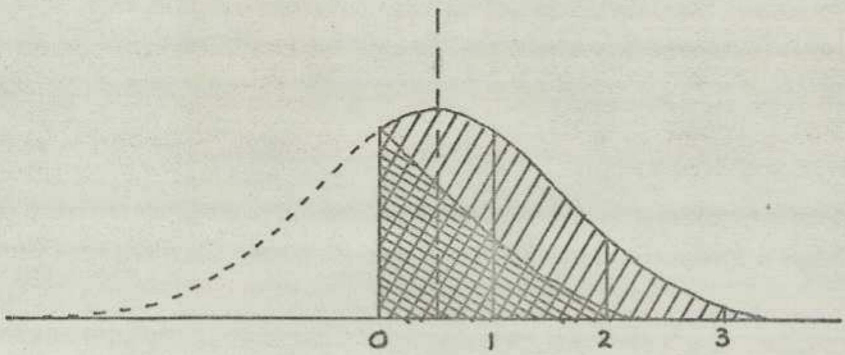
Fig 32.

Total dispersion arises from uncertainty of each observer and variations between observers.

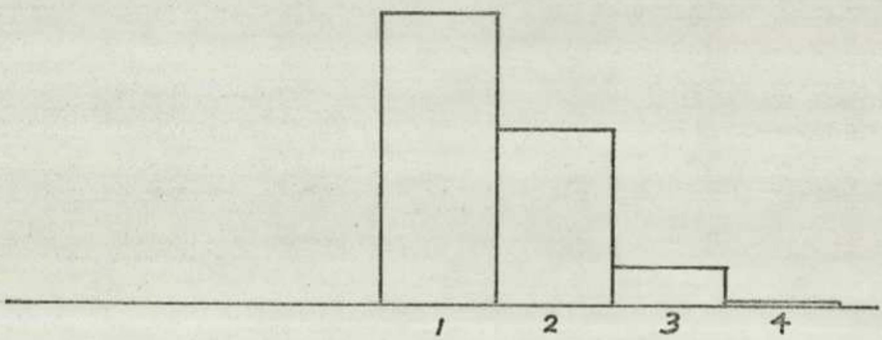
led to the above decision is given in Paper 3 pt 2 p 109 et seq. However, since it is so crucial to further discussion, some of it is repeated here.

Because they have been shown in the past to fit situations involving categorical judgements, the assumptions of Thurstone (see Section 3.1.1.) were adopted for the analysis together with the additional assumption that, in the region of the threshold the widths of the categories were constant. Thus applying them to the problem of estimating colour difference, we assume that a certain stimulus gives rise to a discriminal process. In these experiments, the stimulus was the apparent difference in colour between similar coloured samples illuminated by different lamps. To each discriminal process is assigned a position on a linear continuum. If, for example, the stimulus lay along the Blue-Red axis (figure 31) and the right hand sample appeared redder than the left hand one, a positive number would be assigned; if it appeared bluer, then a negative number would be assigned. Since the continuum stretches far in each direction towards the red and blue ends respectively and "No difference" is in the middle, then Thurstone's model applies.

It is assumed that if the same sample were shown a number of times, the discriminal processes would be distributed normally on the psychological continuum. It is also assumed that, if one comparison were presented once to a number of different



(a)



(b)

Fig 33.

Distribution with modal value 0.5 is folded about the zero axis (a) resulting in categorized distribution (b).

observers, the discriminal processes would also form a normal distribution on a single continuum. It is unfortunate that no method is available to distinguish the contribution of each factor to the total variance. This cannot be done since both sources of variation lead to Gaussian distributions. Thus the total variance of discriminal processes is

$$\sigma^2 = \sigma_{i0}^2 + \sigma_{a0}^2$$

where σ_{i0}^2 is the variance of a single observer and σ_{a0}^2 is the variance among observers of a single stimulus.(Figure 32)

Furthermore, we cannot estimate the mean value of the position of the discriminal processes from the mean value of the scores. This is evident from the skewness of the distributions. The reason for this is that there is a translation by the observer between the psychological continuum and the category scale. This, it is suggested, occurs as follows: Consider the diagram figure 33. The horizontal line represents the continuum on which judgements are made. 0 is the zero point. Owing to fluctuations in the organism, the observer receives a range of apparent stimuli which form a distribution on this continuum. Taking 1 as the scale value of the just noticeable difference, if the discriminal process lies between 0 and 1, the observer will say "No difference" and will be scored 0.5 (the mean value of the category). If the discriminal process lies between 1 and 2, he will say "Just noticeable." and will be scored 1.5 etc. If, however, the discriminal process lies between 0 and -1, he will still say "No difference." and be scored 0.5, since close to the threshold he has no way of knowing in which direction the colour difference lies. This certainly seems to apply to the first two categories, i.e. the observer is conscious of a difference before he can analyse his experience and say whether the colour of the sample is shifted towards the blue or to the red relative to the reference sample. The same applies to discriminal processes lying between -1 and -2. This will be rated as "Just noticeable." and will score 1.5. Thus the scores available for analysis are the result of a folding process about the zero. This is illustrated in figure 33(a). It is thus seen that, if this theory is correct, the distribution of scores will be very skew, particularly near the condition in which both sides are identical. The resulting category scores are shown in figure 33 (b). (It is here appropriate to refer to the method of Coombs referred to in section 2.1. It is seen that the two concepts have no relation to each other apart from the use of the concept of unfolding in an attempt to obtain a workable distribution from an unworkable one.)

A further assumption had to be made namely that, when both sides are similarly lit, the mode of the distribution before folding will be at zero on the psychological continuum. This assumes that neither the room nor the observer has a bias to one or the other side. It also assumes that the time-order error (if Stevens' explanation is correct (section 4.2.3.)) is zero. This is consistent with the model used. It should be noted that the shape of the distribution of scores on the category scale depends on the discriminal dispersion and on the mean value. If the mean value is large, there will be few discriminal processes on the negative side of zero and hence the distributions will be normal. If however, the mean value is small, the contribution from the negative side will be large and hence the total distribution will be markedly skew. Equally if the discriminal dispersion is large, the contribution from the processes on the negative side will spread into several categories and again make the resultant distribution skew.

It is thus clear that no methods of analysis of the types described in Section 3.2 can justifiably be applied to this model to translate the raw data into a probability distribution and hence to determine the discriminal dispersion and the category boundaries. A new method of analysis had to be developed.

It should be noted that other methods of analysis had been rejected a priori because they did not apply to the physical model used. It is here necessary to emphasize the importance of a physical model as a guide. It would have been possible to process the data according to one of the sets of rules which apply to one of the mathematical models and no doubt some sort of results would have been obtained, However no further light would have been thrown on the original investigation and the explanation of the results could have been even more confused than the original data.

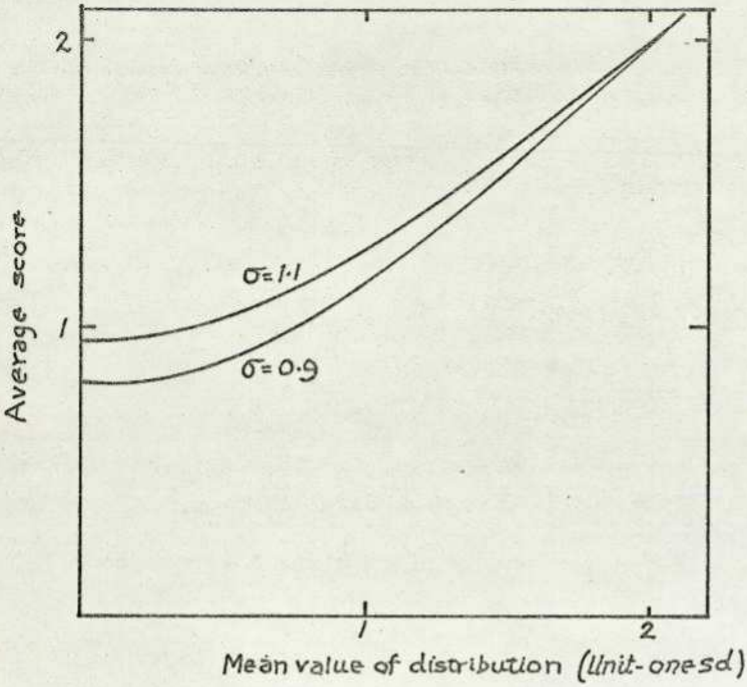
In this instance the model was not assumed to be true, but it was at least a hypothesis consistent with a possible physical mechanism of the observer. It could then further be investigated. The investigations showed considerable agreement between predictions and results, and there is every reason to believe that the hypothesis has some truth in it.

6.3 Techniques for Analysis of Folded Data.

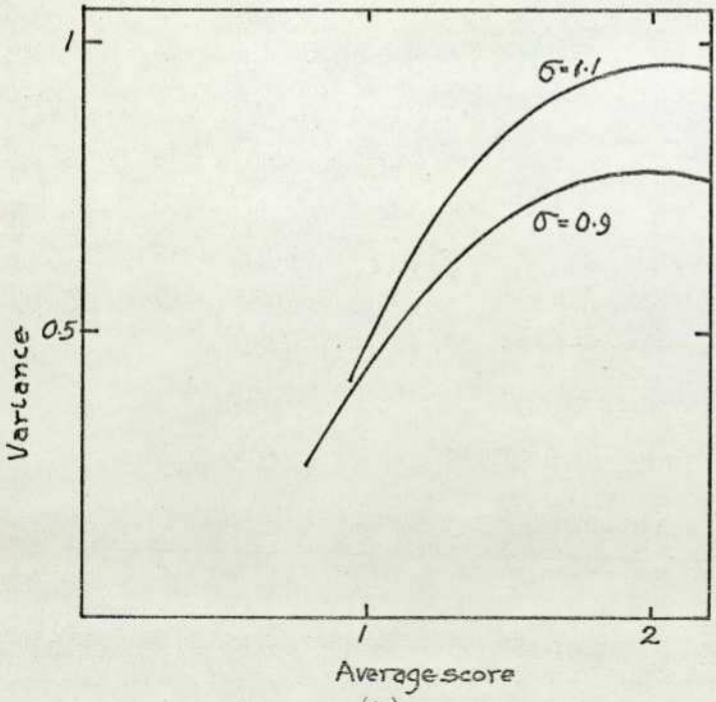
In what follows we assume that Thurstone's postulates apply, that observers use equal category widths for all categories and that the observer is unable to distinguish direction in estimating differences. On this basis, a model is set up and the predictions of the model are compared with the data.

Because of the 'folding' process outlined above, in which the scores in two different categories on the psychological scale contribute to scores in a single category on the judgement scale, there was no possibility of analysing the data directly. Instead, the model had to be used to make theoretical predictions. These predictions were then compared with the experimental results.

The program PROC (Appendix 1) assumes that a normal univariate distribution having a certain value of discriminal dispersion exists on the psychological continuum. The distribution is supposed to be displaced from the origin by an amount corresponding to the mean value of the discriminal processes. It then calculates the scores to be expected in each category based on the folding process described above and calculates the corresponding mean value and variance. In taking the mean value, no assumption is made that this is the central tendency, therefore no assumption about the



(a)



(b)

Fig. 34.

Calculated relations for the univariate distribution.

distribution of the actual scores among the categories is made. The mean value is merely used as a convenient number to compare with that obtained from observations.

Figure 34(a) shows the relationship of average score to mean value for a range of values of dispersion. It is seen that when the mean value is zero (This corresponds to no difference between the colours), the average value is finite and increases with increasing dispersion. This value was used to find the value of dispersion for each test colour. By comparing lamps of the same type i.e. giving zero colour difference, the dispersion could be estimated. The relation of this minimum value to the dispersion is found to be substantially linear. (Paper 3 (2) fig. 9 (a))

The relation between variance and average score is predicted to be as in figure 34 (b). It is seen that the variance is a maximum when the average score is halfway between the category limits. This effect is due to using a limited number of categories. Also the variance increases as the dispersion increases. For each value of dispersion, there is a minimum variance. This minimum value increases approximately linearly with dispersion as shown in Paper 3 (2) fig 9(b).

