UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT

Using Colour As A Tool In Discrete Data Analysis

Colin Ware
John C. Beatty

CS-85-21

August, 1985

Using Colour as a Tool in Discrete Data Analysis

COLIN WARE[1]
and
JOHN C. BEATTY[2]

Abstract

The advent of cheap high quality raster computer graphics opens up new possibilities for the display of numerical data. One of the advantages of computer graphics is that it enables an unprecedented degree of control over colour. Since colour vision is three dimensional this provides us with three additional perceptual dimensions, which may be used in addition to the two spatial dimensions available on a flat display to present multidimensional data. An experimental plotting package was constructed to explore the possibilities of using colour to examine discrete multidimensional data. The general insights gained in using this package are presented. Also, psychophysical reseaich is presented which probes the usefulness of colour in enabling human observers to perceive clusters of points in a multidimensional space. Comparing the resolution of clusters in colour and in space, the results show that colour is an effective extension of space for conveying information about data dimensions. However, the perceptual space defined by colour and space is not homogeneous and resolution is poor in a few specific directions. For this reason, the use of multiple views is advocated whenever colour is used as a tool in exploratory data analysis. An especially attractive aspect of colour is that it can be used to convey information to subjects without any special training.

---

[1] This work was carried out as part of the requirements for a masters degree in the Deptartment of Computer Science at the University of Waterloo. Colin Ware is currently an Assistant Professor in the School of Computer Science at the University of New Brunswick, Box 4400, Fredericton, New Brunswick, Canada, E3B 5A3.

[2] John C. Beatty is an Associate Professor in the Department of Computer Science at the University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

Acknowledgements

The authors are grateful for the assistance and suggestions of Dr. E. Jernigan, Dr. Kellogg S. Booth, Dr. Jane Gentleman (of Statistics Canada), Dr. William B. Cowan (of the National Research Council of Canada) and Dianne E. Ramey.

Table of Contents

Acknowledgements	2
1 INTRODUCTION	3
1.1. Properties of Colour Monitors	4
1.2. Properties of Human Colour Vision	5
1.3. Using Colour in Exploratory Data Analysis	9
2 AN EXPERIMENTAL PLOTTING PACKAGE	10
2.1. Colour Spaces Implemented	12
2.2. Perceiving Correlations	18
2.3. Perceiving Clusters	19
3 EXPERIMENTAL STUDIES	23
3.1. Choosing a Colour Space	24
4 EXPERIMENT ONE	26
4.1. Display Parameters	29
4.2. Instructions to the Subjects	30
4.3. Results and Discussion for Experiment 1	31
5 EXPERIMENTS TWO AND THREE	36
5.1. Cluster Creation	39
5.2. Procedure for Experiment 2	42
5.3. Results from Experiment 2	42
5.4. Procedure for Experiment 3	48
5.5. Training Session	50
5.6. Results and Discussion for Experiment 3	53
5.7. Individual Differences	56
6 MODELLING THE DATA	59
6.1. Model 1	59
6.2. Model 2	62
7 DISCUSSION	71
7.1. Generality of the Results	72
7.2. Practical Considerations	73
7.3. Summary and Conclusions	74
APPENDIX 1	76
APPENDIX 2	77
References	83

1 INTRODUCTION

The use of colour has long been popular with geographers, who use colour in maps to identify features of a terrain such as crops or rock types. Typically in these applications colour is given only a nominal function; it labels regions, which can then be identified by an external key. Another common mapping application uses a colour scale to encode a single continuous data dimension, such as height above sea level. And recently the use of colour dimensions to encode various kinds of information ranging from stress patterns in metal to the the distribution of energy in the universe has become popular. In a notable recent application of colour coding the 1975 US census chose to introduce a colour system encoding two data dimensions in chromatic codes [Meyer1981]. However, the colour coding scheme used by the census has been strongly criticised both on conceptual grounds [Trumbo1981] and by an experimental study which showed that subjects made an unacceptable number of errors when attempting to read data from graphs using the census scheme [Wainer1980].

Yet colour should be capable of conveying this information, and more. Research into human colour vision shows that the visual system contains three more or less independent chromatic "channels" through which we obtain all our information about the visual environment [Hurvich1981]. This raises the possibility that three data dimensions might be encoded in chromatic dimensions, in addition to the two spatial dimensions conventionally used [Sibert1980]. The most common examples of the use of colour to encode three physical variables (in addition to two variables encoded on the X and Y axes) is the pseudo-colouring of Landsat images of the earth's surface. These take data sampled at visible and non-visible wavelengths and map them into a colour space to make all the information visible. However, there has been very little attention given to the possibility of using colour to assist the visualisation of features in discrete data, as opposed to continuous maps.

The problem of interest here is the discovery of structures in multidimensional scatter plots. In this paper we shall argue both on conceptual and empirical grounds that colour can be a valuable tool for exploring this kind of data.

Probably the chief reason for the neglect of colour is that using colour in a controlled way has only recently become technologically feasible. Since print media are hard to manipulate and pigments mix in highly non-linear ways, they make the presentation of precisely defined scales of colours difficult in the extreme. All this has changed with the advent of cheap high quality computer graphics. In images displayed on a colour monitor it is possible to produce complex patterns in which each component is specified precisely, if necessary in terms of internationally standardised colour coordinates [Cowan1983]. Moreover, it becomes possible to investigate different colour mapping schemes with considerable freedom.

1.1. Properties of Colour Monitors

The most widely used colour output device employed for computer graphics is the colour monitor. These contain colour cathode ray tubes (CRT) of the kind found in a colour television sets, though they are capable of a much sharper image. The image displayed on the screen of a CRT is made up of a mosaic of glowing phosphor dots. There are only three kinds of phosphor in a given tube; these glow red, green and blue respectively, producing light as a function of the intensity of the beam of electrons which strike the back side of the face of the tube. All of the other colours which appear on the screen are produced as blends of the light from these three phosphor types. (See [Beatty1983] for further description of the capabilities of this kind of device).

1.2. Properties of Human Colour Vision Relevant to Data Display[3]

It is worth considering what we may expect from human colour vision in the light of results from the century old study of the subject by vision scientists.

The most important fact for the present discussion is that colour space is three-dimensional, a property arising from the presence of cone receptors in the retina having three different sensitivity functions. What this means in practice is that any colour can be matched by a mixture of three coloured lights; and this fact underlies the three dimensional nature of all colour spaces. This includes the red, yellow and blue primaries used by the artist, the red, green and blue primaries of the television monitor, the Munsell hue, chroma and value colour order system, etc. Because of this three dimensionality we can portray three distinct data dimensions through the medium of colour since any change in any of the data dimensions (mapped onto a colour axis) will result in a perceptually different colour.

Nevertheless the fact that colour is three dimensional does not necessarily mean that it provides the equivalent of three additional spatial dimensions in which to display information. Before deciding how colour may be useful in displaying information we must first consider some aspects of human colour perception. A convenient framework for discussion about what colour can and cannot be expected to give us is the classical taxonomy of measurement scales into nominal, ordinal, interval and ratio [Stevens1946] .

[3] There are a number of quite technical terms used in this discussion. The reader may wish to consult a general textbook such as [Boynton1979] for clarification.

Nominal: (The labelling function). Colour is useful in perception primarily to help us determine the location and surface properties of objects in our environment. For example we can discriminate an orange from the leaves surrounding it in a tree primarily by its colour. Perhaps because of this function, colour is extremely effective in isolating visual objects in an immediate and compelling way. Empirical support for this comes from experiments by cognitive psychologists [Kahneman1981] which show that colour has the advantage of being analyzed "preattentively" by the human visual system. That is, colour information is extracted in parallel, not serially as when reading text. This can be understood by considering an experiment in which an observer is required to locate the word "red" in a body of text. To do so he must sequentially read the text. Conversely, when he wishes to locate a red spot in a field of variously coloured spots he can move his eye immediately to that point. A relevant study showed that of a number of experimental variables, including shape, size and colour, colour was the most effective at capturing visual attention [Williams1966]. Because of this "preattentive" visual analysis, colour can be extremely effective in labelling information of different types.[4]

Using colour to label different classes of points on a graph is one nominal application of colour. Another is the labelling of regions on maps. A third might be highlighting different types of information in different colours in a data base information retrieval system. A fourth nominal application of colour is in systems of indicator lights. For these nominal applications the user should be warned that he can only effectively use a small number of colours - somewhere between 5 and 15, depending on the application. The problem is that if colours are not very distinct then simultaneous contrast effects [Cleavland1983] can cause confusion between them. Also, a map or graph which uses too many different colours will likely suffer from visual clutter.

[4] For a discussion of the general function of colour perception and a discussion on the relation between colour and attention see Chapter 5 in [Cowan1984] .

Ordinal: (The display of ordered information). The visual system is more effective in extracting relationships between adjacent coloured areas in the visual field than in determining absolute values. The two phenomena of adaptation and simultaneous contrast are relevant here. Adaptation is a gain change in the receptors of the eye that enables humans to function visually in environments which differ by as many as 7 orders of magnitude in the average level of illumination. Receptors respond to global level changes in the ambient illumination by adjusting their sensitivity range to be most effective in that particular environment. This gain change is gradual, which is why it takes us a while to adjust to bright light after coming out of a darkened room. Simultaneous contrast refers to the fact that the eye does a local differencing operation at the edges of objects, extracting the relative colour change between adjacent areas and sending this information to the brain, while absolute colour information is lost [Ware1983,Cowan1984].

The implication of adaptation and simultaneous contrast from the point of view of displaying ordered numerical data is that they should make it easy for the eye to detect the relative values of adjacent data objects (i.e. ordinal information) in the sense of whether a particular group of points is redder than another or darker than a third. Thus colour should be effective in displaying ordinal information so long as the points to be judged relative to one another are in close physical proximity. However, points which are far apart are likely to have different colours in their immediate surroundings. Because of this, the effects of local contrast may distort the ordinal relationship of the points.

An additional problem in the display of ordinal information using colour is obtaining a colour scale which is perceptually ordered. For example a grey scale from black to white is perceptually ordered, but a scale based on the physical spectum is not - people do not intuitively know that green lies between red and blue. A full analysis of what constitutes a perceptually ordered sequence of colours is an ongoing research area. Some further discussion can be found in [Trumbo1981]. We return to this issue in the description of the colour spaces implemented in the plotting package.

Interval: (The preservation of distance information). There is a considerable body of psychophysical data [Wyszecki1982] which is intended to provide "uniform colour spaces", i.e. a mapping from the physical description of the stimulus into a space in which equal distances correspond to equal perceptual distances. Given such a space it should be possible to make statements like "red 'A' is as different from pink 'B' as yellow 'C' is from brown 'D'." Considered out of context the availability of uniform colour spaces makes it seem possible to portray interval information. Unfortunately, the uniform colour spaces are all derived from experiments with colours viewed in isolation in very simple patterns. When these colours are surrounded by other colours the effect is to badly distort the colour space. In so far as they enhance differences the processes of adaptation and contrast will tend to distort the perceived size of distances between data values. Thus, for the most part, colour will be bad at presenting interval information for the same reasons that it is good at portraying ordinal information.

Ratio: (The preservation of ratio information). Given that colour will be poor at representing interval data we can anticipate that it will be hopeless at conveying ratios to the eye. This is because of the inability of the eye to perceive absolute light or colour values.

Although the above observations suggest that colour will be poor at conveying interval and ratio information, this need not overly concern us. In general, graphs are useful for displaying trends and structures in data, not in displaying actual values, which are often better presented in the form of a table. An important distinction to be made here is between the ability of a graph to display data *values* and its effectiveness in displaying data *structures*. We can anticipate that colour will be useful in showing data structures such as correlations or the presence of discrete clusters, because of the enhancement of colour differences by the visual system. The brain is continually looking for meaningful patterns in the distribution of light on the retina, and it can undoubtedly synthesize colour information and spatial information in the detection of objects. Thus for cluster analysis, colour should help us decide how many clusters there are in a data set, since points having a common colour and some measure of physical proximity should readily form visual objects. However, it may not be at all easy to discern where in the multidimensional space the individual points and clusters lie.

1.3. Using Colour in Exploratory Data Analysis

There is no statistical package which can take raw data and analyze it *a priori*. The statistician must have ideas about the kinds of structures present in the data before applying even the most general analytical procedures. Possible the most powerful tool for this initial exploratory data analysis is the graph. If the data appears to consist of an elongated structure he can apply a trend analysis or a curve fitting procedure. If it appears to be partitioned into discrete groups of points he can apply a clustering algorithm. However, since a conventional graph only displays two dimensions of data, special techniques are required to deal with data of more than two dimensions.

There have been a number of techniques devised to enhance the bandwidth of graphs for displaying multidimensional data. Of these the most common is the generalised draftsman's plot, which shows all pairwise combinations of variables in an array of separate two dimensional graphs. Unfortunately, although this method does display all the information, in order to perceive multidimensional structures the viewer must somehow integrate the content of many graphs - and this is difficult. Other graphical techniques for dealing with multi-dimensional data have been reviewed by [Gentleman1983] and [Chambers1983]. However, excepting for its use in pseudo-colouring satellite images, colour has been neglected in this application.