It is satisfactory that the curve of average score against position on the continuum (figure 34 a) is of the 'threshold type'. The threshold has

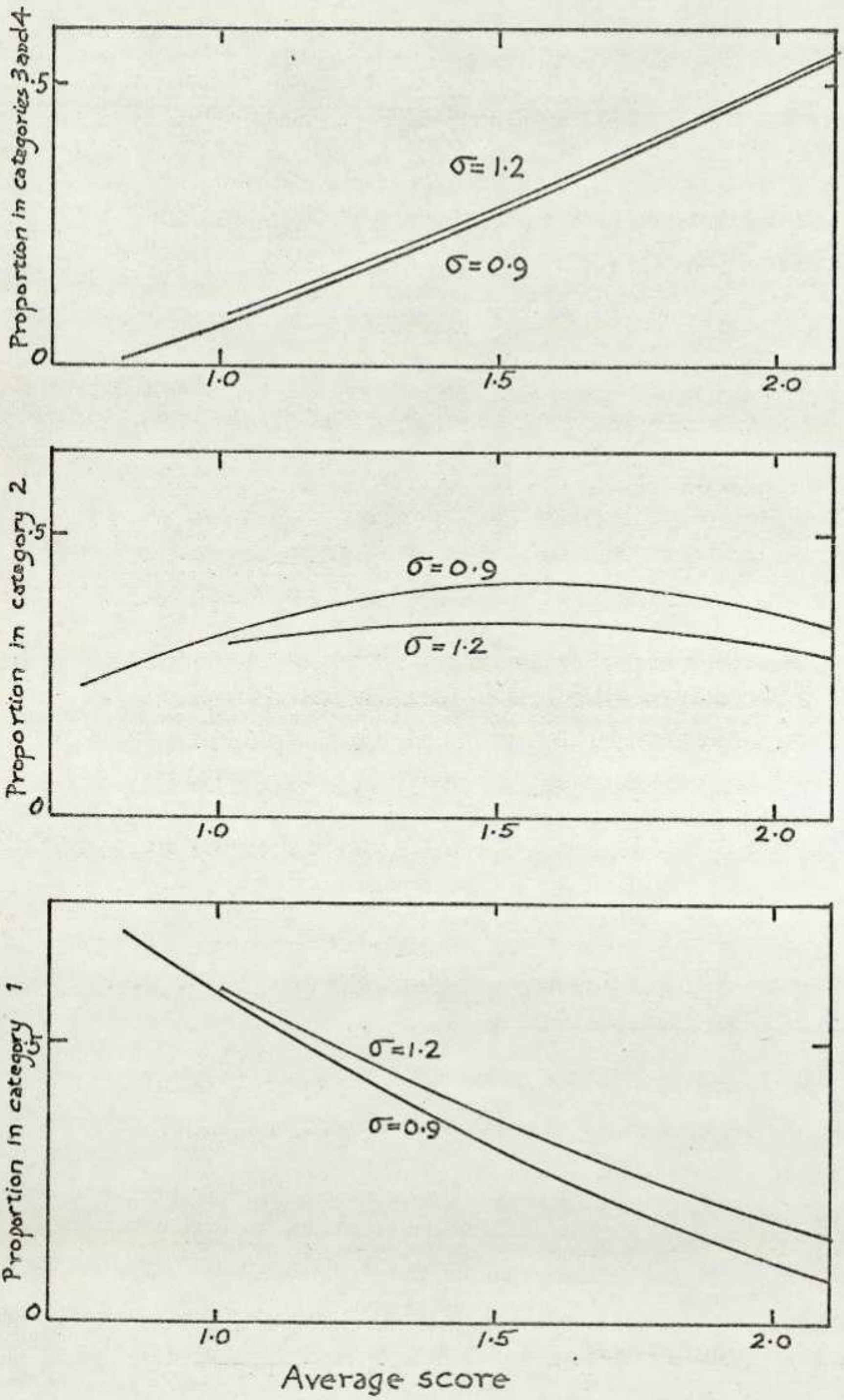


Fig. 35.

Proportions of scores in each category.
(Univariate distribution.)

appeared naturally out of the physical model and has not had to be put there by some special process or by the addition of an arbitrary constant. There is a steady score when the stimulus is zero. As the stimulus is increased, the score increases initially with zero slope. Thus qualitatively the predictions are in accordance with experimental findings.

We also note that the experimental plots of variance against average score (Paper 3 (3) fig 2(a)) indicate a relationship of the type predicted in figure 34 (b) and show considerable agreement for some test colours. However for others (Paper 3 (3) fig 2 (b)) the shape is reasonable, but the points do not agree in value. In particular the values of variance found for some colours when similar lamps are compared (zero signal) are lower than the theoretical minimum. When it is remembered that only 36 or 48 observers are involved, this agreement between prediction and experiment for a second order function, a function of the squares of the deviations, seems remarkable.

The proportions of scores to be expected in each category (categories 3 and 4 are lumped together as the score in category 4 was usually small) are shown in figure 35 as a function of average score. This data can be used to predict the actual number of scores to be expected in each category. The χ^2 test can then be applied. This has indicated

χ^2 values well within an acceptable level of significance. Results of this test will be given in section 6.4, when the bi-variate distribution is discussed.

The conclusion from the analysis in Paper 3 was that the model agreed with the data for half the test-colours, but that the agreement was less good, notably in respect of the variance, for the other half. The actual variance found was nearly always less than that predicted.

However the analysis enabled some suggestive results to be obtained indicating linear relationships between the position on the psychological continuum and the colour shift of the 1960 C.I.E. uniform chromaticity scale (u,v). Further assumptions, this time on the physical side had had to be made, these were

(a) That the discriminal dispersion was constant for one test colour for all colour shifts. This was known to be an approximation since the so-called 'uniform chromaticity scale' is not uniform.

(b) That an adaptation formula used in Publication 13 was valid. This was suspect at the time and has been now replaced.

(c) That all colour shifts were along a single axis and that dispersion normal to that axis was not significant.

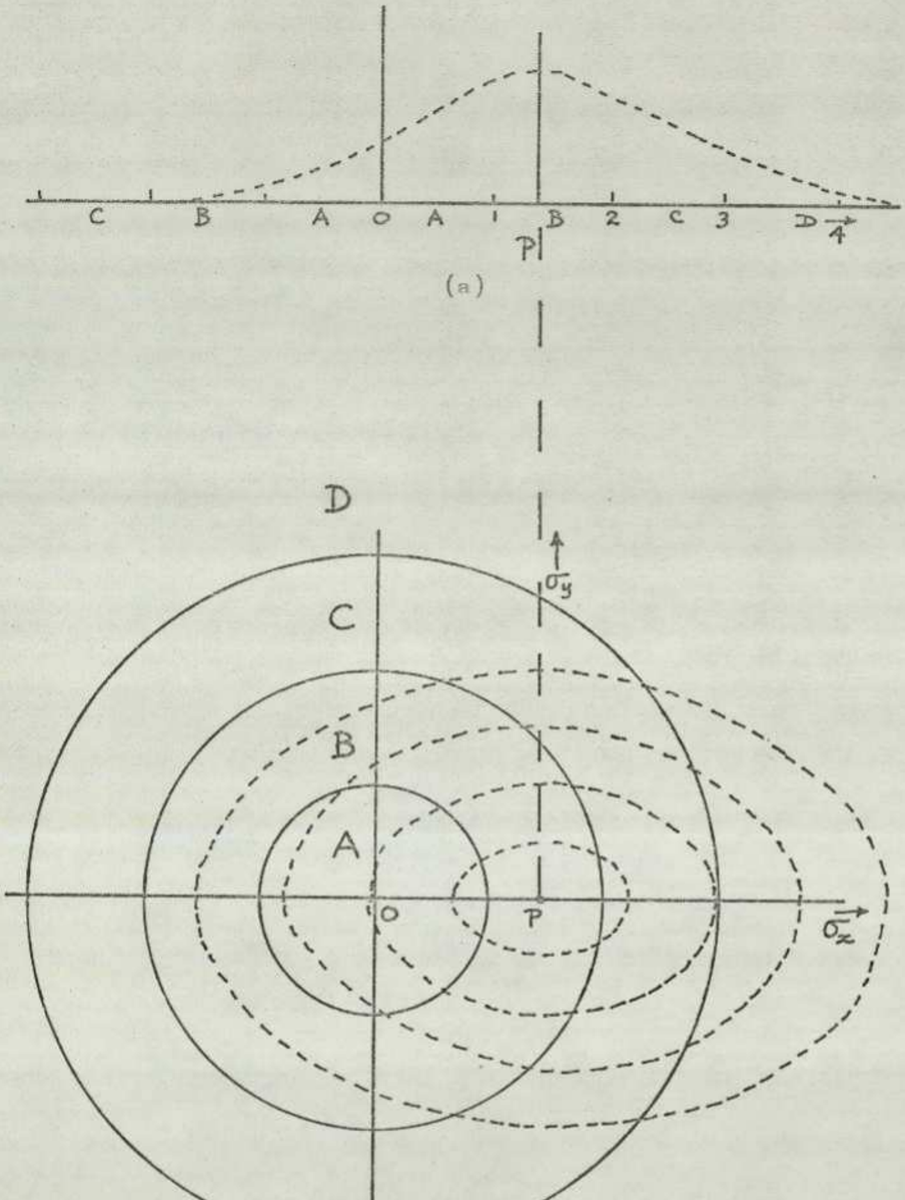
However, despite these assumptions useful results were obtained which enabled tolerances to be proposed for application to special colour rendering indices.

6.4 Bivariate and Univariate Distributions.

In the discussion of Paper 3, it was suggested by C.E.Williams that a bivariate normal distribution should be used instead of the univariate one. This would take into account the observers' variability along a colour axis at right angles to that along which the actual colour difference occurred. That such an allowance was desirable had been discussed among the authors of that paper, but at that time no analysis had been made, since the model had worked well enough to indicate significant relationships in colour rendering and had indicated some of the deficiencies of the method of Publication 13; the investigation of that document had of course been the primary purpose of the study.

Since the time that Paper 3 was read, the Author has investigated the use of the bivariate distribution numerically, using methods similar to those used in the first investigation. It was assumed that corresponding to the u.v. colour surface, there is a two dimensional continuum on the psychological plane. Thus the observer's variability has two degrees of freedom. The sensitivity in the two directions may not be the same. In the calculation it was assumed that the discriminal dispersion in the direction corresponding to the vector shift was σ_x and that at right angles to it was σ_y.

The corresponding probability distribution



(b)

Fig. 36.

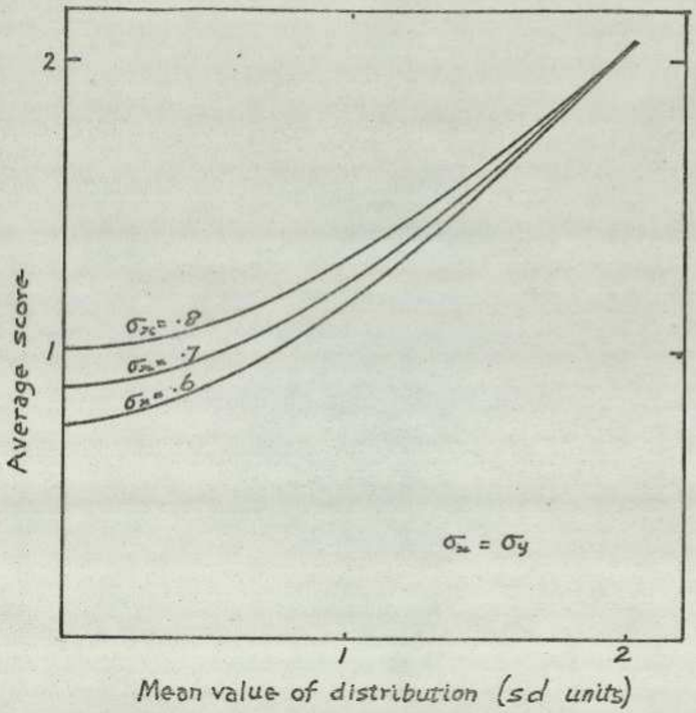
Bivariate distribution.