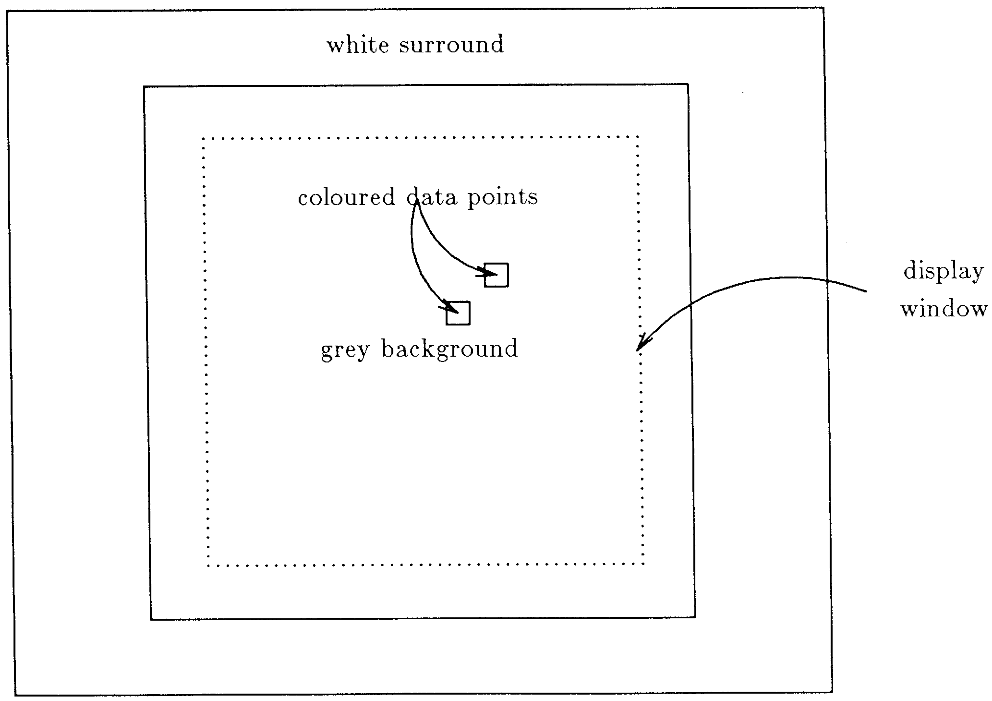


Figure 1: Scale drawing showing the components of the display. See text for size and colorimetric parameters.

2 AN EXPERIMENTAL PLOTTING PACKAGE

To investigate the utility of using colour as an aid in discerning structures in a multidimensional space we developed a display package to informally examine a number of data sets.

The plotting package was designed to read in data files and then allow the user to manipulate them in a number of ways. The basic display is given as a diagram in Figure 1. It consisted of a square display window into which the data was mapped, surrounded by a white band designed to provide a visual reference field for the viewer. The output device used was always a colour monitor. The following list describes some of the features and comments on their usefulness.

Scaling: The data were always scaled to fill the display space on all display dimensions. That is, the maximum and minimum values were found for each data dimension and these values were used to scale the data so that it fitted just within the X and Y display window and similarly filled the colour space being used.

Variable point size: The size of the data points was placed under user control. The size of the data points is especially important since the colour carries so much information. As a rule of thumb when there are many data points the points should be small and when there are few they should be large.

Variable background colour: By default the background colour was a neutral grey midway between black and the surround white. However, the user was able to change the background if desired. Since the background provided the reference relative to which the data points were judged, this could be used to selectively enhance or weaken features of the data. For example, if the data were divided into two clusters this would be more apparent if the background colour were placed roughly midway between the characteristic colour of the clusters.

Permutation vector: A permutation vector was provided to control which data points were mapped to which display dimensions. Since colour and space are not isomorphic visual dimensions, changing this mapping emphasises different features of the data.

Generalised draftsman's plot: A facility was provided for combining colour display with the generalised draftsman's plot. This is a plot in which all pairwise combinations of variables are plotted in a matrix of small graphs. This plot was enhanced by mapping the left over variables to data point colours, in the way described above, for each graph in the matrix. By giving multiple views of the data space, features which are obscured in one view are revealed in another.

Colour spaces: The user was provided with a choice of three colour spaces through which the data could be mapped. These are illustrated in Figure 2 and described below.

2.1. Colour Spaces Implemented in the Package

HSV

H, S and V, refer to Hue, Saturation, and Value, respectively. Hue refers to whether a colour is red, green, yellow or blue, etc. Loosely, hue relates to the position on the spectrum held by a given colour. Saturation refers to how far from neutral a colour is; a pale pink has low saturation, an intense red has high saturation. Value refers to how dark or light a colour is. The mapping from HSV into the RGB monitor coordinates was developed from Alvy Ray Smith [Smith1978]. This mapping is monitor dependent and only makes a very rough approximation to the psychophysically determined Hue, Saturation and Value quantities. However, it does effectively use the colour space obtainable from a colour monitor. Figure 3 shows equal saturation contours in the space which is obtainable from a colour monitor. It also shows equal saturation contours derived using

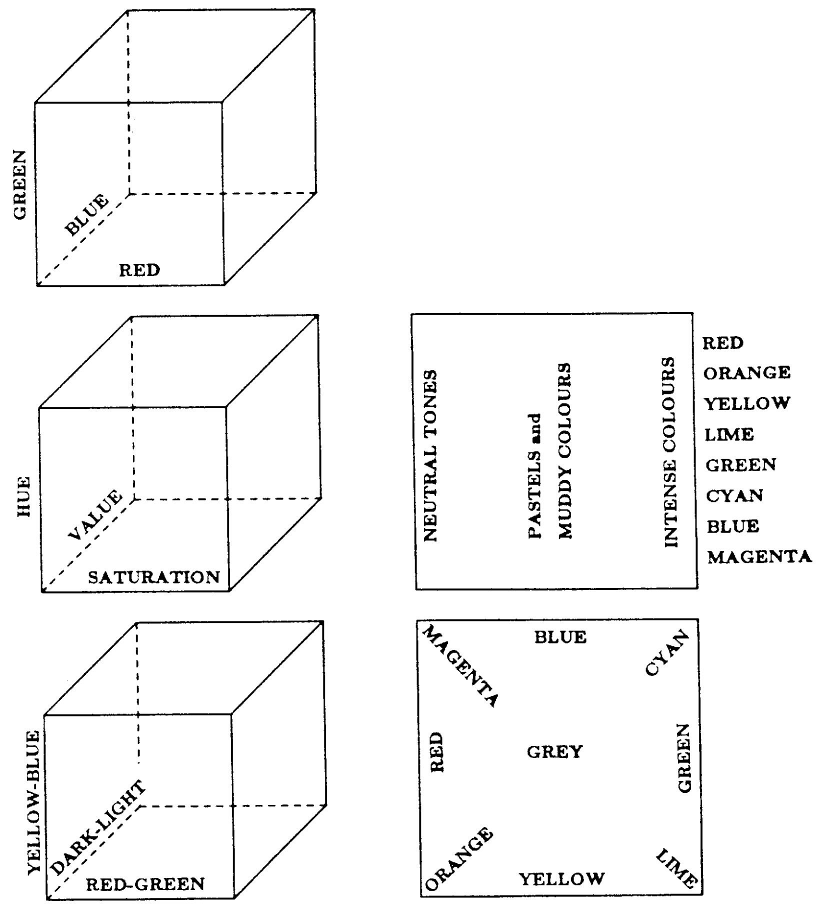


Figure 2. The axes of the three colour spaces through which data can be passed to form the colour of data points. Top: RGB. Middle: HSV. Bottom: Opponent Colour. For the HSV space the arrangement of colours on a plane defined by HUE and SATURATION axes, is shown. For the Opponent Colour space the arrangement of colours on a plane defined by RED-GREEN and YELLOW-BLUE axes, is shown.

Smith's mapping. The HSV display space suffers from extreme perceptual non-uniformities in the following sense, when the variable which is mapped to Value is near zero, structures present in the variables mapped to Hue and Saturation are not observable. Put simply, when a colour is close to black it is hard to tell how red or green it is. Also, when Saturation is low Hue is not discriminable. The HSV colour space does come close to creating valid ordinal scales, at least for two of its dimensions. The lightness dimension can be used to create a scale, running from black through intermediate greys to white. Or saturation can be used to create a scale from white through intermediate pink values to red. However, the Hue dimension of the HSV system is arguably not a psychologically valid colour ordering. This is evident when we sample it coarsely. For example, if we select the primary colours, red, green, yellow, and blue, it is not obvious perceptually how these should be ordered. However short sections of the colour space, e.g. from red, through orange may be perceived as continua. Thus a sequence of shades of orange is ordered.

RGB

The RGB colour space is one in which data dimensions are mapped directly to the amount of light generated by the red, green and blue phosphors of the colour monitor. In creating create the RGB colour space it is usually desirable to correct for nonlinearities in the relationship between voltages sent to the electron guns in the monitor and the luminance output function of the phosphors. This function can be measured and the non-linearity corrected by the use of an appropriately scaled colour look-up table (this type of calibration is called *gamma correction* and is described in [Cowan1983]).

One advantage of the RGB colour space is that if the monitor is calibrated it is relatively simple to convert all colour values displayed on the monitor to CIE tristimulus values - an internationally standardised way of specifying colour. A second advantage of the RGB space for present purposes is that it effectively uses the entire gamut of the display device, unlike perceptually defined colour spaces, which may have arbitrarily shaped boundaries and are likely to be a poor fit with the colour space of the display device. This point is elaborated in Figure 3. A disadvantage of the RGB colour space is that it does not result in a perceptually ordered sequence of colour if a value on a single axis is varied. For example, consider the following scale created from the RGB space. The Red and Blue values are fixed at 50% and Green is varied from zero to 100%. The result will be a continuum from purple through grey to a pale green. People do not perceive the purple to grey section of this continuum as a variation in greenness and thus will not correctly read data encoded as change in a single data variable.

Opponent Colour.

This was an attempt to adapt the Opponent colour space of Hurvich and Jameson [Hurvich1981] to the use of information display. The mappings created were not intended to be psychophysically precise, but were intended to approximately create red-green and yellow-blue opponent colour axes from the colour gamut obtainable from the monitor. Considerable problems were encountered fitting a meaningful scheme into the RGB cube, which is the colour gamut of a monitor. Visual non-uniformities resulted which were even more severe than those arising from the use of the HSV space. There is, unfortunately, a bad mismatch between the axes of the opponent space and the gamut of colours obtainable from a monitor. This problem is illustrated in Figure 4, which shows one attempt to adapt the opponent colour space to the constraints of a monitor.[5]

[5] The space implemented in the package is very close to one independently arrived at by Naiman [Naiman1985], although it differs from his in that the package implementation mapped values in the range 0.0 - 1.0 in the three opponent axes to the range 0.0 - 1.0 in the RGB axes. The C code for implementing the Opponent colour space is reproduced in Appendix 1.

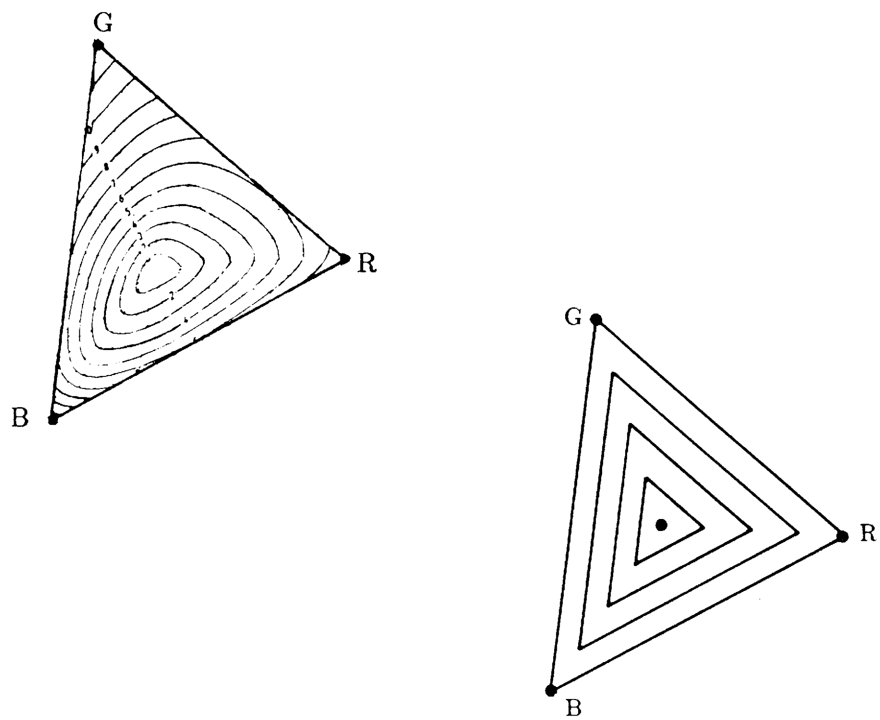
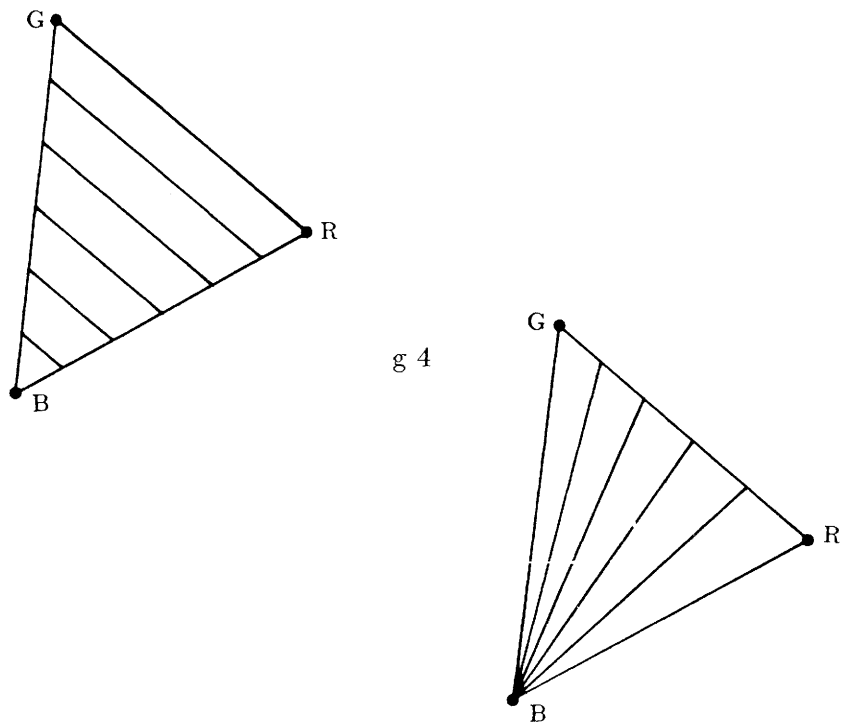


Figure 3. In the upper graph the triangular region represents the range or *gamut* of colours obtainable using an RGB monitor plotted in CIE chromaticity coordinates. At the corners are the colours of the three phosphors and inside the triangle are colours produced by mixtures of the phosphors. Neutral colours are in the central region of the triangle. The roughly elliptical contours represent psychophysically determined equal saturation contours, taken from [Wyszecki1982] p. 512. Saturation is the "vividness" of a colour. Thus, grey has zero saturation, a muddy red has low saturation and a brilliant red has high saturation. To construct a psychophysically valid colour space with the same range of saturation available irrespective of hue, the gamut of usable colours would have to be restricted to those colours which lie within one of the untruncated equal saturation contours. Only part of the available colour space would therefore be usable.

The lower graph shows equal saturation contours from Alvy Ray Smith's HSV space plotted in the same way.



g 4

Figure 4. The triangle in these diagrams represents the gamut of a colour monitor plotted in CIE chromaticity coordinates. This diagram shows equal Y-B contours in the upper plot and equal R-G contours in the lower plot

Again, with respect to providing perceptually ordered information the Opponent colour space is worse than the HSV space. The dark-light axis remains, but we now have saturation and value replaced by the two colour opponent axes. The problem with these opponent axes is that it is not perceptually obvious that grey lies between red and green; neither is it perceptually obvious that grey lies between yellow and blue. Furthermore, the Opponent colour space suffers from the same non-uniformities when there is a low value on the dark-light axis as does the HSV colour space.

2.2. Perceiving Correlations

Using the package described above we made a number of observations concerning the perceivability of correlations and clusters artificially created in multidimensional space. The following facts emerged about perceiving correlations. These are derived from the comments of a number of observers.

1) It is easy to perceive a correlation between a variable which is mapped onto space and a variable which is mapped onto a color dimension. Human sensitivity to this structure is comparable with sensitivity to perceiving correlations between a variable plotted on the X-axis and one plotted on the Y-axis.

2) It is very hard to perceive correlations between two variables, both of which are mapped to colour dimensions. For example, if a variable mapped to Saturation is correlated with a variable mapped to Value (all other variables being random), dark points will tend to have low saturation, while light points will tend to have high saturation. The net result will be an increase in the number of highly saturated colours. A very high degree of correlation was necessary before this effect could be observed.

3) RGB space is good for allowing the existence of a correlation to be perceived, but it is confusing when the question, "which variables are correlated?" is asked.

4) Using HSV space it is usually possible to perceptually determine which variables were correlated. However, a higher degree of correlation is necessary before its existence can be recognised.

2.3. Perceiving Clusters

The following points emerged from informal observations of the perceivability of clusters in a five-dimensional space.

1) RGB space is good for allowing observers to perceive the number of clusters which exist in a five-dimensional space, but it is confusing when the question, 'Where are those clusters located?' is asked.

2) Using HSV space it is usually possible to perceptually determine where clusters are located, but greater cluster separation is necessary before clusters can be distinguished.

3) Opponent colour space is as poor as the HSV space in terms of identifying the number of clusters present, and lies somewhere between HSV and RGB in terms of identifying the location of clusters.

Figures 5 and 6 illustrate results of our initial informal investigation of clusters. In Figure 5 three clusters were created in a five dimensional space and displayed using three existing techniques devised for looking at multidimensional data, in addition to our new technique using colour. The three existing techniques are Chernoff faces [Chernoff1973], Star plot and the generalised draftsman's plot [Chambers1983]. For the colour plot the five dimensions are mapped to displacement along on the X-axis, disaplacement along the Y-axis, luminance of the red phosphor, luminance of the green phosphor and luminance of the blue phosphor, respectively. The three cluster structure is only discernible using the colour plot, where it appears as three clouds of points, each having a different colour.

The advantages of colour become more apparent when there are more data points involved. Stars and faces become impractical and colour becomes a valuable enhancement of the generalised draftsman's plot. Figure 6 shows a generalised draftsman's plot with and without colour showing 200 data points, far too many for the faces or star plotting techniques. With the aid of colour is it apparent that there are four clusters present in the data.

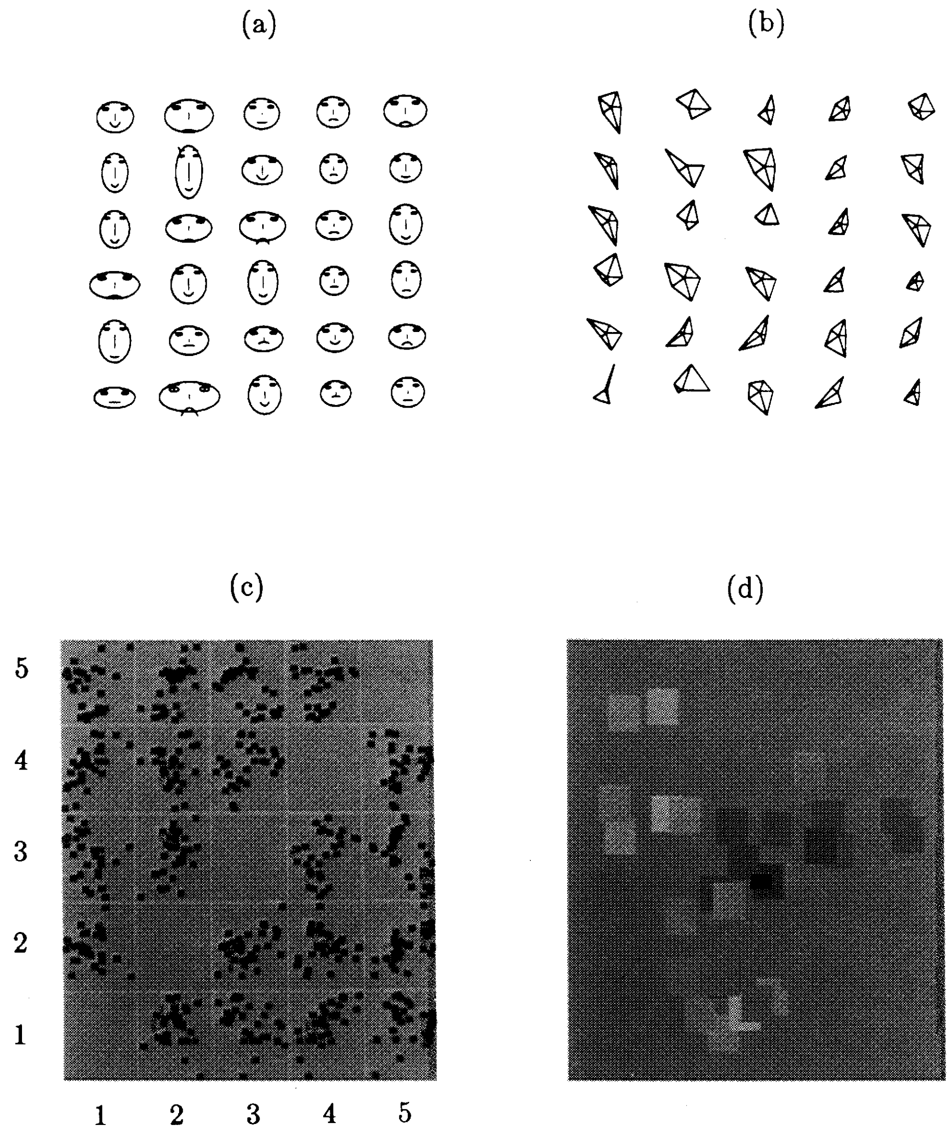


Figure 5. The same 30 data points plotted in 4 different ways: (a) faces, (b) star plot (c), generalized draftsman's plot, (d) colour enhanced plot showing view (5,4) of the generalized draftsman's plot. Only in (d) is the three cluster structure perceivable. The three clusters are: 1) the 6 pale blue and grey points in the upper left quadrant, 2) the 12 green points in the left central region, 3) the 12 pink points in the lower central region. Notice that two of the points in cluster 3, a pink point and an orange point, belong spatially to cluster 2. It is only because of their colour that they can be correctly allocated.

The pictures of the display reproduced in this and the next figure are taller than they are wide. This is an artifact of the reproduction process. The display space was square on the face of the colour monitor.

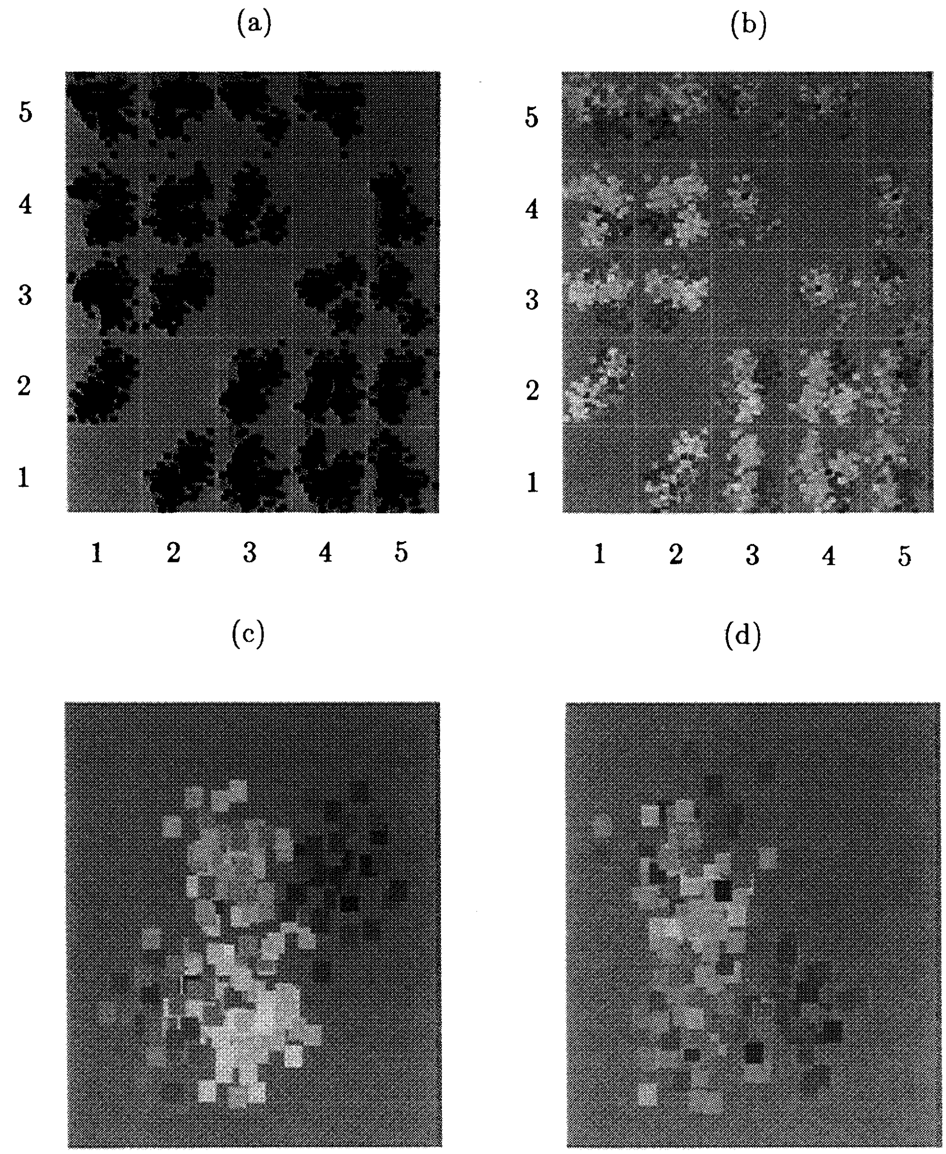


Figure 6. The usefulness of colour increases as the number of clusters and the number of points increase. These plots show 200 points scattered in a 5-dimensional space. The points are grouped into four clusters. (a) shows a generalised draftsman's plot. (b) shows a colour enhanced generalised draftsman's plot. (c) shows view (2,3) of the generalised draftsman's plot. (d) shows view (4,5) of the generalised draftsman's plot.