Stimulus is at P. Areas A, B, C, and D are the categories.

would then be an elliptical shaped solid whose section in the x or y directions is a Gaussian distribution. Figure 36 shows such a distribution in elevation and in plan. It is seen that the section has the same shape as that used for the univariate case. O is the condition of zero stimulus. P is the location of the stimulus. Thus the stimulus-value is the vector OP. If the subject responds to magnitude only and not to direction, then the circle of unit radius contains the location of all 'non-noticeable' stimuli. The circle of radius two contains all the 'just-noticeable' and 'non-noticeable' stimuli. These are circles because the unit of measurement in both dimensions is the discriminal dispersion.

The apparent variability of the stimuli about the point P is represented by an elliptical solid whose section along the major and minor axes is a normal distribution. The angle between an axis of the ellipse and the vector OP will depend on the sensitivity of the observer to different colours. Without some additional information, it is not possible to deduce this from the observations.

From figure 36, we see that the probability of an observer's seeing no difference is the proportion of the elliptical volume bounded by a cylinder of unit radius (A). Similarly the probability of a just noticeable difference being observed is the proportion of the elliptical volume enclosed between the cylinder of unit radius and that of radius 2.



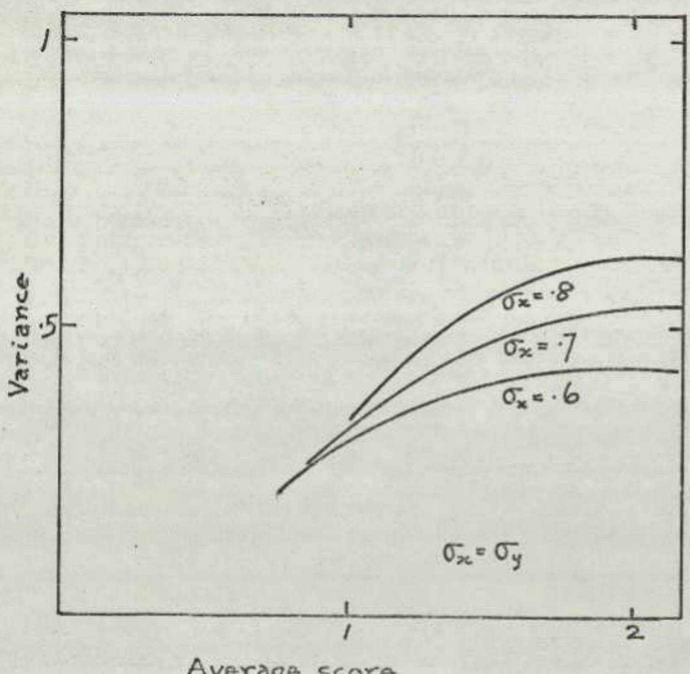


Fig.37.

Calculated relations for bivariate distribution.
$(\sigma_x = \sigma_y)$

(area B in the diagram). Similarly the probability of a noticeable difference is proportional to the volume, region C. The fourth category extends indefinitely.

Using the Program TRYA (see Appendix II), the proportions of scores to be expected in each category, the mean and the variance were calculated for any value of the radius vector OP, for different values of the dispersion in the x direction and different ratios of σ_x to σ_y. By altering the correlation coefficient, the axes of the ellipses could also be rotated.

Calculations were first made for $\sigma_x = \sigma_y$. The relations between the average score to be expected and the mean value of the distribution A were first examined. These are shown in figure 37. Comparison with figure 34 shows that the curves are of a shape similar to those calculated for the univariate distribution. It is also seen that the curves having the same average score at zero mean value for the two distributions, e.g. the univariate curve for $\sigma = 0.9$ and the bivariate curve for $\sigma_x = \sigma_y = 0.6$, have almost identical shape. This means that there is nothing to choose between them if averages only are considered. Thus in Paper 3, the conclusions about the relations between colour shifts and average scores would be unchanged whichever distribution was chosen; only the magnitudes of σ would be altered.

Referring now to the variances, it was stated

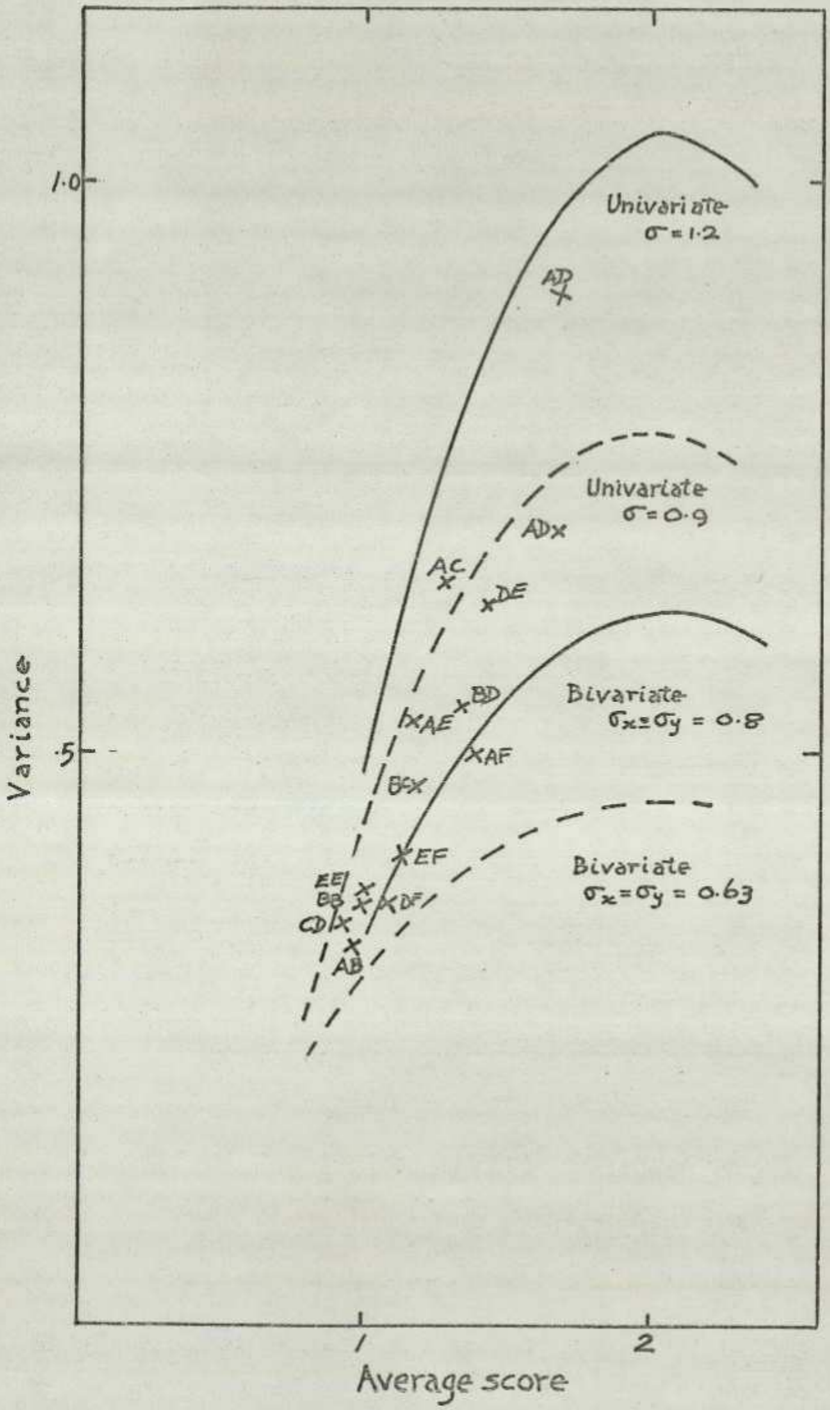


Fig. 38.

Comparison of variances for two distributions.
(Points relate to Test Colour 7)

in that Paper (p114) that reasonable agreement was obtained between the variance obtained for samples 1, 2, 3, and 6, but that less good agreement was obtained for the other test colours. These two types of curve are illustrated in Paper 3 (3) figure 2. It is therefore appropriate to consider the curves of variance as a function of average score for the two types of distribution, since it is only in second order terms that the two distributions will differ. In figure 38, two pairs of curves of variance against average score are plotted. The values of σ have been chosen so that the same average score is obtained for zero mean value (Compare figures 34 and 37). Thus $\sigma = 1.2$ (univariate) and $\sigma_x = \sigma_y = 0.8$ (bivariate) yield the same average score of 1.0 when the mean value is zero. Likewise $\sigma = 0.9$ and $\sigma_x = \sigma_y = 0.63$ yield the same average scores. Thus comparing the variance for $\sigma = 1.2$ with that for $\sigma_x = \sigma_y = 0.8$, it is seen that the variance for the bivariate distribution is always less than that for the univariate and in particular the variance for zero mean value is lower.

Also plotted on figure 38 are experimental points of the variances obtained for test colour 7. The estimate of σ for the univariate distribution had been 1.25, based on the best fit to the average mean score curves. Comparing the variances for $\sigma = 1.2$ (univariate) with the corresponding values for the bivariate distribution, we see that the points are scattered about a line intermediate between the two. This suggests that a bivariate distribution having lower dispersion in the y direction would possibly fit better. It seems remarkable that such good agreement is obtained when it is remembered that σ involves the squares of the <u>deviations</u> from the average scores.

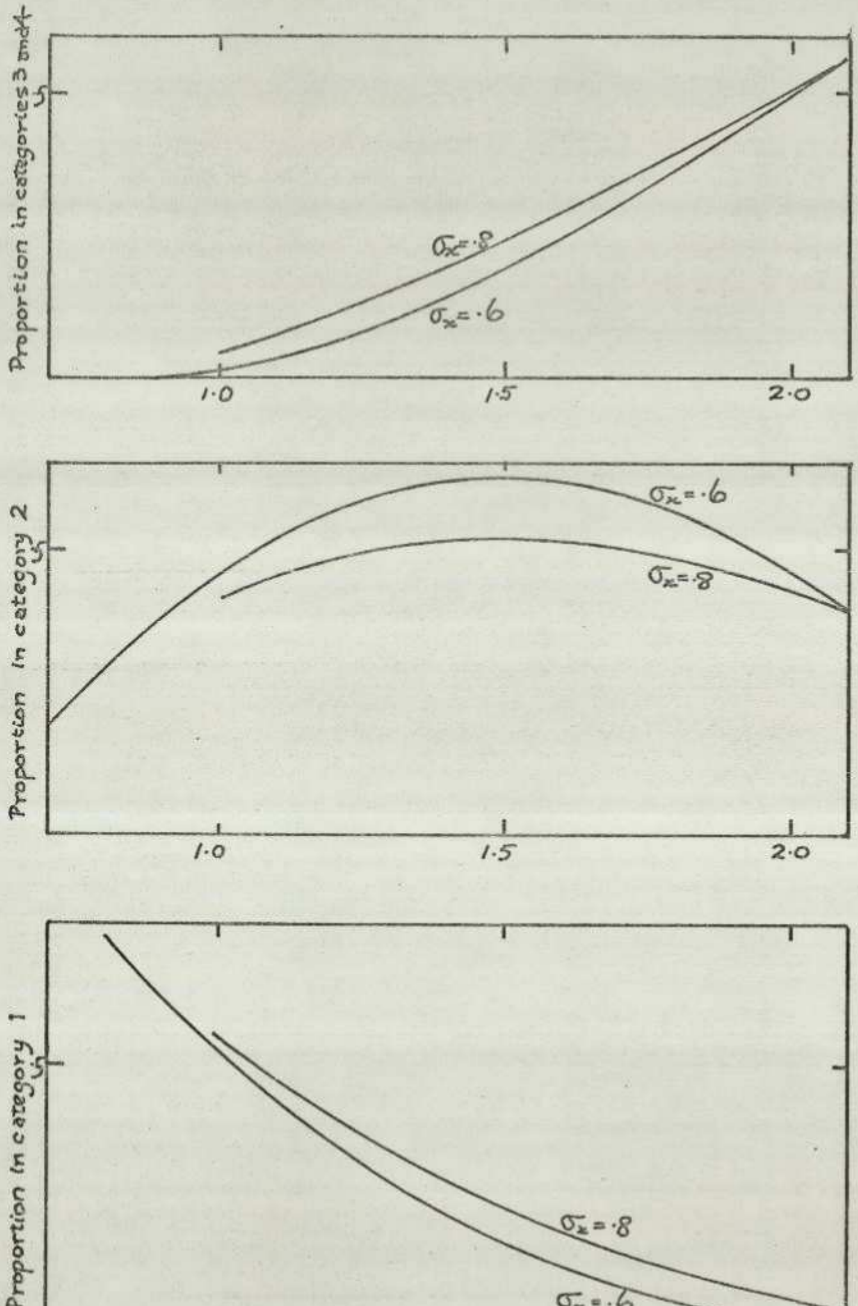


Fig. 39.