3 EXPERIMENTAL STUDIES

In initiating this investigation we were concerned with the general problems attending the use of colour as a tool in exploratory data analysis. It would have been useful for this investigation to have a classification of the kinds of meaningful structures for which the statistician may wish to look, given discrete data. With aid of such a classification we could systematically ask the question, "will colour help us in detecting this structure and how sensitive will it be?" Unfortunately, we do not have such a taxonomy, and in its absence it is impossible to make general statements about how useful colour is in conveying an arbitrary structure to the eye. Therefore, for the purpose of this investigation we decided to arbitrarily pick a particular type of structure, namely, clusters with normal density functions (forming hyper-ellipsoids in a five-dimensional space) and experimentally investigate the utility of colour in facilitating the detection of these structures.

We chose to investigate cluster perception for three reasons. Firstly, since colour is used perceptually to define visual objects (vis the "preattentive" property mentioned above), groups of points having a similar colour should stand out perceptually as visual objects. Secondly, in looking for clusters we are generally more interested in how many there are, not in exactly where they are (this can be discovered later using different techniques). Thus the problems of reading colour coded data values - which were anticipated in the earlier discussion of the properties of human colour vision - are not important here. Thirdly, the distortion produced by simultaneous contrast will be much reduced in scatter plots, where the data are discrete, as compared to coloured maps, where the data are continuous. In scatter plots the data points all share a common background, which can act as a visual reference. This can be deduced from the empirical observation that contrast effects are a function of the separation of stimuli.

The series of empirical studies we now describe was intended to experimentally examine the capacity of human colour vision and spatial vision to convey information to the brain about clusters in a five-dimensional space.[6] The first study was intended as a simulation of the situation which might be faced by a statistician looking for an unknown number of clusters in a data set. The second and third studies, which are more critical, look at the ability of colour to enable subjects to discriminate clusters along various vectors in a five-dimensional perceptual space defined by X and Y spatial axes and R, G and B colour axes.

3.1. Choosing a Colour Space

It is necessary to say a few words about the particular choice of colour space used, namely the RGB colour space. The RGB colour space is instrumentally defined rather than perceptually defined. Perceptual colour spaces are derived from a conceptual model of human colour perception. (See for example, [Meyers1981]). The HSV and Opponent colour spaces discussed above fall in the category of perceptual colour spaces. Unfortunately, as we have seen, perceptual colour spaces are often severely nonuniform with respect to perceived differences. For example, most perceptual colour spaces have a "lightness" or "brightness" axis and two chromatic axes. A consequence of this is that when there is a low value on the lightness axis there can be little resolution on the chromatic axes. Perhaps the most important point here is that we were investigating the use of colour as a tool in *exploratory* data analysis. Thus it is not possible to make any assumptions about the distribution of data. If we did know something about the data in advance then it would be possible to align the colour axis in ways that optimally displayed specific features. In the absence of such foreknowledge, the best strategy

[6] The object of study was individuals with normal colour vision. All subjects were screened for colour anomalies using standard pseudoisochromatic plates [Ichikawa1978].

is to choose a colour space with generally useful properties. Although RGB is not an optimal colour space, it was chosen because it uses the entire display space of the colour monitor and it does not have the extreme nonuniformities of the colour spaces which use brightness as an axis.

4 EXPERIMENT ONE

The purpose of the first experiment was to determine the extent to which colour helps in identifying how many clusters exist in a five-dimensional space. A secondary purpose was to determine the effects of learning by comparing the results obtained from 14 subjects, each of whom had little experience with the task, with the performance of 1 subject (the author C.W.) who performed the experiment 14 times and had considerable experience with experiments concerning colour perception, but did not, of course know the sequence of stimuli, or how many clusters were going to appear at a given time.

The algorithm used to construct the stimuli used on a given trial was designed to simulate a variety of data sets. This algorithm contained a number of random parameters which determined how many clusters would appear (between 1 and 6), how large their standard deviations were in each of the five data dimensions, and how their centres were distributed on each of the five dimensions. The subject's task was simply to estimate how many clusters there were on a given trial. On half the trials colour was used to display three of the data dimensions and space (i.e. position relative to the X and Y axes) was used to present the remaining two. On the other half of the trials only spatial information was presented and all the points were black. The performance measure used was how much the subject's estimate of the number of clusters present on a given trial differed from the number of centres randomly allocated to that trial. Note that this measure does not relate directly to any statistical measure of clustering; since allocation of points to centres was also random there may have been cases of clusters created with no points in them, or cases of clusters created with centers which were nearly coincident in the 5 space. Nevertheless, in so far as a subject is capable of making guesses which are close to the number of cluster centres created on a given trial this will reveal that the subject is using information about the data. What follows is a more detailed description of the experiment.

The stimulus generation algorithm for experiment 1: The clusters generated in the first experiment consisted of 40 points divided into between one and six clusters, distributed randomly in a five-dimensional space.[7] Given below is the sequence of steps used to generate the stimuli used in experiment 1. Essentially, this consisted of two phases; one in which the stimulus pattern was generated and a second in which it was scaled to fill the display space. The algorithm contains a number of arbitrarily chosen parameters determining the distribution and spread of clusters. These were adjusted to optimise the number of response errors and thus make it possible to discriminate between experimental conditions. Since errors were the independent variable it was important that they be kept at a fairly high level. The procedure used was as follows.

1) With equal probability, select a number of clusters between one and six.

2) Choose centres randomly (and uniformly) distributed in a 5-dimensional space. Position p_i (i = 1,2,3,4,5) is chosen independently for each of the five dimensions such that $0 \leq p_i \leq N$, where N is the number of clusters selected for a given trial.

3) Choose standard deviation values independently for each cluster and each dimension. The average standard deviation was 4.5 and the standard deviation of the standard deviation values was 0.1 (this was to ensure that clusters would have a variety of sizes and shapes).

4) Generate clusters according to the parameters chosen above.

[7] The alogrithm used to generate uniformly distributed pseudo random numbers for all experiments is from [Box1958] while the algorithm used to generate normally distributed random numbers is from [Pike1965] . The authors are grateful to Victor Klassen for the C code implemenation of these algorithms.

Cluster Generation Algorithm

```
/* make number of clusters */
number_of_clusters = (a random integer from 1 to 6)

/* make centres */
for i=1 to number_of_clusters
    for j=1 to number_of_dimensions
        centre(i,j) = (a random number uniformly distributed
                  between 0 and the number_of_clusters)

/* make standard deviations */
for i=1 to number_of_clusters
    for j=1 to number_of_dimensions
        standard_deviation(i,j) = normal() * 0.1 + 0.45
        /* the standard deviations have an average value
        of 0.45 and they vary about that with a standard
        deviation of 0.1*/

/* make data points */
for k=1 to number_of_points
    i = (a random integer from 1 to number_of_clusters)
    for j=1 to number_of_dimensions
        point(k,j) = centre(i,j)
            + standared_deviation(i,j) * normal()

end Cluster Generation

/* Note that normal() is a function which returns a random
number from a population with a mean at zero and a standard
deviation of one. */
```

Figure 7. The procedure used to generate clusters for experiment 1 is given in the form of pseudocode.

5) Scale the data so that it filled the display space by finding maximum and minimum values for each display dimension and scaling appropriately.

6) Display the data, mapping the five data dimensions to X, Y, R, G and B display dimensions, so that the data points exactly filled the display window on the screen and also were scaled to be linear between the minimum and the maximum signal on the red, green and blue phosphors respectively. As a result of this procedure clusters could be of various shapes and sizes, elongated or roughly spherical.

The data generation model is given in the form of pseudocode in Figure 7.

4.1. Display Parameters

The following display parameters were common to all three experiments. The stimuli were presented on a 13 inch Electrohome ECM1301 monitor in a darkened room where the only light source other than the display monitor and the terminal was light from a dimmed table lamp set to provide illumination just sufficient for reading the keyboard. The stimulus arrangement is shown in Figure 1. Data points consisted of 0.6 cm squares placed in a 11.2 cm square display window, on a neutral grey background having a luminance of approximately 45 cd/m^2 measuring 14 cm square. A white surround was included to provide a visual reference and a constant state of adaptation. Its luminance was approximately 80 cd/m^2. The subject was located 270 cm from the screen. These measurements can be converted to degrees of visual angle by multiplying each centimeter value by 0.212.

To fully specify the stimuli it is necessary to know the chromaticity coordinates of the monitor phosphors. To this end the spectral energy distribution of the phosphors was measured spectroradiometrically and from these measures the following CIE chromaticity coordinates were derived

	x	y	z
Red Phosphor:	0.620	0.330	0.05000
Green Phosphor:	0.210	0.675	0.11500
Blue Phosphor:	0.150	0.060	0.79000

The maximum luminance values for the red, green and blue phosphors were 1.93, 9.30 and 1.98 respectively.[8]

4.2. Instructions to the Subject

Instructions were delivered on the terminal with illustrative examples appearing on the colour monitor. The subject was first required to enter his or her name and then to read through "pages" of information appearing on the terminal screen. The subject had control over when a new page would appear and could also turn pages back to view earlier information.[9]

Space considerations prohibit the full reproduction of these instructions here but in essence they covered the following points.[10]

1) The subject was told that a cluster was a group of points, and was given three examples of clusters with colour and three examples of clusters without colour. In each case a written commentary described the position and the colour (if it was used) of the clusters in the example.

2) The subject was told that he or she would be shown between 1 and 6 clusters and that the task would be to estimate the number and enter it on the keyboard.

[8] The luminance values were determined by measuring the luminance of the green phosphor using a photometer and measuring the relative values of the red and green phosphors using a psychophysical technique (the minimally distinct border - see Cowan, 1982).

[9] The authors are grateful to M.W.Schwarz for providing the code which was used in this part of the user interface

[10] The instructions are reproduced in full in Appendix 2.

3) The subject was told to expect 20 trials with colour and 20 without colour, and was told which to expect first.

4.3. Results and Discussion for Experiment 1

The data from 14 relatively inexperienced observers are shown in Figure 8. This shows three measures of performance: the percentage of correct identifications of the number of clusters generated; the average absolute error - that is the average size of the departure from a correct estimate, irrespective of whether this was an underestimate or an overestimate; and also, the average size of the error taking direction into account, with a positive value indicating an overestimate and a negative value indicating an underestimate.

Consider first the number of trials for which the observer correctly identified the number of clusters: the data for the inexperienced observers shows a clear advantage in the use of colour for all conditions except that in which there was only one cluster, when overall the data show a slight advantage for no colour. The overall advantage given by colour was found to be highly significant ($p < 0.001$) by the Wilcoxon matched pairs signed ranks test. The benefits of colour increase with the number of clusters in the data set, with the maximum benefit occurring for trials where there were six clusters present. In this case with the help of colour subjects correctly identified 6 clusters in 38 percent of the trials while without colour no correct identifications occurred.

A possible interpretation for the disadvantage offered by colour on the single cluster trials is that, given a lack of experience with colour, observers tended to overinterpret the data and identify more clusters than actually existed.

Consider next the data on the absolute size of the error in each of the cluster categories. This data shows the same pattern of increasing benefit for colour as the number of clusters increases, with a slight disadvantage for colour when only one cluster was present. Overall these data

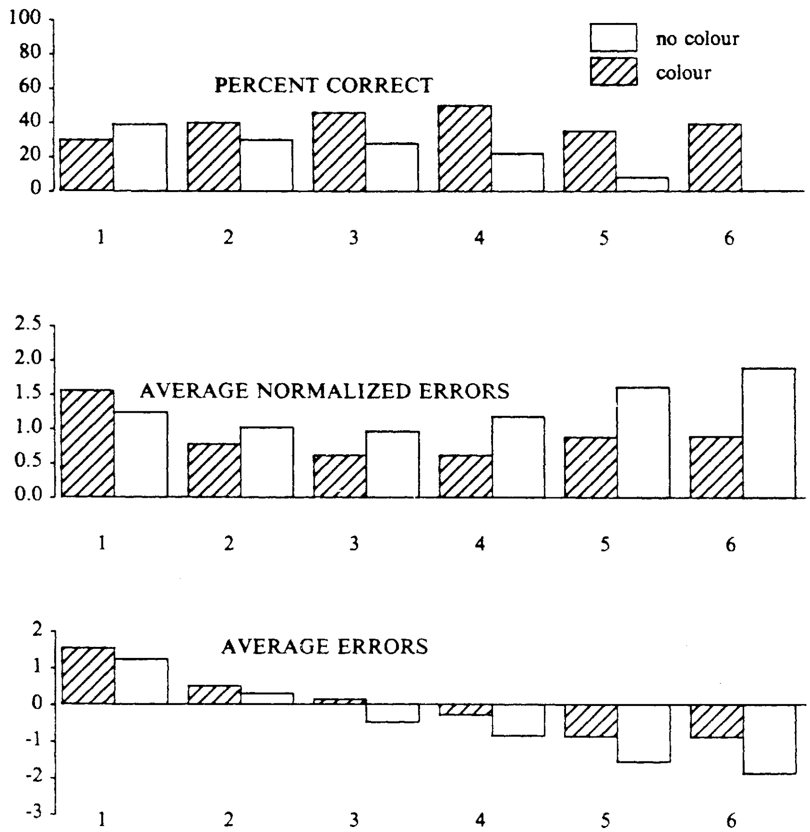


Figure 8: Data obtained from 14 relatively inexperienced observers who participated in experiment 1. Shaded bars show data obtained with colour while open bars show data obtained without colour. In the top graph the height of each bar represents the percent correct for between 1 and 6 clusters. The middle graph shows the average size of the discrepancy between the number of clusters present and the number the subject estimated. This data is normalized by taking the absolute values of the errors. The lower graph shows the direction of the error; a positive value indicating an overestimate on average.

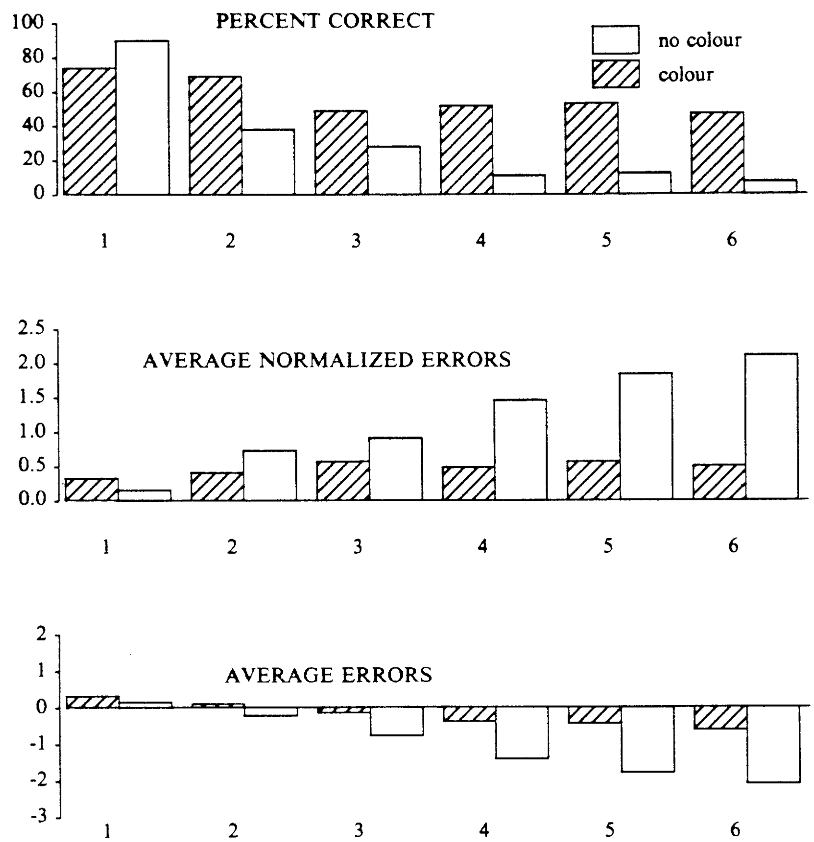


Figure 9: Data from a single experienced subject. See legend of Fig. 8 for graph conventions.

present a U-shaped function with the best performance (fewest errors) in the middle of the range and the worst at each end.

The data giving the direction of error shows that on average, observers tend to overestimate the number of clusters when there are few and underestimate the number when there are many. This underestimation is markedly greater without the assistance of colour.

Comparing the data obtained from the 14 inexperienced observers with that of the single subject run 14 times (Figure 9), the most dramatic difference appears when there are a small number of clusters. For the number correct measure we can see a large improvement in performance when there are one or two clusters, both with colour and without colour. For example, where without colour inexperienced observers identified a single cluster 40% of the time the experienced observer identified a single cluster 90% of the time. This difference is also reflected in the error size measure. Presumably, these differences reflect tendencies for inexperienced observers to perceive structures in noisy data which are not in fact present. Thus both colour and lack of experience contribute to the tendency to overinterpret data.

An encouraging aspect of the results from the perspective of using colour as a general purpose tool in data analysis is that the inexperienced observers did better in the conditions where colour coding was used; they were able to use colour information even though they had been provided with only a single training session lasting approximately five minutes. This ease of learning is supported by the observation that the experienced observer really shows very little improvement in his performance over inexperienced observers with the aid of colour when there are more than two clusters present. This suggests that colour coding does not need to be an esoteric tool of the specialist but can be used in presenting information to a largely untrained audience.

Although this experiment suggests that colour may be useful in displaying extra data dimensions it can yield us no information about *how* useful it is. The problem is the large number of variables involved in the experiment, each of which may affect cluster perception in unknown ways. Clusters may have a number of shapes, be large, small or elongated. And we cannot know how all these variables affect cluster perception. This is the cost of attempting to simulate 'real' data. In order to make precise statements about the relative efficiency of colour in conveying information about clusters we need a much simpler experimental paradigm. The experiments which follow give more meaningful answers to the issue of cluster resolution, but only in the restricted context of a situation in which there can be only one or two clusters present in the display.

5 EXPERIMENTS TWO AND THREE

The purpose of experiments 2 and 3 was to explore the threshold envelope for distinguishing clusters which differ along various vectors in the five-dimensional perceptual space defined by two spatial and three colour dimensions. By doing detailed measurements in a tightly controlled situation it was possible to make direct comparisons between the ability of observers to resolve colours in XY space (the *de facto* norm) and their ability to resolve clusters separated along other vectors in the five-dimensional space, including clusters which are separated in colour alone and clusters which are separated in both colour and space. Experiments 2 and 3 were two investigations of the same problem; experiment 2 was designed to investigate the lowest possible thresholds which might be obtained by intensive study of a single individual, whereas experiment 3 was a pared down version of the same experiment, designed to get an idea of what these thresholds are with a number of different subjects under somewhat more realistic conditions.

To make these experiments manageable the experimental situation was simplified to one in which the task faced by the subject was to determine whether there were one or two clusters in a particular display. Thus, the goal was to assess the envelope for discrimination of two clusters from one cluster in the 5-dimensional perceptual space. Unfortunately, even with this simplified experiment, to sample this envelope with any appreciable density would require an enormous number of trials.

For these experiments we decided to sample the envelope of thresholds for cluster separation by measuring separation along vectors defined in the following way. Allow each of 5 coordinates to have a value of -1, 0 or +1. Taking all combinations this yields 3^5 or 243 different vectors. Since we are not interested in the zero vector this leaves us with 242 vectors. In addition, each two cluster stimulus involved clusters separated along opposite vectors. Thus the number of two cluster stimulus conditions is given by the expression $(3^5-1)/2 = 121$. In what follows, when a direction of cluster separation is given, only one vector is specified for reasons of economy. However, the reader should understand that the opposite colinear vector is always also involved. We represent the display space by the 5-tuple (X Y R G B). Thus, (1 0 0 0 0), represents separation on the x-axis only, and has an implicit opposite vector (-1 0 0 0 0). The vector (1 0 1 -1 0) represents clusters which are separated so that they are more red and less green on the right hand side of the screen and less red and more green on the left hand side of the screen; it has the implicit opposite vector (-1 0 -1 1 0). Observe that this notation is intended to capture only the direction of cluster separation. The magnitude of cluster separation is specified independently.

Since the purpose of these experiments was to compare colour and space, it is more meaningful to break down the set of experimental conditions in terms of spatial components and colour components. We can consider the set of conditions as the set derived by taking all combinations of the set of possible spatial separations (Table 1) with the set of possible distinct colour separations (Table 2).

Table 1

X-axis	Y-axis
0	0
1	0
-1	0
0	1
0	-1
1	1
1	-1
-1	1
-1	-1

Table 2

	RED	GREEN	BLUE
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1
8	1	-1	0
9	1	0	-1
10	0	-1	-1
11	1	1	1
12	1	-1	1
13	1	1	-1
14	1	-1	-1

Observe that in Table 1, the complementary of each entry is present whereas the complementaries of the entries in the colour table are not represented. The reason for this is to eliminate complementary vectors from the set which results when the two tables are combined. This can best be made plain with an example. Suppose for the moment, that the complementary values were present in both Table 1 and Table 2. We will have entries (0 1) and (0 -1) from Table 1, and entries (1 0 0) and (-1 0 0) from Table 2. Taking all combinations we get (0 1 1 0 0), (0 -1 1 0 0), (0 1 -1 0 0) and (0 -1 -1 0 0). The last two of these are redundant because they are colinear with the first two. Instead of removing complementaries from Table 2 we could have achieved the same effect (though less economically) by keeping complementaries in Table 2 and removing them from Table 1.

Removing the complementaries from Table 2 has not entirely eliminated all redundancy. Taking the product of the 9 XY possibilities with the 14 RGB possibilities we get 126 conditions. There are still 5 redundant conditions introduced by adding the first entry in Table 2 to all the enties in Table 1. The condition represented by (0 0 0 0 0) is spurious since clusters are not separated for this condition. Furthermore, one member of each of the following vector pairs is redundant because they are colinear: (1 0 0 0 0), (-1 0 0 0 0); (0 1 0 0 0), (0 -1 0 0 0); (1 -1 0 0 0), (-1 1 0 0 0); (1 1 0 0 0), (-1 -1 0 0 0). This reduces the total number by 5 to give the 121 stimulus conditions.

5.1. Cluster Creation

Centres for the two clusters were defined by taking the difference vector for that trial multiplied by both $+\frac{1}{2}$ and $-\frac{1}{2}$. When the difference vector was zero then the two clusters had the same centre. Having created the cluster centres, 40 points were randomly allocated to the two centres and normally distributed about them on all display dimensions. Thus all clusters were of the same average size, that is, they had a uniform constant probability density function which was normally distributed and the same in all directions.

As in the first experiment the data were first generated and then scaled to the display space. However, this scaling resulted in an unforeseen and undesirable effect. In the preliminary design stages, it was discovered that thresholds for cluster discrimination were significantly lower for clusters separated obliquely at 45 degrees on the screen than for clusters which were separated either horizontally or vertically. The reason for this was the scaling. The earliest manifestation of separation of points into two clusters is an elongation of the cloud of points. Scaling along x and y to fill the display space removes horizontal or vertical elongations of the data, but oblique elongations remain. Figure 10 makes this point graphically. Since we were specifically interested in perception of discrete clusters (i.e. separation of the cloud of points into two distinct groups), the fact that subjects could use elongation as a clue to cluster separation before they appeared as separate groups was deemed undesirable. The solution adopted was to scale along the vector of cluster separation prior to scaling to the display space. This preliminary scaling was designed to (on average) give the cloud of points the same oblique extent in the direction of cluster separation as in other oblique directions. The corrected procedure was as follows:

1) Generate clusters separated by amount specified for that trial and as part of the generation procedure scale down along the vector of cluster separation.

2) Scale to fill the display space.

All display parameters, such as point size and background colour, were the same as used in experiment 1.