Proportions of scores in each category.
(Bivariate distribution)

Facing p.97

6.5 χ^2 Analysis of Distributions.

The proportions of scores to be expected in each category were next examined. These are shown in figure 39 for the bivariate distribution. The figure may be compared with figure 35 for the univariate distribution. It is seen that, for each distribution, the proportion in the second category rises to a maximum when the average is about 1.5, but that its value is higher for the bivariate distribution. This difference between distributions was used as a criterion in considering whether the univariate or bivariate model was the more appropriate.

To carry out the χ^2 analysis, it was necessary to calculate the expected value of scores in each category for each distribution. The process was as follows. For a given colour sample, the average score(Paper 3 (3) table 1 p.113) for the comparison of similar lamps was used to establish a corresponding value of σ. As the case when $\sigma_x = \sigma_y$ was chosen, only one value is required. For each lamp comparison, the proportions expected in categories 1, 2 and (3 & 4 combined) were calculated by examining the tables (or using graphs Fig. 35 and Fig. 39). Only in the case of lamp combination AD, the extreme pair, was it necessary or useful to separate categories 3 and 4. From the proportions, the expected number in each category could be calculated for 48 or 36 observers.

Table I

Experimental results.							Univariate						Bivariate					
Comparison	No. of Obs.	Average	Variance	Cat. 1.	Cat. 2.	Cat. (3 & 4)	Variance	Cat. 1.	Cat. 2.	Cat. (3 & 4)	χ^2	df.	Variance	Cat. 1.	Cat. 2.	Cat. (3 & 4)	χ^2	df.
BD	48	1.33	.54	18	21	9	.77	22	16	10	2.4	1	.49	16	24	8	.7	1
								+	–	+				–	+	–		
AB	48	.96	.32	28	20		.49	29	19		.1		.34	26	22		.3	
								+	–					–	+			
AC	36	1.22	.65	17	13	6	.70	18	12	6	.1	1	.37	15	17	4	2.2	1
								+	–	0				–	0	+		
AD (Expt 2)	36	1.69	.69	8	16	12	.99	11	12	13	2.2	1	.59	6	18	12	.5	1
								+	–	+				–	+	0		
BC	48	1.17	.47	22	20	6	.64	23	16	9	2.0	1	.42	21	22	5	.4	1
								+	–	+				–	+	–		
CD	36	.92	.35	23	13		Average is below threshold											
AD (Expt 4)	36	1.69	.90	10	12	14	.99	11	12	13	.2	1	.59	6	18	12	5.0	1
								+	0	–				–	+	–		
AE	48	1.17	.53	22	21	5	.66	24	16	8	2.8	1	.42	20	23	5	.4	1
								+	–	+				–	+	–		
AF	36	1.39	.50	12	16	8	.84	15	12	9	2.0	1	.60	11	18	7	.4	1
								+	–	+				–	+	–		
DE	48	1.42	.63	16	21	11	1.0	20	16	12	2.4	1	.52	14	24	10	.8	1
								+	–	+				–	+	–		
DF	36	1.08	.37	16	20		.57	20	14		1.4		.38	17	19		.1	
								+	–					+	–			
EF	48	1.13	.41	22	22	4	.62	26	15	7	5.2	1	.46	22	22	4	0	1
								+	–	+				0	0	0		
EE	36	.94	.33	20	16		.49	22	14		.5		.34	19	17		.1	
								+	–					–	+			
BB	36	.94	.33	21	15		.49	22	14		.1		.34	19	17		.4	
								+	–					–	+			

Comparison of distributions using χ^2 test.

Colour 7.

The actual number in each category was given by the original data. The value of χ^2 could then be calculated, given the number of degrees of freedom. Here a difficulty was encountered. The appropriate value of σ to assume had been obtained from the average value of the scores when two similar lamps were compared. Thus one degree of freedom was removed. However, when the number of scores in each category was calculated, the expected value was frequently less than 5 in one category. It is usually considered that for a number less than 5, the test cannot be applied and those scores must be included with those of a neighbouring category. This was in fact done. For example, the scores in category 4 were added to those of category 3 (with one exception). Also for small values of σ, the scores in categories 3 & 4 combined were frequently added to the scores of category 2. However, in calculating the average score for zero signal, all categories had been used. It is thus seen that for a given distribution of scores between say category 1 and categories 2, 3 & 4 combined there could be different values of average score and hence a whole degree of freedom could not have been absorbed. However, in the analysis that follows, it is assumed that a whole degree of freedom had been lost. Thus when more than 5 scores occurred in each of three categories, one degree of freedom was assumed. χ^2 was calculated on this basis.

Table II

Experimental results.							Univariate						Bivariate					
Comparison	No of Obs.	Average	Variance	Cat. 1.	Cat. 2.	Cat. (3 & 4)	Variance	Cat.1.	Cat. 2.	Cat. (3 & 4)	χ^2	df.	Variance	Cat.1.	Cat.2.	Cat. (3 & 4)	χ^2	df.
BD	48	1.92	.83	9	15	24	.92	9	17	22	.4	1	.54	5	22	21	5.9	1
								0	+	−				−	+	−		
AB	48	1.00	.29	26	22		.46	29	19		.8		.34	26	22		0	
								+	−					0	0			
AC	36	1.60	.96	11	14	11	.84	11	13	12	.1	1	.52	7	20	9	4.5	1
								0	−	+				−	+	−		
AD (Expt 2)	36	2.47	.42	*8	21	7	.84	11	13	12	7.7	1	.53	9	18	9	1.1	1
								+	−	+				+	−	+		
BC	48	1.50	.63	13	24	11	.84	16	18	14	3.2	1	.50	11	26	11	.5	1
								+	−	+				−	+	0		
CD	36	1.17	.44	16	20		.58	18	18		.4		.38	15	21		.1	
								+	−					−	+			
AD (Expt 4)	36	2.36	.69*	14	10	12	.87	13	13	10	1.2	1	.58	11	18	7	8.0	1
								−	+	−				−	+	−		
AE	48	1.23	.50	19	24	5	.64	23	17	8	4.7	1	.45	18	25	5	.1	1
								+	−	+				−	+	0		
AF	36	1.97	.65	6	12	18	.92	6	12	18	0	1	.54	3	16	17	4.1	1
								0	0	0				−	+	−		
DE	48	1.81	.83	9	20	19	.90	11	17	20	.9	1	.56	6	23	19	1.9	1
								+	−	+				−	+	0		
DF	36	1.22	.38	13	20	3	.64	17	13	6	6.2	1	.42	14	18	4	.5	1
								+	−	+				+	−	+		
EF	48	1.35	.68	18	21	9	.73	20	18	10	.8	1	.52	15	26	7	2.1	1
								+	−	+				−	+	−		
EE	36	.89	.36	23	13		.35	24	12		.1		.28	23	13		0	
								+	−					0	0			
BB	36	.94	.33	21	15		.44	23	13		.5		.30	21	15		0	
								+	−					0	0			

Comparison of distributions using χ^2 test.

Colour 8.

*Scores given in categories 2, 3 and 4.

Two hypotheses were tested, viz (1) that the distributions were consistent with having been derived from a univariate distribution and (2) that they were consistent with having been derived from a bivariate one. The results are shown in Tables 1 and 2 for test colours 7 and 8 respectively. For each lamp combination, the variance and the scores in each category are shown. Where there were less than 5 in category (3 + 4), scores in that category were added to those of category 2. This removed a further degree of freedom and χ^2 could no longer be estimated. The predicted variance, the scores in each category, the values of χ^2 and the number of degrees of freedom are shown for univariate and bivariate distributions.

It is seen that the value of χ^2 varies from zero for AF (colour 8) to 8 for AD (colour 8 (Expt 4)). Since there is only one degree of freedom, neither of these values is by itself significant. Assuming that a single colour sample has a constant value of σ and always the same type of distribution, we can add the values of χ^2 and obtain combined values as follows

	Colour 7		Colour 8	
	Univariate	Bivariate	Univariate	Bivariate
χ^2	19.7	10.4	25.2	28.7
df	9	9	10	10

For colour 7, the result is inconclusive, but for colour 8, the values would lead us to reject both hypotheses. For neither colour is there a strong preference for one type of distribution or the other. However, when we examine the results for colour 8, we see that major contributions were made by the two AD results. In one case they strongly favoured the univariate distribution and in the other they strongly favoured the bivariate one. These results are also both unusual because they included the fourth category. Comparing the two AD values for colour 7, we see the same pattern. In Experiment 2, the bivariate distribution is favoured, whilst in Experiment 4, the univariate distribution is favoured. No explanation is put forward for this result. It can only be concluded that there was something in the different arrangement of these two experiments which affected the observers when large colour differences were presented. If these two lines are excluded, we find that the results for colour 8 are as follows

Colour 8 for 8 df

	Univariate	Bivariate
χ^2	16.3	19.6

This result is similar to that obtained for colour 7.

The conclusion that neither hypothesis is preferable is surprising in view of the form of the variance curves. It may well be because the averages included the contribution from the extra category and consequently more than one degree of freedom existed. However, even with twenty degrees of freedom, neither hypothesis is consistently favoured.

A possible conclusion is that the type of distribution required, varies depending on the lamp combination. This seems the most likely explanation on the basis of the form of the colour ellipses shown in the paper.

To give an indication of the preferred type of distribution, a plus sign or a minus sign has been placed under the corresponding predicted value for each type of distribution to indicate whether the predicted value is greater or less than that actually obtained. It is noticed that the pattern is frequently +,-,+ for the univariate distribution and -,+,- for the bivariate one, indicating that a distribution intermediate between linear and circular would give a better fit.

To make further progress in this direction, it would be necessary to recalculate scores and averages on the basis of fewer categories so that the number of degrees of freedom was correct and consistent; also to do this for several distributions. From the levels of significance obtained above, it seems unlikely that any stronger conclusions could be drawn using the data available. Clearly many more observations would be needed. It is thus concluded that either hypothesis is acceptable, but neither has a high level of significance.

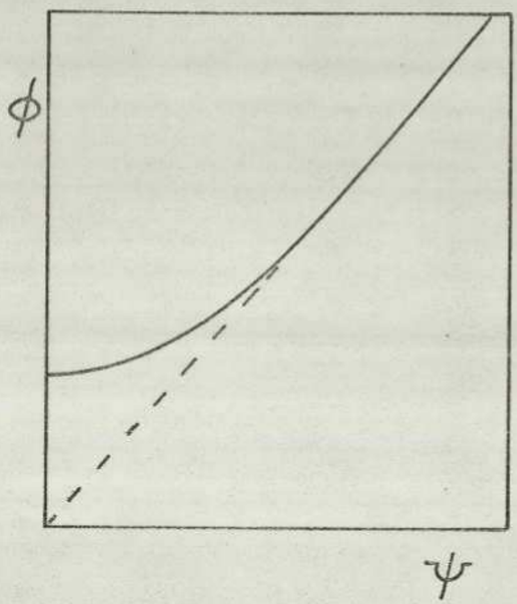


Fig. 40.

Mathematical model for detection close to threshold.

7. FURTHER APPLICATIONS OF THE MODEL.

The model which has been used in the colour problem is clearly applicable to any problem in which observers are asked to detect a difference between a sample under test and a standard, but are not asked to state the direction in which the difference lies. This type of problem is not uncommon. Observers are frequently in agreement that two samples or situations differ, but would be in acute disagreement on how they differed. The sense of taste would be such a continuum for which we have few words to communicate our sensations. The use of the model enables experiments to be designed in which differences and the degree of difference are studied even though the observer could not analyse the differences. The fact that the shape of characteristic is unchanged for a two dimensional continuum means that it would not be necessary to know the dimensionality before the experiment was made.

In addition to the above type of experiment, it is suggested that the model is applicable in the investigation of the limits of human or animal sensitivity i.e. the problem of the absolute limen. Poulton (14) lists the following sensations which have all yielded curves of the "threshold" type, (figure 40), tactile vibration on the arm, subjective warmth and cold, loudness, subjective brightness, taste of salt, detection of shades of grey. He states that Stevens had used a power law of the type

$$\psi = k(\phi - \phi_0)^n$$

where ψ is the subjective magnitude, ϕ is the stimulus magnitude and ϕ_0 is the threshold magnitude, to fit the experimental curves. He says that ϕ_0 can serve two functions. Firstly as a mathematical device to convert the curved part of the relation on a log-log plot into a straight line and secondly to allow for adaptation. ϕ_0 may be varied to account for the adaptation level of the sensor under test. In fitting the equation to data, selection of k arranges that the middle of the curve should fit the data, the value of n makes the upper part agree within experimental limits and the value of ϕ_0 can then be selected to make the curve agree at the lower end. Poulton remarks that if the expression with its three arbitrary constants is to be preferred to any other formulae, it must be shown to fit the data better than any other function having three constants e.g. a quadratic polynomial, and concludes with the remark that even two arbitrary constants can give the curve fitter considerable scope. With the quantity of data usually available, it is seldom possible to advocate one particular function in preference to another.

But even if an excellent fit has been obtained, one is still left with the question:- "What is the threshold and how does it arise?" The formula does no more than postulate a threshold. It is thus a mathematical expression rather than a mathematical model.

In all physical measuring instruments, the limit of sensitivity is set by a statistical process. It may be a statistical process in the instrument itself arising for example out of Brownian motion or random motion of charge carriers,it may arise out of a statistical process in the observer when he makes decisions about the coincidence of pointers. These are factors which limit the precision of physical measurement. It seems reasonable to assume that statistical processes limit the sensitivity of psychophysical measurements. Furthermore all information on the biological processes involved in producing an observation from a stimulus indicate that these are of a statistical nature. It is therefore natural to seek the origin of the threshold in a random process. A mathematical formula does not become a model unless a factor in a physical process can be associated with each arbitrary constant in the formula.

The signal detection theory (section 5) would at first sight seem to offer the possibility of a satisfactory explanation. However, because it is assumed that the observer understands the statistics of the problem in order to make his decision on a maximum liklihood basis, his answer becomes fully determined and ceases to be statistical. Thus, if he has to decide whether a grey is darker or lighter than a standard, he has only to compare the grey with his criterion and his answer must follow. As pointed out in section 5.2, the model reduces to Thurstone's model if the fluctuations are assumed to be Gaussian. There will thus be no threshold in the comparison, provided that greys both darker and lighter are available. Close to the standard, the observer may often say that he can see no difference or he may see a difference in the wrong sense. However, since from a large number of observations, the probability of the grey being lighter or darker can be determined, a curve crossing the 50% probability line approximately where the two greys are equal will be obtained. There may be some displacement from the equality point (see Section 4.2.3, Time-order effect) but the effect will be small.

If however the observer is presented with a black standard and is told that there are no greys darker than the standard, then he can only express differences in a positive sense. It will now be found that there is a "threshold", since if a "mistake" is made, it must always be in a positive direction. By a "mistake" is meant that the observer sees a difference when none exists. Let us consider possible physical processes that could lead to this. If we assume that the eye is dark adapted, there will be a background of pulses transmitted from the nerve cells. If integration is carried out over a long period, it will be found that that the average frequency of the pulses is constant. But it is quite possible for a high proportion of cells to transmit at the "same time", i.e. within the integrating time of the decision system. If that time corresponds to the time when the observer is asked whether a signal is present, he will say "Yes, there is a signal", because the momentary level of pulses differs from the background average by a significant amount. Equally he could be asked the question when a high proportion of cells were quiescent. In that case the momentary background would be lower than normal. It is now assumed that he can detect this as a difference before he can detect whether the number of pulses is greater or less than the average. He will therefore say as before "Yes, there is a signal" without sensing that the difference is in the opposite direction. Indeed because he knows that differences can only occur in one direction, he will think that it is positive.

Such a process, although speculative, is not inconsistent with what is known of animal mechanisms. A similar process might apply to sensors other than the eye. If such a process does exist, then a model for it is the one discussed in previous sections and in Paper 3. The discriminal dispersion represents the variation of the rate of transmission of background pulses relative to the mean value. This will depend on the sensors themselves. Their sensitivity will depend on the signal which is to be observed.

It is therefore suggested that functions of the type discussed in Paper 3 and in this thesis form appropriate mathematical models to represent threshold phenomena. In particular, the normalized average score A for a large number of categories will be given by

$$A = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{M^2}{2}\right) + M(2P(M) - 1)$$

where A = average/σ, and

M = mean value/σ, the normalized value of the stimulus mapped on the psychological continuum. $P(M)$ is the probability integral for a mean value M. This function is shown in figure 6 of Paper 3 (2). It will be seen that this curve is of the threshold type, similar in form to figure 40. It has zero slope when $M = 0$; the slope then increases until the curve becomes linear when M is large. It will in particular be noted that when M is zero, i.e. when the signal is zero, the average normalized score is $\sqrt{2/\pi} = 0.8$. This means that the expected score is 0.8σ. For example, if the dispersion was found to be equal to two category widths, the average score for zero signal would be expected to be equal to 1.6. This analysis assumes an infinite number of categories, but the numerical analysis using the program PROC shows that the curve near the zero is little changed when few categories are used. The limitation of the categories at the top end has of course a very considerable effect and completely alters the shape of the curve.

It should be noted that the above results are independent of the sensitivity of the observer. It would be interesting if we could also evaluate the probability that the signal would be detected, correctly or incorrectly. i.e. the probability of a positive response. Such a relation would give a convenient relation to convert score ratios to mean value of the stimulus on the continuum. Unfortunately this is not possible without knowledge of the width of the categories. In Paper 3, the limit of the first category was arranged to be the just-noticeable difference and the category widths were assumed to be constant. The probability that the observer would say that there was a difference when none was presented was then shown to be (Paper 3(2) Fig.10) 40% for $\sigma = 1.2$. It is thus seen that the frequently quoted criterion for a just-noticeable difference of 50% of positive scores is quite unsatisfactory when large values of discriminal dispersion exist. These high values of discriminal dispersion are frequently found in psychophysical studies.

8. SUGGESTIONS FOR FURTHER WORK.

The experiments of Paper 3 were carried out for a specific purpose, namely the investigation of colour rendering. Out of that work arose the model discussed above. It is recognized, however, that the experiments were far from ideal for investigating models and their performance.

Further investigations are required along two lines. The first is an investigation of observer's ability to detect differences before being able to detect a sense of direction. An experiment would need to be devised aiming first at a purely one-dimensional continuum. As an example the difference of lightness of greys from a standard might be examined. The continuum of lightness should be one-dimensional and experiments could be made with several values of grey as standards. Other investigations might be on the intensity of acoustic noise or on differences of pitch.

If the model survived the first experiment, it could be tried for two dimensions e.g. by introducing controlled variations of colour along various axes of the u-v diagram or in the case of the acoustic experiment by varying the quality of the noise by a controlled filtering or synthesizing process.

The second study might be the investigation of the application of the same model to the absolute threshold. As was suggested in the last section, this might be applied to a comparison of black with greys. Equally it could be applied to the detection of acoustic signals.

However, each of these studies involves a great deal of labour in assembling observers, in verifying that the conditions are as they are supposed to be, in avoiding mislesding effects due either to the arrangement or due to the way in which the experiment was administered. No doubt much of the necessary data exist in the logbooks of the laboratories where such work has been in progress for many years. It is a pity that it is seldom possible from published papers to find sufficient detail of the exact arrangement of an experiment to know whether the data would be of interest. It would be most desirable if experimenters could devise a standard format for publication of results. It is envisaged that full details of the experimental arrangement, of the conditions for each trial, of the skill age, motivation and characteristics of each observer and so on could be documented in a standard form. The full detail of every observation would also need to be recorded. It would then be possible for workers to compare the work of others with their own, and would enable the same data to be examined in the light of different theories.

9. CONCLUSIONS.

All analysis of subjective data implies the existence of a model. Before analysis starts it is essential to answer certain questions about the continuum on which stimuli exist:- How many dimensions does it have? Is it likely to be the same over the whole range of investigation, or are there likely to be different regimes in which sensitivity will differ? Does it have a zero in the range? What are the sources of variability? Are they likely to introduce normal or log-normal distributions? What is the size of the variability compared with the scale values? What side effects are being introduced by the experimental procedure? If these cannot be eliminated, are they being balanced out?

To answer these questions, much must be known about the stimulus continuum which is being measured. For physically well defined stimuli, such as sensation of warmth, brightness and so on, it is fairly straightforward to analyse them and to devise appropriate models.

However, for purely qualitative stimuli, such as pleasantness and colour preference, the problem is fraught with so many difficulties that the Author doubts the validity of any of the methods discussed for producing a result with a consistent meaning. Nevertheless, these are the sort of continua that frequently interest an environmental engineer.

This is illustrated by the problem of the colour rendering index. An attempt had been made to provide a figure of merit for lamps, based on the way in which they changed the appearance of eight test colours. The result was a single number which could be arrived at in an infinite number of ways. Thus the same colour rendering index could be obtained by a lamp which would give a perfectly acceptable colour rendering of a general scene and by another which would give a completely unacceptable one. Clearly the index is not a unique function of the subjective parameter "colour rendering".

A similar problem is faced by the environmental engineer in his attempts to apply a cost effectiveness analysis to an environmental scheme. It is easy enough to calculate costs of purely physical changes; but where more subjective values are concerned, such as the destruction of an ancient church or of a beautiful landscape, money values are no longer applicable since they are no longer unique functions of the operation proposed. Alternatively we may say that they mean such different things to different observers, one may be strongly in favour whilst another may be violently opposed, that there is no meaning in the expression "average observer".

To apply any of the models discussed in this thesis, there must be a monotonic one-valued relationship between the stimulus and the response, and this relationship must not be violently affected by other factors outside the control of the experimenter.