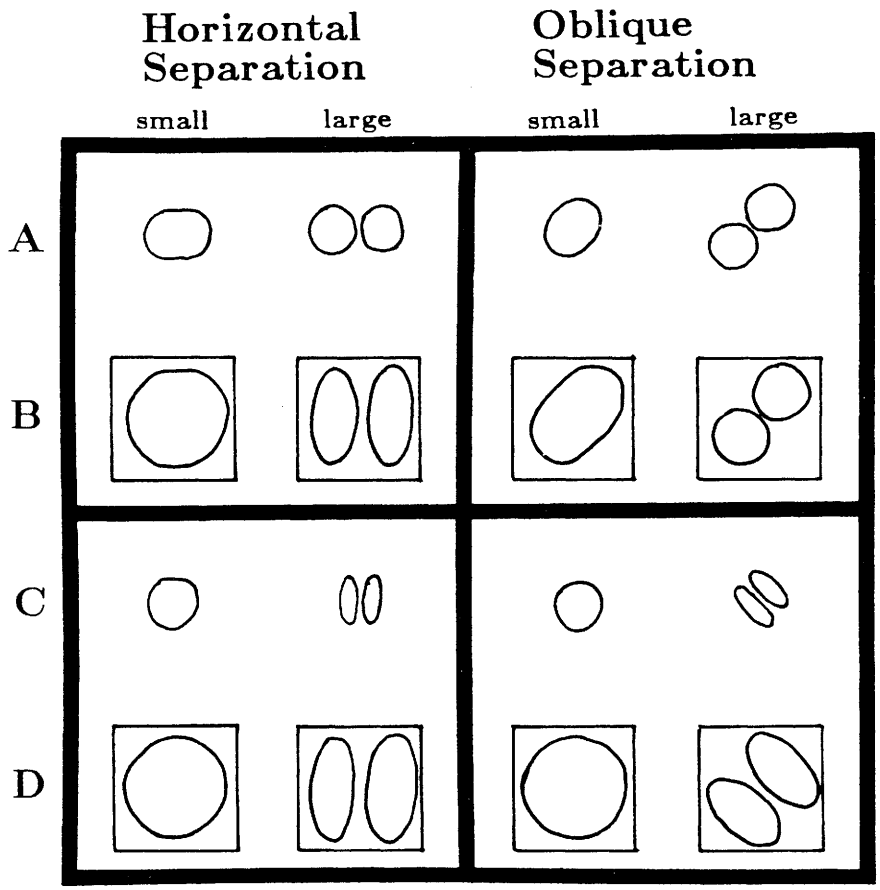


Figure 10. In the early stages of experimentation a problem was discovered in that clusters separated along oblique axes were perceived as distinct when they were closer together than clusters separated either vertically or horizontally. This figure shows the reason for this. The blobby shapes in "A" represent clusters separated by a small amount and a large amount respectively in horizontal and oblique directions. "B" shows what happens to these clusters when the data are scaled to fit the display space. The small horizontal separation becomes indiscernible, while the small oblique separation apears as an elongation in the oblique direction. Since we were interested in perception of distinct clusters, not in perception of elongated blobs we decided to scale in the direction of cluster separation. "C" and "D" show the effects of this scaling. Now the oblique situation and the horizontal situation are equivalent; clusters can only be discerned when the cloud of points separates into two distinct groups.

5.2. Procedure for Experiment 2

The method used in experiment 2 to establish the threshold for discriminating two clusters was the double staircase procedure [Hake1966]. This is a technique for measuring thresholds which is designed to reduce the effect of experimenter bias. The procedure involves initially presenting a stimulus which is well above threshold, in this case two clusters which are clearly discriminable. The subject responds "double" by pressing the designated key. On the next trial the distance between the two clusters is *decreased* and the subject again responds "double" if he can tell there are two. This sequence is repeated until he responds "single" at which point the difference between clusters is *increased* for the next trial. The separation is then increased iteratively until the subject responds "double" again. The net effect of this procedure is that when cluster separation is plotted against trial number an oscillating path around the threshold is traced out. The version of the procedure used for this experiment had some additional subtleties. One was that the step size (by which the clusters increased or decreased their separation) was reduced at each reversal. This had the effect of causing the subject to "home in" on the threshold. In addition, trials in which there really were two clusters were randomly interspersed with trials for which there was only one cluster. In this way the subject could never know if he was actually seeing one cluster or two. Pseudocode for the double staircase algorithm is given in Figure 11.

5.3. Results from Experiment 2

The variance of the data was found to correlate highly with the threshold. A log transformation made the variance more homogeneous across experimental conditions (see Figure 12). Thus all statistical analysis was done on the log of the raw data scores.

Double Staircase Algorithm

```
while (reversals < CRITERION) do

    in random order do
        Test( 0 )
        hightest = Test( highval )
        lowtest  = Test( lowval )

    if new reversal on hightrace
        half stepsize
        reversals = reversals + 1

    if new reversal on lowtrace
        half stepsize
        reversals = reversals + 1

    if ( response to hightest is "two clusters" )
        decrease highval by high_stepsize
    else increase highval by high_stepsize

    if ( response to lowtest is "two clusters" )
        decrease lowval by low_stepsize
    else increase lowval by low_stepsize

    if (false positive)
        increase lowval
        increase highval
end while

end Double Staircase
```

Figure 11. The procedure used to determine thresholds for experiment 2 is given in the form of pseudocode.

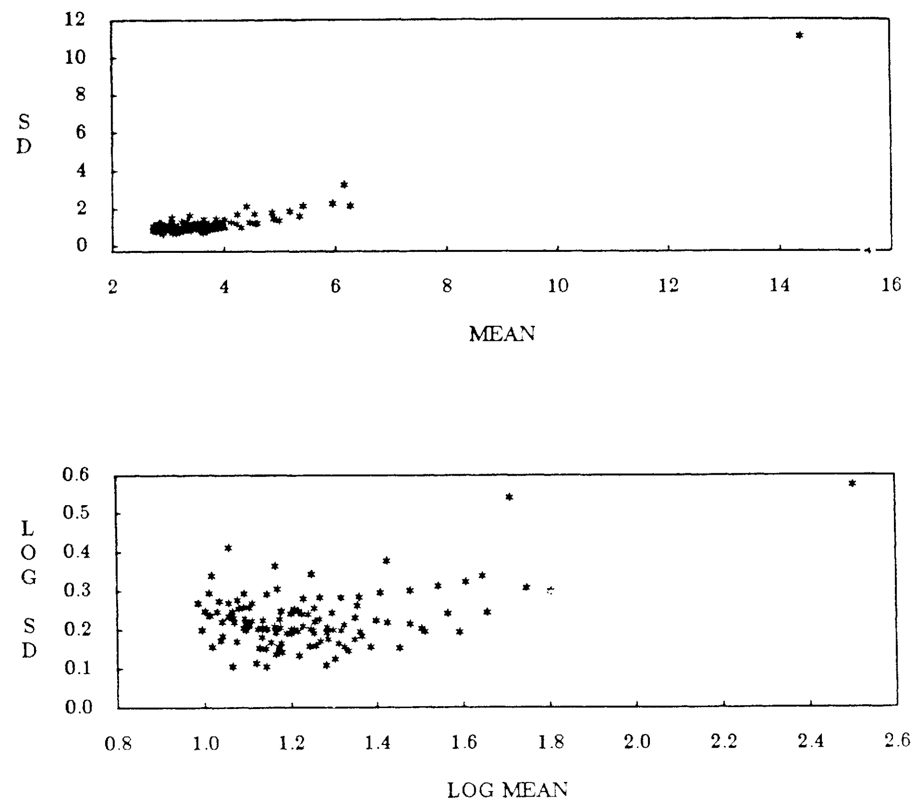


Figure 12. The top graph shows the means from experiment 2 plotted against their standard deviations. The standard deviations correlated highly with the means. In the lower graph the same data are plotted but the means and standard deviations are based on log transformed raw data.

Since there was no discernible bias in terms of thresholds for left oblique cluster separation versus right oblique cluster separation these categories were collapsed into one 'oblique' category. Similarly, vertical and horizontal scores were combined to give one 'major axes' category.

This leaves three spatial categories in place of the 9 shown in Table 2. They are: 1) separation in colour only, 2) separation partly in colour and partly along a major spatial axis, and 3) separation partly in colour and partly along an oblique spatial axis. Figure 13 shows a histogram of the mean thresholds for the 14 colour categories in each of these spatial categories. Values in this plot represent thresholds for perceived cluster discrimination. These are measured in units which correspond to the standard deviation of the generic cluster.

Summarising the major effects which appear here: the threshold for separation on major axes only (i.e. X or Y) is 3.09 while the threshold for separation on obliques is 3.02. There is no significant difference between these two values; thus we take their average 3.055 to represent the threshold for discriminating clusters separated only in space. This value is marked by the dotted horizontal line on the bar plots.

It can be seen that thresholds for clusters separated only in terms of their colour are somewhat higher than those where clusters are separated in terms of colour and space components. In particular, the threshold is much higher when the separation is along one specific colour direction, namely when clusters are separated only by the magnitude of the signal sent to the blue gun. Also, when clusters are partly separated in colour and partly in space, the highest threshold for cluster separation was obtained when the colour component was blue gun separation.

How much worse is cluster resolution in colour only as compared to separation in colour and space? To make this comparison we plotted the log threshold of clusters separated by colour only against the log threshold for clusters separated by colour and major axes. (See Figure 14). These points form a roughly linear distribution, with the exception of one outlier

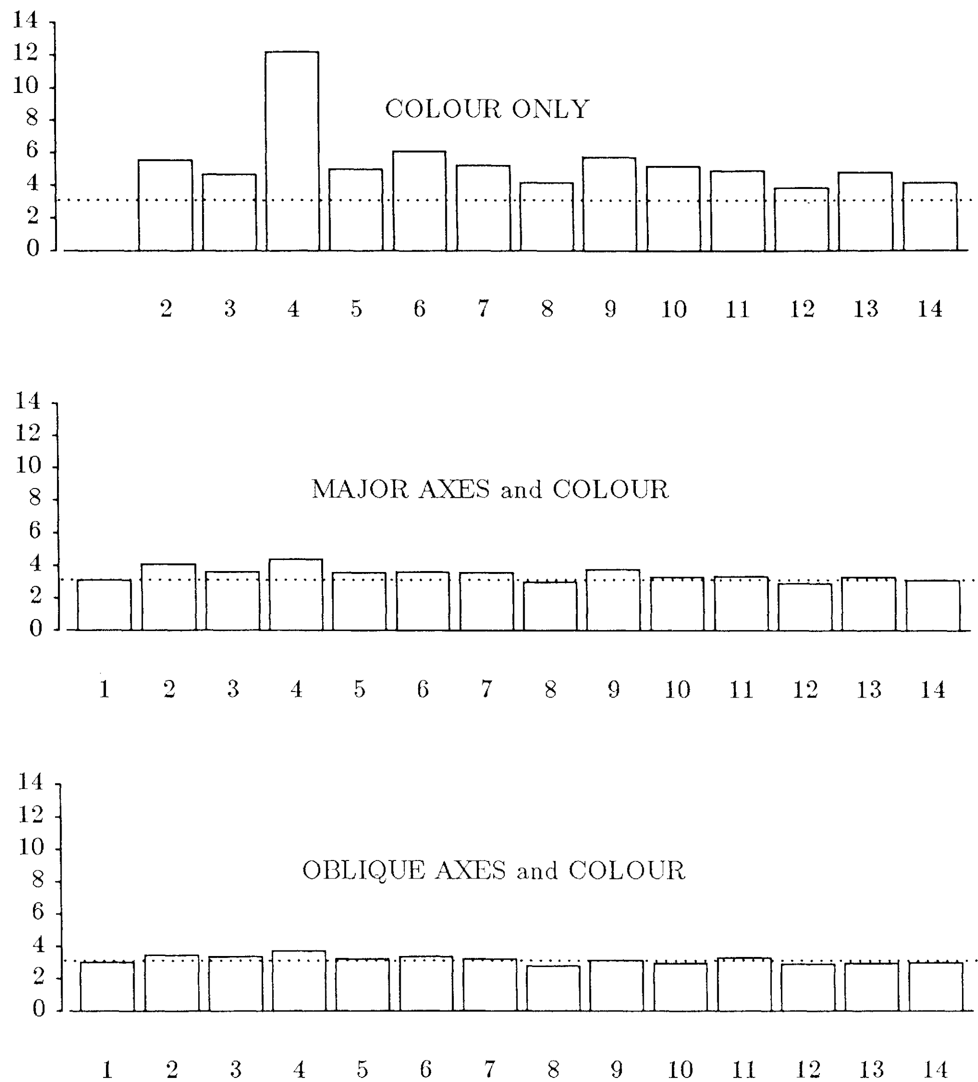


Figure 13. A bar plot summarising the data from experiment 2. Bars show the mean thresholds for discriminating clusters separated along various vectors in XYRGB space. The top plot shows cluster discrimination thresholds for clusters separated by colour only. Labels underneath each bar correspond to entries in Table 2. There is no bar for the first position in the top plot because this is the condition corresponding to zero separation. The middle plot shows cluster discrimination thresholds for clusters separated partly along a major display axis (that is, either vertically or horizontally) and partly in colour. The lower plot shows cluster discrimination thresholds for clusters separated partly by colour and partly along an oblique axis. The horizontal dashed line denotes the average threshold for distinguishing clusters which are separated by space alone.