Papers included in this thesis

1. Mainwaring, G. and Stainsby, A. G., The Effect of Lantern Light Distribution on Subjective Assessments in a Street Lighting Installation. Trans. Illum. Eng. Soc. (London) 33 98-116(1968)

2. Halstead, M. B., Morley, D, I,, Palmer, D. A. and Stainsby, A.G. Colour Rendering Tolerances. Proc. A.I.C. Congress 'Color 69', 2 839 (1969)

3. Halstead, M. B., Morley, D. I., Palmer, D. A. and Stainsby, A. G. Colour Rendering Tolerances in the C.I.E. System. Lighting Research and Technology, 3, 99 (1971)

4. Halstead, M. B., Morley, D. I., Stainsby, A. G. and Palmer, D. A. Determination of Tolerances for Colour Rendering Indices. C. I. E. Proceedings, XVII session - Barcelona P - 71.19 (1971)

5. Morley, D. I., Halstead, M. B., Palmer, D. A., and Stainsby, A. G. Subjective Assessments of Colour-rendering Indices. Proc. A. I. C. Congress 'Color 73' 359 (1973)

List of References

1. Aitchison, J. and Brown, J.A.C. The Lognormal Distribution. Cambridge University Press, (1957) Chap.2.

2. Bernoulli, D. Exposition of a New Theory on the Measurement of Risk. 1738. Translation in Econometrica 22, 22-35, 1954.

3. Commission Internationale d'Eclairage. Publication no. 13.

4. Coombs, C.H. Psychological scaling without a unit of Measurement. Psychological Review 57, 145, 1950.

5. Crawford, B.H. The Colour Rendering Properties of Illuminants. Br. J. Appl. Phys. 14, 319, 1963.

6. Cramer, G. (1728) letter (1732) to Nicholas Bernoulli cousin of Daniel Bernoulli. Quoted in note appended to Bernoulli(2)

7. Eijkman, E. and Vendrik, A.J.H. Detection theory applied to the Absolute Sensitivity of Sensory Systems. Included in Swets(25) Chap. 18, p.392.

8. Fitzhugh, R. The Statistical Detection of Threshold Signals in the Retina. Included in Swets(25) Chap. 17, p366.

9. Guilford, J.P. Psychometric Methods. McGraw Hill, New York. 1954.

10. Lieberman, B. (Editor) Contemporary Problems in Statistics. Oxford University Press, 1971.

11. Merkel, J. Die Abhängigheit Zwischen Reiz und Empfindung. Philosophische Studien 4, 541-594, 1888.

12. Miller, G.A. Sensitivity Changes in the intensity of white noise and its relation to masking and loudness. J. Acoust. Soc. Amer. 19, 609-619, 1947.

13. Plateau, J.A.F. Sur le mesure des sensations physiques et sur la loi qui lie l'intensite de ces sensations a l'intensite de la cause excitante. Bulletin de l'Academie Royale de Belgique. 33 (ser2) 376-388, 1872.

14. Poulton, E.C. The new psychophysics: six models for magnitude estimation. Psychological Bulletin 69, 1 - 19. 1968.

15. Stevens, J.C., Mack, J.D. and Stevens, S.S. Growth of sensation on seven continua as measured by force of handgrip. Jour. Exper. Psychology. 59, 60-67, 1960.

16. Stevens, J.C. and Guirao, M. Individual loudness functions. Jour Acoust. Soc. Amer. 36. 2210-2213, 1964.

17. Stevens, S.S. On the Theory of Scales of measurement. Science 103, 677-680, 1946. (This and many other papers on the same topic are collected in Ref. 10.)

18. Stevens, S.S. (Editor) Handbook of Experimental Psychology. pp 1-49. John Wiley, New York.(1951)

19. Stevens, S.S. On the brightness of lights and the loudness of sounds. Science 118. 576, 1953. (Abstract)

20. Stevens, S.S. The measurement of Loudness. Jour Acoust. Soc. Amer. 27, 815-820, 1955.

21. Stevens, S.S. On the psychophysical law. Psychophysical Review. 64, 153-181, 1957.

22. Stevens, S.S. Measurements, Statistics and the Schemapiric View. Science 161, 849-856 1968.

23. Stevens, S.S. Issues in Psychophysical Measurement. Psychological Review. 78, 426-450, 1971.

24. Stevens, S.S. Psychophysics and social scaling. General Learning Press, Morristown, N.J. 1972.

25. Swets, J.A. Signal detection and recognition by human observers. John Wiley, New York. 1964.

26. Thurstone L.L. A law of comparative judgement Psychological Review 34 273-286, 1927. See also Torgerson (Ref 27) for extended discussion.

27. Torgerson, W.S. Theory and Methods of Scaling John Wiley, New York 1958.

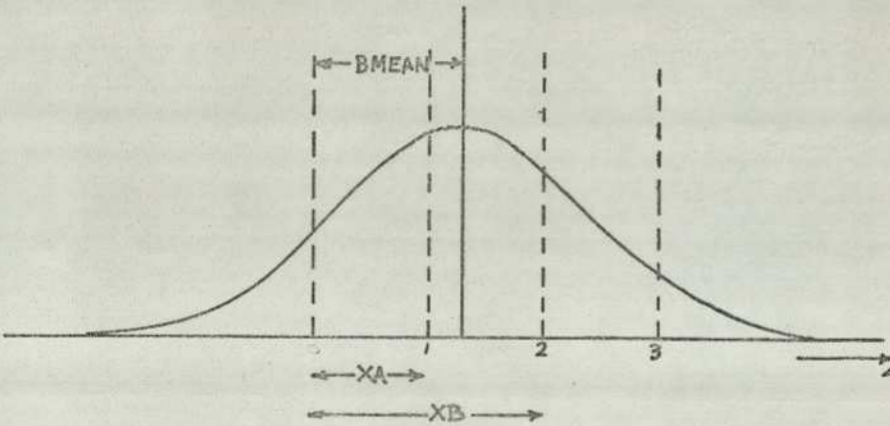


Fig. A1. 1.

Facing p.119

APPENDIX I

Program PROC (Fortran IV)

The program calculates the average score to be expected if a univariate normal distribution is folded about the zero line and the results categorized in up to 7 categories of equal width. Various other functions of the scores are tabulated.

Description

It is assumed that observers are presented with a signal BMEAN on the category scale, i.e. one unit is the just noticeable difference, two units, the noticeable one and so on. The observations are assumed to be spread normally on the category scale with a dispersion SIGMA. The program calculates the proportion of the probability integral, folded about zero, that lies between two category boundaries.

Referring to figure A.1.1, the positions of the category boundaries are put equal to $1/\sigma$, $2/\sigma$, and so on. Then for a distribution having a given value of BMEAN the proportions of the probability integrals lying between 0 and 1 (XA) and between 1 and 2 (XB) are calculated. These are subtracted to give the expected scores in categories 1 and 2. The process is continued for the other categories. The integration uses an I.C.L. library subroutine F4ERFN which supplies the error function.

From the proportions in each category (note that the last category always has a limit of infinity) the following output is calculated and tabulated.

Output

AVER the average response on the category scale.

VAR the variance of the response on the category scale.

RTVAR the square root of VAR (not used)

SLOPE an approximate value of the slope of the curve of AVER as a function of BMEAN (not used, intended for weighting)

PF1, etc. the proportions of scores in each category up to a maximum of seven.

Operation

To run the program, the following cards are required. The input data is tabulated immediately in front of the output.

Card 1 NCAT (I1) The integer number of categories. (not more than 7)

Card 2 MINB, MAXB, INTB (3I1)
MINB is the lowest integer part of the mean value. MAXB is the highest integer part of the mean value. INTB is the interval between successive decimal parts of B in the table. Thus INTB can take the values 5, 2 or 1 depending on the detail required.

Card 3 SIGMIN, SIGMAX, SIGINT (3I3)
SIGMIN is 100 times the maximum value of σ
SIGINT is 100 times the interval between successive values of σ (thus for the table to cover the range $\sigma = 0.90$ to 1.10 in steps of 0.02, SIGMIN = 90, SIGMAX = 110, SIGINT = 2.)

```
      MASTER PROC
      DIMENSION PF(7),PX(7),PO(7),TF(7)
      INTEGER SIGMIN,SIGMAX,SIGINT
      READ(1,300)NCAT
      READ(1,301)MINB,MAXB,INTB
      READ(1,302)SIGMIN,SIGMAX,SIGINT
      WRITE(2,305)
      WRITE(2,306)NCAT,MINB,MAXB,INTB,SIGMIN,SIGMAX,SIGINT
      WRITE(2,215)
C          NCAT IS NUMBER OF CATEGORIES , ONE CARD
C          MINB,MAXB ARE INTEGER PARTS OF MEAN VALUE
C          INTB IS INTERVAL OF B IN TENTHS
C         SIGS ARE ENTERED AT 100 TIMES THEIR VALUE IF FOR 0.9 PUT 90
      DO 405 K1=SIGMIN,SIGMAX,SIGINT
      SIGMA=0.01*FLOAT(K1)
      NA=1
      IF(INTB-2)10,11,12
   10 MTOP=9
      GO TO 15
   11 MTOP=8
      GO TO 15
   12 MTOP=5
   15 DO 400 L=MINB,MAXB
      DO 395 M=0,MTOP,INTB
      BMEAN=0.1*FLOAT(10*L+M)
      PXA=0
      AV=0
      AMEAN=BMEAN/SIGMA
C          AMEAN IS LOCATION OF BOUNDARY ON NORMAL DEVIATE SCALE
      DO 100 J=1,NCAT
      IF(J.LT.NCAT)GO TO 40
      PXB=1.0
      GO TO 90
   40 X=FLOAT(J)/SIGMA
C          X IS LOCATIOH OF BOUNDARY ON NORMAL DEVIATE SCALE
      XPOS=AMEAN+X
      XNEG=AMEAN-X
      XNAB=ABS(XNEG)
      I=1
C          CONVERT X TO Y FOR ERROR FUNCTION LIBRARY SUBROUTINE F4ERFN
      Y=XPOS*0.70711
   50 CALL F4ERFN (Y,FRE,FREC)
C          FRE IS ERROR FUNVTION, FREC IS COMPLEMENT (NOT USED)
      IF(I.EQ.2)GOTO 60
      PXPOS=0.5*(1.0+FRE)
C          CONVERTS ERROR FUNCTION TO NORMAL PROBABILITY INTEGRAL
      PO(J)=PXPOS
      Y=XNAB*0.70711
      I=2
      GO TO 50
   60 PXNAB=0.5*(1.0+FRE)
      IF(XNEG)80,70,70
   70 PXB=PXPOS-PXNAB
C          70 AND80 GIVE AREA BETWEEN ZERO AND BOUNDARY
```

```
      GO TO 90
   80 PXB=PXPOS+PXNAB-1
   90 PX(J)=PXB
      PF(J)=PXB-PXA
      PXA=PXB
      AV=FLOAT(J)*PF(J)+AV
  100 CONTINUE
      AVER=AV-0.5
      VAR=0
      SUMP=0
      DO 200 J=1,NCAT
      VAR=PF(J)*(FLOAT(J)-AV)**2+VAR
      SUMP=PF(J)+SUMP
  200 CONTINUE
      RTVAR=SQRT(VAR)
      GOTO(220,221,222),NA
  220 AV2=AVER
      GOTO250
  221 AV3=AVER
      SLOPE=0
      GOTO225
  222 AV1=AV2
      AV2=AV3
      AV3=AVER
      SLOPE=5.0*(AV3-AV1)/FLOAT(INTB)
  225 WRITE(2,210)BM1,AV2,SG1,RTV1,VR1,SLOPE,TF
      IF(NA.EQ.4)GOTO 395
  250 BM1=BMEAN
      SG1=SIGMA
      RTV1=RTVAR
      VR1=VAR
      DO 252 IA=1,7
  252 TF(IA)=PF(IA)
      IF(NA.LT.3)NA=NA+1
      IF(L.EQ.MAXB.AND.M.EQ.MTOP)NA=4
      IF(NA.NE.4)GOTO 395
      SLOPE=10.*(AV3-AV2)/FLOAT(INTB)
      GOTO 225
  395 CONTINUE
  400 CONTINUE
      WRITE(2,211)
  405 CONTINUE
  210 FORMAT(F6.1,F9.4,F6.2,F7.3,9F8.3)
  211 FORMAT(1H0)
  215 FORMAT(2X,5HBMEAN,3X,4HAVER,2X,5HSIGMA,2X,5HRTVAR,4X,3HVAR,4X,
     15HSLOPE,4X,3HPF1,5X,3HPF2,5X,3HPF3,5X,3HPF4,5X,3HPF5,5X,3HPF6,5X,
     13HPF7)
  300 FORMAT(I1)
  301 FORMAT(3I1)
  302 FORMAT(3I3)
  305 FORMAT(3X,4HNCAT,6X,4HMINB,6X,4HMAXB,6X,4HINTB,5X,6HSIGMIN,4X,
     16HSIGMAX,4X,6HSIGINT)
  306 FORMAT(I6,6I10)
      STOP
      END

END OF SEGMENT, LENGTH   573, NAME  PROC
```