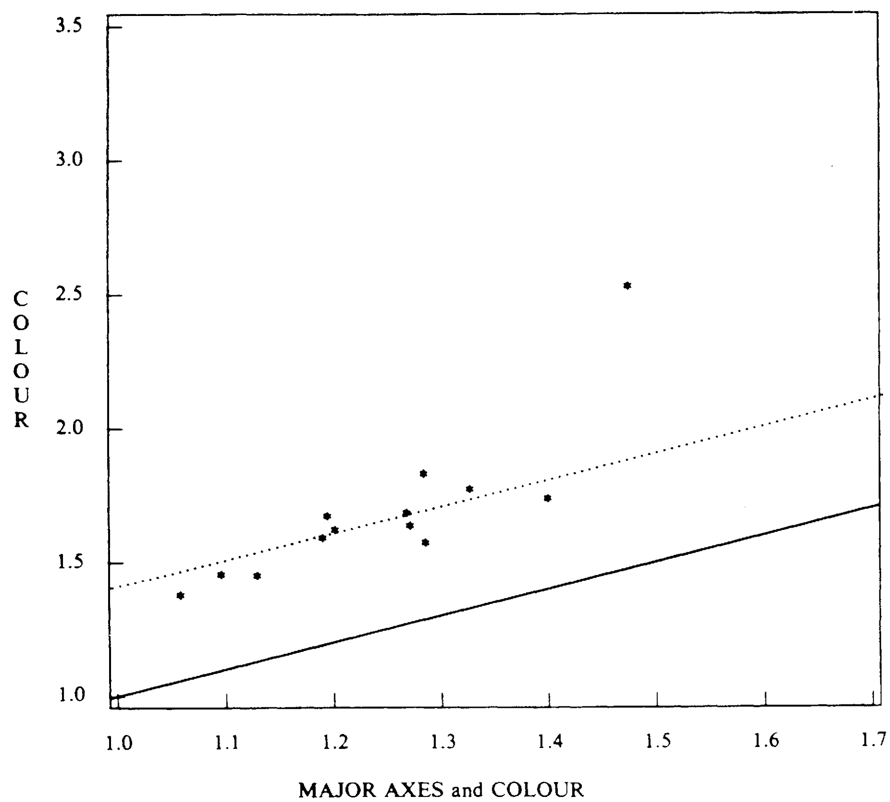


Figure 14. Shows a different representation of the data in the top and the middle bar graphs of the previous figure. The log of the threshold for resolving clusters separated by colour alone is compared with the log of the threshold for resolving clusters separated by space and colour. Thus the y coordinate of data point i is the log threshold for clusters separated along vectors corresponding to the ith bar in the top graph of Figure 13 (separation in COLOUR ONLY), while the x coordinate is the log threshold for clusters separated along vectors corresponding to the ith bar in the middle graph of Figure 13 (separation in MAJOR AXES and COLOUR).

The data points would fall on the oblique solid line, representing identical thresholds for colour alone and colour plus spatial separation, if there were no penalty for viewing clusters separated only in colour. To give an indication of the kind of penalty paid for having clusters separated only in colour, the parallel dotted oblique line shows a multiplication factor of 1.5 over the solid line. Thus the threshold is approximately 50% worse if clusters are separated in colour alone than if they are separated by colour and space.

(the point representing separation on the blue gun alone). Because this is a log plot, adding a constant value on a given axis represents a constant multiplication factor. The oblique diagonal line drawn though the points represents a multiplication factor of 1.5 over the situation of perfect agreement between conditions. This line was placed by eye to give a very rough indication of the perceptual penalty of having clusters separated only in colour. It can be seen that to a first approximation clusters separated only by colour have to be 50% further apart than clusters separated by colour and space before they can be discriminated.

Before attempting to interpret what these results mean in practical terms we present experiment 3.

5.4. Procedure for Experiment 3:

The important differences between experiments 2 and 3 were:

1) The data were obtained from multiple subjects, thus making more general conclusions possible. A corollary of this is that the experiment was considerably pared down since it was not feasible to obtain 60 hours of observation per subject from 8 subjects.

2) The experimental setup was more representative of the situation of a statistical analyst faced with unknown data. To achieve this the double staircase procedure was abandoned in favour of a procedure in which clusters were tested for distinctness at a fixed set of separations, namely 2, 4, 8, 16 and 32 standard deviations apart. In this experiment the subject could have no idea in advance of how clusters would be separated in the five dimensional space because all conditions were randomised. (Note that in the double staircase procedure used in the previous experiment the subject could expect either no separation or separation along a particular vector). Thus he had a clearer idea of what to look for.

To reduce the the number of hours per subject over experiment 2 we took advantage of the finding that certain classes of conditions were found experimentally to be equivalent. Consider, for example, clusters separated along (1 0 1 0 0) and (1 0 -1 0 0). In the first instance we have redder clusters on the right and less red clusters on the left. In the second we have redder clusters on the left and less red clusters on the right. In other words the conditions are mirror symmetric about a vertical axis. Since the results from experiment 2 showed no left-right asymmetries the two conditions were collapsed to one (which consisted of a random alternation between the above two), as were all analogous pairs. However, we did not go so far as to assume that vertical is the equivalent of horizontal or oblique. This has the effect of reducing Table 2 from a nine entry table to the four entry Table 3.

Table 3

	X-axis	Y-axis
1	0	0
2	1	0
3	0	1
4	1	1

Reading the table from top to bottom, it represents 1) a set of conditions in which cluster separation is by colour only, 2) a set of conditions in which cluster separation is by colour and horizontal displacement, 3) a set conditions in which cluster separation is by colour and vertical displacement, and 4) a set of conditions in which cluster separation is by colour and oblique displacement.

Taking the cross product of table 3 with the 14 entry table of colours (Table 2) we get 56 conditions, from which we subtract the zero separation condition, giving 55 conditions. This is less than half the 121 conditions of experiment 2.

For each of the 55 conditions we tested cluster discriminability for clusters separated by 2, 4, 8, 16, and 32 standard deviations. This yields 275 conditions. Add 75 conditions for which there was only one cluster and we get the 350 trials which were presented to each subject in a random order in a single experimental session.

There were 10 such sessions for each subject. Note that although the set of values for cluster centres was the same for all experimental sessions, the normally distributed clouds of points around those centres were created independently for each trial.

5.5. Training Session:

Prior to the experiment proper the subject was given a training session which consisted of a run through the entire (randomised) sequence of 350 trials, but with the essential difference that after responding (one cluster or two) feedback was given on each trial concerning actual cluster separation.

This feedback took the form of a visual indicator in the lower left corner of the display which showed the colour direction and spatial direction of cluster differences as well as a bar whose length indicated the size of cluster differences. In addition to this every 10 trials the subject was told the percentage of "hits" (correct responses when there were two clusters) and the percentage of "false positives" (identification of two clusters when there was only one).

Stabilising response criterion: In the task described above the subject could adopt a conservative strategy of minimising errors at the expense of hits, or the opposite strategy of optimising hits. In order to enable subjects to adopt a reasonably consistent response strategy the feedback on overall performance (hits and false positives) was maintained throughout the series of experimental sessions and the subjects were requested to maintain false positives at 30% while maximising hits. Subjects evidently found this easy to achieve and produced remarkably stable performance. This feedback had the additional advantage of maintaining the subjects' level of motivation over the course of the experiment.

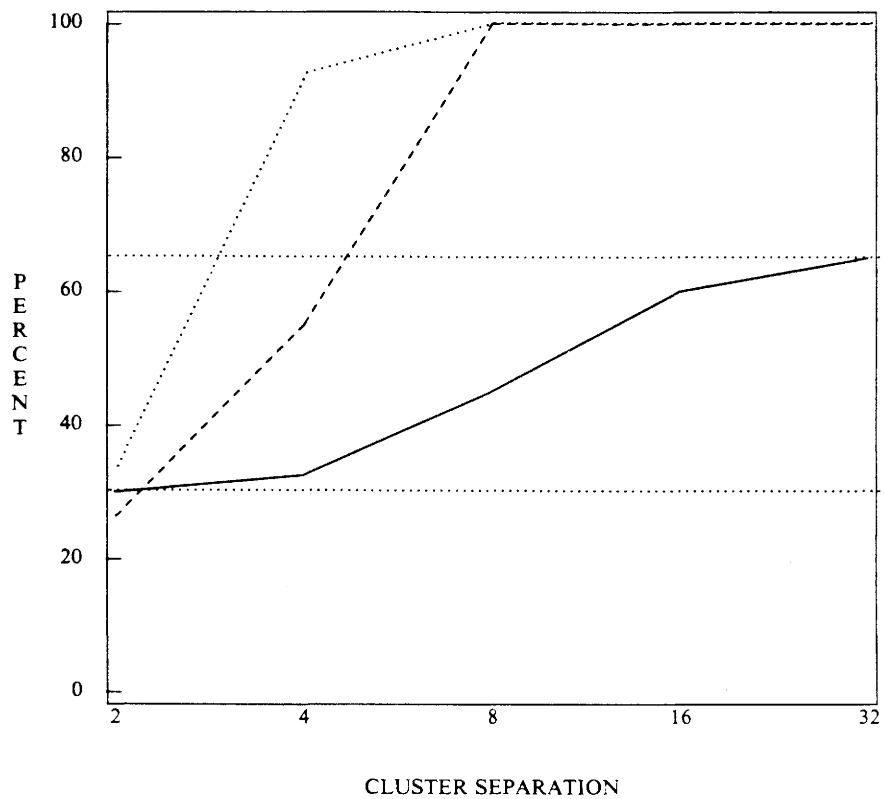


Figure 15. Percent correct is plotted against cluster separation for three of the cluster separation directions tested in experiment 3: (0 0 1 0 1) is the solid line; (1 0 1 0 -1) is the dashed line; (1 1 1 -1 -1) is the dotted line; The upper horizontal dotted line shows the 65% threshold criterion. The lower horizontal dotted line shows the average false positive rate at 30%.

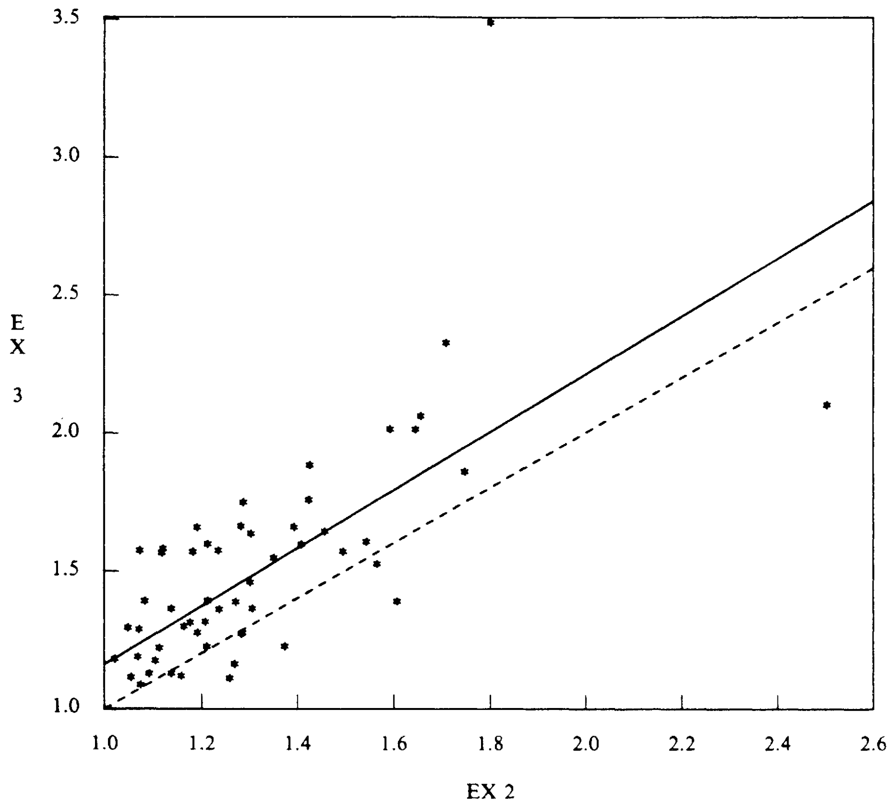


Figure 16. The log thresholds from each of the 56 categories of experiment 3 is plotted against the same data from experiment 2 which has been collapsed to create the same set of conditions. The solid line is a linear regression fit to the data. The dashed line is the fit which would be obtained if the scores in the two experiments were identical. The fact that the solid line is above the dashed line represents a threshold which is approximately 20% higher for experiment 3 than for experiment 2.

5.6. Results and Discussion for Experiment 3:

Thresholds were derived from the data of experiment 3 by first extracting the total number of 'hits' (correct identifications of two clusters) for each of the conditions. The threshold was taken to be 65%, midway between the error rate at 30% and 100%. To determine the threshold for a given condition the scores at the various separations (2, 4, 8, 16, 32 standard deviations) a piecewise linear interpolation was used. The point where the interpolated line crossed the 65% threshold criterion determined the threshold separation. Figure 15 illustrates the threshold determination for three cluster separation directions.

Figure 16 shows the results of experiment 2, plotted against those of experiment 3 in the form of a scatter plot of the log transformed data. This was done by collapsing the results of experiment 2 to the same categories as those used in experiment 3. The correlation between the two studies accounts for 47% of the variance and is highly significant ($p < 0.01$), showing that there is substantial agreement between the two studies. A linear regression of experiment 3 data onto experiment 2 data produced a line with a gradient very close to 1. The vertical displacement of this line on the log log plot corresponds to a threshold which is approximately 22% higher for experiment 3. This 22% difference in threshold may be attributable to two factors; part may be due to the different experimental technique used in the two experiments and part may be due to the relative inexperience of the subjects involved in experiment 3. Unfortunately, there is no basis for discriminating between these two factors with the given data. Nevertheless, the fact of the overall agreement adds credibility to the results of both experiments since two radically different experimental techniques point to the same conclusions.