DOCUMENT PROC , NORMLPROC ; LP00 ON 22/04/75 AT 20.12

NCAT	MINB	MAXB	INTB	SIGMIN	SIGMAX	SIGINT
4	0	2	1	90	130	5

BMEAN	AVER	SIGMA	RTVAR	VAR	SLOPE	PF1	PF2	PF3	PF4	PF5	PF6	PF7
0.0	0.7936	0.90	0.513	0.263	0.000	0.733	0.240	0.025	0.001	0.000	0.000	0.000
0.1	0.7976	0.90	0.517	0.267	0.079	0.731	0.242	0.026	0.001	0.000	0.000	0.000
0.2	0.8094	0.90	0.527	0.278	0.156	0.722	0.248	0.029	0.001	0.000	0.000	0.000
0.3	0.8289	0.90	0.544	0.296	0.233	0.707	0.258	0.033	0.001	0.000	0.000	0.000
0.4	0.8560	0.90	0.566	0.320	0.307	0.688	0.271	0.040	0.002	0.000	0.000	0.000
0.5	0.8904	0.90	0.592	0.350	0.379	0.663	0.287	0.048	0.003	0.000	0.000	0.000
0.6	0.9318	0.90	0.620	0.384	0.448	0.634	0.304	0.058	0.004	0.000	0.000	0.000
0.7	0.9799	0.90	0.650	0.422	0.512	0.601	0.323	0.070	0.005	0.000	0.000	0.000
0.8	1.0342	0.90	0.680	0.462	0.573	0.565	0.343	0.085	0.007	0.000	0.000	0.000
0.9	1.0944	0.90	0.709	0.503	0.629	0.527	0.362	0.102	0.010	0.000	0.000	0.000
1.0	1.1600	0.90	0.738	0.544	0.680	0.487	0.379	0.121	0.013	0.000	0.000	0.000
1.1	1.2304	0.90	0.764	0.584	0.726	0.446	0.395	0.142	0.017	0.000	0.000	0.000
1.2	1.3052	0.90	0.789	0.622	0.767	0.405	0.408	0.164	0.023	0.000	0.000	0.000
1.3	1.3838	0.90	0.811	0.657	0.803	0.364	0.417	0.189	0.029	0.000	0.000	0.000
1.4	1.4658	0.90	0.830	0.689	0.834	0.325	0.423	0.215	0.038	0.000	0.000	0.000
1.5	1.5506	0.90	0.847	0.717	0.860	0.287	0.424	0.242	0.048	0.000	0.000	0.000
1.6	1.6377	0.90	0.860	0.740	0.881	0.251	0.421	0.268	0.060	0.000	0.000	0.000
1.7	1.7268	0.90	0.871	0.758	0.897	0.217	0.414	0.295	0.074	0.000	0.000	0.000
1.8	1.8172	0.90	0.878	0.771	0.909	0.186	0.402	0.321	0.091	0.000	0.000	0.000
1.9	1.9086	0.90	0.883	0.779	0.916	0.158	0.386	0.345	0.111	0.000	0.000	0.000
2.0	2.0004	0.90	0.884	0.782	0.919	0.133	0.367	0.367	0.133	0.000	0.000	0.000
2.1	2.0924	0.90	0.883	0.780	0.918	0.111	0.345	0.386	0.159	0.000	0.000	0.000
2.2	2.1839	0.90	0.879	0.772	0.912	0.091	0.321	0.401	0.187	0.000	0.000	0.000
2.3	2.2747	0.90	0.872	0.760	0.902	0.074	0.295	0.412	0.218	0.000	0.000	0.000
2.4	2.3643	0.90	0.861	0.742	0.888	0.060	0.269	0.419	0.252	0.000	0.000	0.000
2.5	2.4523	0.90	0.848	0.719	0.869	0.048	0.242	0.421	0.289	0.000	0.000	0.000
2.6	2.5382	0.90	0.832	0.692	0.847	0.038	0.215	0.419	0.328	0.000	0.000	0.000
2.7	2.6217	0.90	0.813	0.661	0.821	0.029	0.189	0.412	0.369	0.000	0.000	0.000
2.8	2.7023	0.90	0.791	0.626	0.790	0.023	0.164	0.401	0.412	0.000	0.000	0.000
2.9	2.7023	0.90	0.767	0.588	0.774	0.017	0.141	0.386	0.456	0.000	0.000	0.000

0.0	0.8294	0.95	0.546	0.298	0.000	0.707	0.257	0.034	0.002	0.000	0.000	0.000
0.1	0.8331	0.95	0.549	0.301	0.075	0.705	0.259	0.035	0.002	0.000	0.000	0.000
0.2	0.8445	0.95	0.559	0.312	0.150	0.697	0.264	0.037	0.002	0.000	0.000	0.000
0.3	0.8632	0.95	0.575	0.330	0.224	0.684	0.272	0.042	0.002	0.000	0.000	0.000
0.4	0.8892	0.95	0.595	0.354	0.295	0.666	0.282	0.049	0.003	0.000	0.000	0.000
0.5	0.9223	0.95	0.620	0.384	0.365	0.643	0.295	0.057	0.004	0.000	0.000	0.000
0.6	0.9621	0.95	0.647	0.419	0.431	0.617	0.310	0.068	0.006	0.000	0.000	0.000
0.7	1.0085	0.95	0.676	0.457	0.494	0.587	0.325	0.080	0.008	0.000	0.000	0.000
0.8	1.0609	0.95	0.705	0.497	0.553	0.554	0.341	0.095	0.010	0.000	0.000	0.000
0.9	1.1190	0.95	0.734	0.539	0.607	0.519	0.356	0.111	0.014	0.000	0.000	0.000
1.0	1.1823	0.95	0.763	0.581	0.658	0.482	0.371	0.129	0.018	0.000	0.000	0.000
1.1	1.2505	0.95	0.789	0.623	0.703	0.445	0.383	0.150	0.023	0.000	0.000	0.000
1.2	1.3230	0.95	0.814	0.662	0.744	0.406	0.393	0.171	0.029	0.000	0.000	0.000
1.3	1.3993	0.95	0.836	0.699	0.780	0.368	0.401	0.194	0.037	0.000	0.000	0.000
1.4	1.4790	0.95	0.856	0.733	0.811	0.331	0.405	0.218	0.046	0.000	0.000	0.000
1.5	1.5615	0.95	0.873	0.762	0.838	0.295	0.405	0.242	0.057	0.000	0.000	0.000
1.6	1.6465	0.95	0.887	0.787	0.859	0.261	0.402	0.267	0.070	0.000	0.000	0.000
1.7	1.7334	0.95	0.898	0.807	0.876	0.228	0.396	0.291	0.086	0.000	0.000	0.000
1.8	1.8217	0.95	0.906	0.821	0.888	0.198	0.385	0.313	0.103	0.000	0.000	0.000
1.9	1.9110	0.95	0.911	0.830	0.896	0.171	0.371	0.335	0.123	0.000	0.000	0.000
2.0	2.0008	0.95	0.913	0.833	0.899	0.145	0.355	0.354	0.146	0.000	0.000	0.000
2.1	2.0907	0.95	0.912	0.831	0.898	0.123	0.335	0.370	0.172	0.000	0.000	0.000
2.2	2.1804	0.95	0.907	0.823	0.892	0.103	0.314	0.384	0.200	0.000	0.000	0.000
2.3	2.2692	0.95	0.900	0.809	0.883	0.085	0.291	0.393	0.231	0.000	0.000	0.000
2.4	2.3569	0.95	0.889	0.791	0.869	0.070	0.267	0.399	0.264	0.000	0.000	0.000
2.5	2.4429	0.95	0.876	0.767	0.851	0.057	0.242	0.401	0.299	0.000	0.000	0.000
2.6	2.5270	0.95	0.859	0.738	0.829	0.046	0.218	0.399	0.337	0.000	0.000	0.000
2.7	2.6088	0.95	0.840	0.705	0.803	0.037	0.194	0.393	0.376	0.000	0.000	0.000
2.8	2.6877	0.95	0.818	0.668	0.774	0.029	0.171	0.384	0.417	0.000	0.000	0.000
2.9	2.6877	0.95	0.793	0.628	0.759	0.023	0.149	0.370	0.458	0.000	0.000	0.000

APPENDIX II

Program TRYA (Fortran IV)

The program calculated the average score to be expected if the fluctuations of observers are distributed according to a bivariate normal distribution. It follows the same method as Program PROC to evaluate the proportion of scores in each category. The program is designed for four categories only.

Description

Referring to fig.36 in section 6.4, BMEAN is the length of the vector OP on the category scale. The observations are assumed to be distributed normally with respect to the dimension x, along the stimulus axis, and with respect to y, normal to it. The dispersions in these two dimensions are σ_x and σ_y, and the correlation coefficient is variable.

The bivariate normal distribution is converted to polar coordinates using a function segment and is integrated using INTEG. INTEG is a subroutine for Gaussian quadrature in two dimensions, and takes the place of the library subroutine used in program PROC. (INTEG was kindly provided by Mr J. Snell of the Mathematics department of the City University.) Because it is an iterative routine, INTEG requires more computer time than was required in PROC.

Output

As for PROC.

In addition PR1, PR2 etc. These are the probabilities of the observations being above a given category boundary (not used).

Input

To run the program, the following cards are required:

Card 1 NCAT, EPS (I1, F8.5)

NCAT a single integer indicating the number of categories (must be 4), EPS a real number indicating the maximum error acceptable after each iteration of INTEG. e.g. 0.1% is entered as 0.001.

Card 2 MINB, MAXB, INTB (3I1)

MINB and MAXB are the integer parts of the mean value. INTB is the interval of BMEAN in tenths. This may be 1, 2 or 5.

Card 3 SIGMIN, SIGMAX, SIGINT (3I3)

These are the minimum, the maximum and the interval of σ_x, and are entered at 100 times their absolute value. Thus for $\sigma_x = 0.9$, write 090 and for an interval of 0.1, write 010.

Card 4 SIGRAT, RHO (2F5.2)

SIGRAT is the ratio σ_y/σ_x, RHO is the correlation coefficient. These are entered at their correct magnitudes.

```
      MASTER TRYA
      DIMENSION PF(8),PR(8),TF(8)
      COMMON BMEAN,SIGX,SIGY,PI,RHO
      INTEGER SIGMIN,SIGMAX,SIGINT
      READ(1,300)NCAT,EPS
      READ(1,301)MINB,MAXB,INTB
      READ(1,302)SIGMIN,SIGMAX,SIGINT
      READ(1,303)SIGRAT,RHO
      WRITE(2,305)
      WRITE(2,306)NCAT,MINB,MAXB,INTB,SIGMIN,SIGMAX,SIGINT,EPS,SIGRAT,RH
     10
      WRITE(2,215)
C           NCAT IS NUMBER OF CATEGORIES , ONE CARD
C           MINB,MAXB ARE INTEGER PARTS OF MEAN VALUE
C           INTB IS INTERVAL OF B IN TENTHS
C       SIGS ARE ENTERED AT 100 TIMES VALUE. THUS FOR 0.9 PUT 90
C          SIGRAT IS RATIO SIGY/SIGX
C          RHO IS CORRELATION COEFF
      PI=4.0*ATAN(1.0)
      DO 405 K1=SIGMIN,SIGMAX,SIGINT
      SIGX=0.01*FLOAT(K1)
      SIGY=SIGRAT*SIGX
      NA=1
      IF(INTB-2)10,11,12
   10 MTOP=9
      GO TO 15
   11 MTOP=8
      GO TO 15
   12 MTOP=5
   15 DO 400 L=MINB,MAXB
      DO 395 M=0,MTOP,INTB
      BMEAN=0.1*FLOAT(10*L+M)
      AV=0
      DO 100 J=1,4
      A=FLOAT(J-1)
      IF(J.EQ.4)GOTO 40
      B=FLOAT(J)
      GOTO 41
   40 B=10
   41 C=0.0
      IF(BMEAN-0.0001)42,42,44
   42 D=0.5*PI
      CALL INTEG(A,B,C,D,VALUE,EPS,N)
      PF(J)=2*VALUE
      GOTO 46
   44 D=PI
      CALL INTEG(A,B,C,D,VALUE,EPS,N)
      PF(J)=VALUE
   46 AV=FLOAT(J)*PF(J)+AV
  100 CONTINUE
      AVER=AV-0.5
      TNCAT=TF(NCAT)
      PRQ=0
      NKIT=NCAT-1
      DO 112 JP=1,NKIT
      PRQ=PRQ+TF(JP)
  112 PR(JP)=1.0-PRQ
      PR(NCAT)=1.0-PRQ-TNCAT
      VAR=0
      DO 200 J=1,NCAT
      VAR=PF(J)*(FLOAT(J)-AV)**2+VAR
  200 CONTINUE
      RTVAR=SQRT(VAR)
      GOTO(220,221,222),NA
```

```
220 AV2=AVER
    GOTO250
221 AV3=AVER
    SLOPE=0
    GOTO225
222 AV1=AV2
    AV2=AV3
    AV3=AVER
    SLOPE=5.0*(AV3-AV1)/FLOAT(INTB)
225 WRITE(2,210)BM1,AV2,SG1,RTV1,VR1,SLOPE,(TF(I),I=1,NCAT),
   1(PR(I),I=1,NCAT)
    IF(NA.EQ.4)GOTO 395
250 BM1=BMEAN
    SG1=SIGX
    RTV1=RTVAR
    VR1=VAR
    DO 252 IA=1,7
252 TF(IA)=PF(IA)
    IF(NA.LT.3)NA=NA+1
    IF(L.EQ.MAXB.AND.M.EQ.MTOP)NA=4
    IF(NA.NE.4)GOTO 395
    SLOPE=10.*(AV3-AV2)/FLOAT(INTB)
    GOTO 225
395 CONTINUE
400 CONTINUE
405 CONTINUE
300 FORMAT(I1,F8.5)
301 FORMAT(3I1)
302 FORMAT(3I3)
303 FORMAT(2F5.2)
305 FORMAT(3X,4HNCAT,6X,4HMINB,6X,4HMAXB,6X,4HINTB,5X,6HSIGMIN,4X,
   16HSIGMAX,4X,6HSIGINT,3X,3HEPS,7X,6HSIGRAT,4X,3HRHO)
306 FORMAT(I6,6I10,F10.5,2F10.3)
215 FORMAT(1X,5HBMEAN,2X,4HAVER,3X,5HSIGMA,2X,5HRTVAR,3X,3HVAR,3X,
   15HSLOPE,6X,3HPF1,4X,3HPF2,4X,3HPF3,4X,3HPF4,6X,3HPR1,4X,3HPR2,4X,
   13HPR3,4X,3HPR4)
210 FORMAT(1H ,F5.2,F8.4,F7.2,3F7.3,F9.3,3F7.3,F9.3,3F7.3)
    STOP
    END

END OF SEGMENT, LENGTH   621, NAME  TRYA
```

TRIBUTION

```
C         FUNCTION SEGMENT TO CONVERT BIVARIATE NORMAL DISTRIBUTION
C         INTO POLAR COORDINATES FOR INTEGTATION
      FUNCTION F(R,TH)
      COMMON BMEAN,SIGX,SIGY,PI,RHO
      X=R*COS(TH)-BMEAN
      Y=R*SIN(TH)
      RO2=1-RHO**2
      SRO2=SQRT(RO2)
      XS=X/SIGX
      YS=Y/SIGY
      OBY2=0.5/RO2*(XS**2-2*RHO*XS*YS+YS**2)
      DIV=PI*SIGX*SIGY*SRO2
      F=R*EXP(-OBY2)/DIV
      RETURN
      END

END OF SEGMENT, LENGTH   113, NAME  F
```

```
      SUBROUTINE INTEG(A,B,C,D,VALUE,EPS,N)
C         SUBROUTINE INTEG EVALUATES DOUBLE INTEGRAL OF FUNCTION (X,Y)
C         BETWEEN LIMITS X=A,B AND Y=C,D GIVING  VALUE (N=NO OF ITERATS
      DIMENSION W(3),X(3)
      W(1)=5.0/9.0
      W(3)=W(1)
      W(2)=8.0/9.0
      X(1)=SQRT(0.6)
      X(3)=-X(1)
      X(2)=0.0
      N=1
      H=B-A
      HH=D-C
      V1=0.0
      DO 1 K=1,3
      DO 1 L=1,3
    1 V1=V1+W(K)*W(L)*F(A+.5*H*(X(K)+1.0),C+.5*HH*(X(L)+1.0))
      V1=0.25*H*HH*V1
    3 N=N+1
      H=(B-A)/N
      HH=(D-C)/N
      V2=0.0
      DO 2 I=1,N
      DO 2 J=1,N
      DO 2 K=1,3
      DO 2 L=1,3
      A1=A+.5*H*(X(K)+1.0)+(I-1)*H
      A2=C+.5*HH*(X(L)+1.0)+(J-1)*HH
    2 V2=V2+W(K)*W(L)*F(A1,A2)
      V2=.25*H*HH*V2
      TEST=ABS(V2-V1)
      V1=V2
      IF(TEST.GT.EPS)GO TO 3
      VALUE=V2
      RETURN
      END

END OF SEGMENT, LENGTH   362, NAME  INTEG
```

DOCUMENT TRYA , NORMLTRYA : LP00 ON 05/03/75 AT 05.04

NCAT	MINB	MAXB	INTB	SIGMIN	SIGMAX	SIGINT	EPS	SIGRAT	RHO
4	0	2	5	60	110	10	0.00100	1.000	0.000

BMEAN	AVER	SIGMA	RTVAR	VAR	SLOPE	PF1	PF2	PF3	PF4	PR1	PR2	PR3	PR4
0.00	0.7532	0.60	0.444	0.197	0.000	0.751	0.245	0.004	0.000	0.249	0.004	0.000	-0.000
0.50	0.8766	0.60	0.512	0.262	0.446	0.637	0.349	0.014	0.000	0.363	0.014	0.000	0.000
1.00	1.1990	0.60	0.597	0.357	0.745	0.373	0.554	0.072	0.000	0.627	0.073	0.001	0.001
1.50	1.6215	0.60	0.636	0.404	0.887	0.138	0.605	0.246	0.009	0.862	0.257	0.011	0.002
2.00	2.0862	0.60	0.651	0.424	0.937	0.030	0.409	0.497	0.061	0.970	0.561	0.063	0.002
2.50	2.0862	0.60	0.637	0.406	0.945	0.004	0.162	0.599	0.234	0.970	0.561	0.063	0.002
0.00	0.8772	0.70	0.518	0.269	0.000	0.640	0.344	0.017	0.000	0.360	0.017	0.000	0.000
0.50	0.9846	0.70	0.569	0.324	0.398	0.551	0.412	0.036	0.000	0.449	0.037	0.001	0.000
1.00	1.2748	0.70	0.654	0.428	0.691	0.347	0.532	0.116	0.004	0.653	0.120	0.004	0.000
1.50	1.6753	0.70	0.708	0.501	0.847	0.154	0.540	0.282	0.024	0.846	0.306	0.024	0.000
2.00	2.1213	0.70	0.725	0.525	0.902	0.046	0.382	0.471	0.099	0.954	0.572	0.101	0.001
2.50	2.1213	0.70	0.700	0.491	0.911	0.009	0.182	0.529	0.279	0.954	0.572	0.101	0.001
0.00	1.0017	0.80	0.583	0.340	0.000	0.542	0.414	0.043	0.001	0.458	0.044	0.001	0.000
0.50	1.0975	0.80	0.627	0.394	0.358	0.476	0.451	0.070	0.002	0.524	0.072	0.003	0.000
1.00	1.3600	0.80	0.711	0.505	0.634	0.321	0.510	0.158	0.012	0.679	0.170	0.012	0.000
1.50	1.7316	0.80	0.772	0.596	0.799	0.163	0.487	0.303	0.046	0.837	0.350	0.047	0.001
2.00	2.1586	0.80	0.790	0.624	0.851	0.061	0.358	0.443	0.139	0.939	0.582	0.139	0.000
2.50	2.1586	0.80	0.752	0.565	0.848	0.016	0.193	0.474	0.315	0.939	0.582	0.139	0.000
0.00	1.1276	0.90	0.646	0.418	0.000	0.461	0.455	0.081	0.004	0.539	0.085	0.004	0.000
0.50	1.2128	0.90	0.685	0.469	0.323	0.412	0.471	0.110	0.008	0.588	0.117	0.008	0.000
1.00	1.4508	0.90	0.765	0.585	0.584	0.294	0.487	0.195	0.025	0.706	0.220	0.025	0.000
1.50	1.7969	0.90	0.828	0.686	0.742	0.165	0.446	0.315	0.073	0.835	0.389	0.073	-0.000
2.00	2.1932	0.90	0.844	0.712	0.803	0.073	0.335	0.414	0.177	0.927	0.592	0.179	0.001
2.50	2.1932	0.90	0.797	0.635	0.812	0.025	0.198	0.428	0.349	0.927	0.592	0.179	0.001
0.00	1.2528	1.00	0.708	0.501	0.000	0.393	0.471	0.124	0.011	0.607	0.135	0.011	0.000
0.50	1.3295	1.00	0.742	0.551	0.293	0.357	0.474	0.151	0.018	0.643	0.169	0.018	0.000
1.00	1.5456	1.00	0.816	0.666	0.538	0.267	0.464	0.225	0.044	0.733	0.269	0.044	0.000
1.50	1.8679	1.00	0.877	0.769	0.689	0.164	0.413	0.320	0.105	0.836	0.424	0.104	-0.001
2.00	2.2349	1.00	0.887	0.786	0.740	0.082	0.315	0.389	0.214	0.918	0.604	0.215	0.000
2.50	2.2349	1.00	0.834	0.696	0.746	0.033	0.199	0.390	0.377	0.918	0.604	0.215	0.000
0.00	1.3773	1.10	0.767	0.588	0.000	0.338	0.470	0.167	0.024	0.662	0.191	0.024	-0.000
0.50	1.4467	1.10	0.797	0.635	0.265	0.311	0.465	0.190	0.034	0.689	0.224	0.034	-0.000
1.00	1.6427	1.10	0.862	0.742	0.487	0.242	0.441	0.249	0.068	0.758	0.317	0.068	-0.000
1.50	1.9332	1.10	0.914	0.836	0.633	0.159	0.385	0.320	0.136	0.841	0.456	0.136	0.000
2.00	2.2755	1.10	0.920	0.847	0.689	0.088	0.297	0.366	0.249	0.912	0.615	0.249	0.000
2.50	2.2755	1.10	0.864	0.747	0.694	0.041	0.197	0.360	0.402	0.912	0.615	0.249	0.000

These articles have been removed for copyright reasons