Figures 17 and 18 show the data from experiment 3 plotted in the same manner as the plots for experiment 2. Overall, the pattern of results appears to be substantially the same. That is, the highest thresholds were obtained for separation by colour only, with one condition in

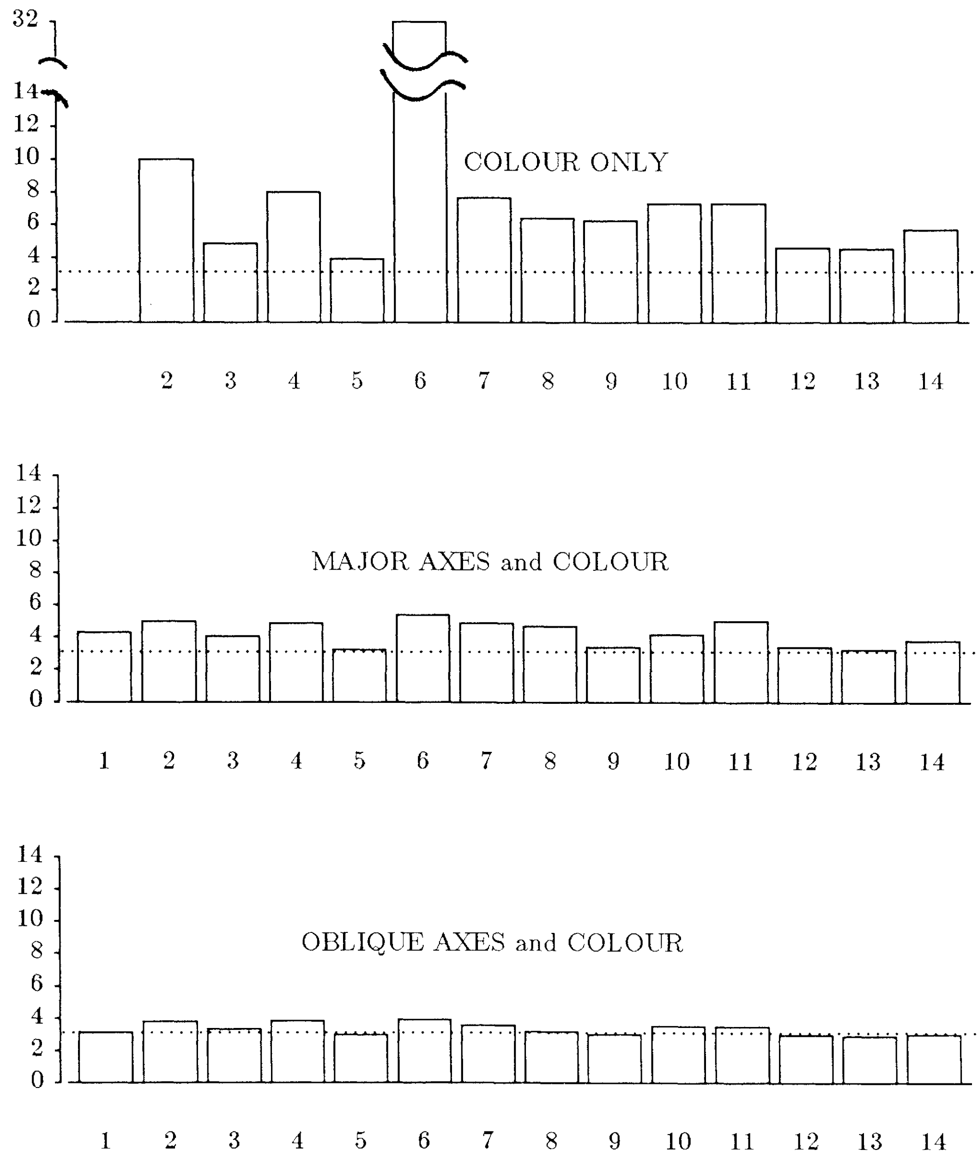


Figure 17. Data from experiment 3 summarised in the same manner as that for experiment 2. See figure 13 for labelling conventions.

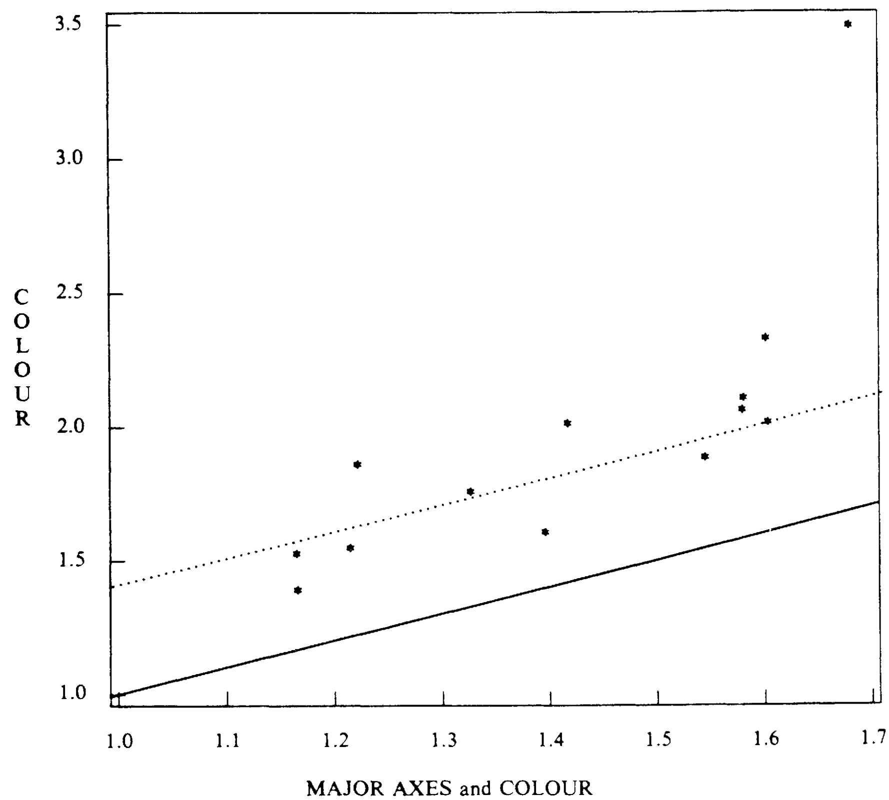


Figure 18. The data from experiment 3 plotted in the same manner as that for experiment 2 for comparison (see Figure 14) The plot compares the log of the threshold for resolving clusters separated by colour alone with clusters separated by space and colour. The dashed oblique line shows a multiplicatation factor of 1.5; it is not a statistical fit but is placed to facilitate visual comparison between this figure and Figure 9. Thus the threshold is approximately 50% worse if clusters are separated in colour alone than if they are separated by colour and space. Note that the penalty is approximately the same as for experiment 3 as for experiment 2.

particular showing a very high threshold. However, the condition for which the threshold is the highest differs in experiments 2 and 3. In experiment 2, clusters separated on the variable sent to the blue gun only gave the highest threshold. In experiment 3 clusters separated on the red and blue guns produced the highest thresholds. The reason why this threshold is so high deserves some note. This is one of the thresholds illustrated in Figure 15 and it can be seen that for this condition the detection rate only just reached the defined threshold. If the threshold criterion had been defined as 60%, for example, instead of 65%, the threshold would be a more modest 16 standard deviations.

Thus, in spite of the above discrepancy between experiments 2 and 3, there is general agreement between the two experiments; the second highest threshold obtained in experiment 2 is the highest for experiment 3, and there is general agreement on which conditions yield a high threshold and which yield a low threshold. Figure 18 shows that the penalty for display using colour only is about a 50% decrement in performance, approximately the same for the 8 subjects of experiment 3 as for the one highly practiced subject of experiment 2.

5.7. Individual Differences

An important issue regarding the utility of colour coding of data is that of individual differences. It might be the case that some people, who are not "colour blind", will have great difficulty in using colour coded information.[11] If such people are common then the utility of colour coding data is substantially reduced.

---

[11] There are, of course, in the population at large a substantial percentage of people who are colour blind in the sense of having one of the well defined colour anomalies - between 8 and 10 percent of males - and these people will unavoidably have difficulty with colour coded information.

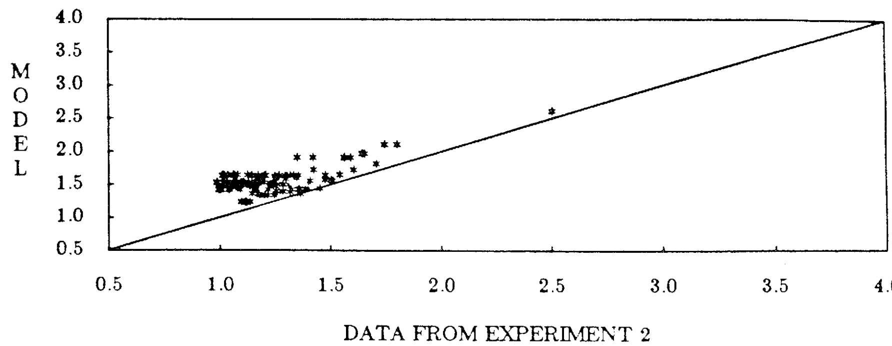
To look for evidence that colour may be harder for some individuals to use we did an analysis of individual differences.

We collapsed the data for each subject into the four categories defined by Table 3. Then we determined thresholds for discriminating clusters in each category and for each subject. The results for 8 subjects are shown in Table 4 where the standard deviation of individual clusters is the unit of measurment.

Table 4

Subjects	Categories			
	1	2	3	4
s1	6.429	4.540	4.157	3.127
s2	7.305	4.563	4.611	3.814
s3	5.608	3.495	4.529	3.169
s4	4.961	3.534	4.207	3.309
s5	7.167	4.393	3.817	3.410
s6	6.977	4.676	4.540	3.487
s7	5.920	4.211	4.031	3.180
s8	5.486	3.626	4.262	3.084

Individual differences were largest in category 1, which represents clusters separated by colour only, and smallest for category 4, which represents clusters separated in colour and an oblique direction. Overall the subjects tested in this experiment differed very little from one another in their scores. Certainly there is no evidence that any of them had a substantial problem with using the colour information.



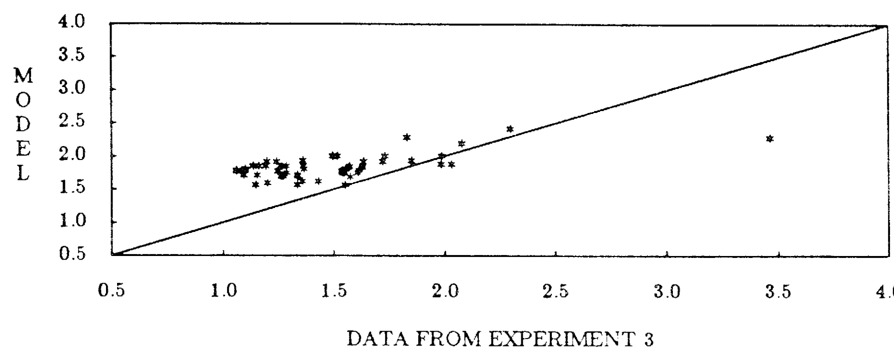


Figure 19. The Predictions of Model 1 are plotted against the 121 data points from Experiment 2 and the 55 data points from Experiment 3.

6 MODELLING THE DATA

The advantage of being able to model the data is that 1) a successful model would allow the prediction of cluster separation thresholds along an arbitrary vector, not just the vectors measured, and 2) with the aid of a model it would be possible to deliberately enhance certain views by choosing to orient the data optimally with respect to the display space.

6.1. Model 1

The simplest model which could be applied to the data would be to assume that the perceptual space is uniform everywhere. This would yield a threshold envelope which is a hyperellipsoidal shell. (N.B. this does not mean that its density is the same in all directions, only that on any straight line through the space in any direction the density would be a constant.)

This ad hoc model gives the threshold as

$$c_1X^2 + c_2Y^2 + c_3R^2 + c_4G^2 + c_5B^2 = 1 \ .$$

A simple way to test the model is by finding the values for the coefficients c_i using the thresholds measured empirically along the five display axis (1 0 0 0 0), (0 1 0 0 0 0), (0 0 1 0 0), (0 0 0 1 0), (0 0 0 0 1). Then using these values to see how good a fit is given for the remaining 116 observations from experiment 2 and 52 observations from experiment 3.

The results of this exercise are shown in Figure 19 for the data from experiments 2 and 3. If there were perfect agreement between the model and the data all the data would lie on the oblique line. Looking at these graphs it is apparent that apart from the data points on which the model is based, most of the data lies above the line. This means that the fit is conservative; lower thresholds were obtained than expected. Statistically, these fits are terrible. However, in order to explain why we must first introduce a metric for evaluating the fit.

A Metric for Evaluating Models

To evaluate the models we compared the sum of the squared differences between the observed data points X_i and the predicted data points P_i with the sum of the squared differences between the observed data points and the data mean $\overline{X}$. Here n is the number of data points.

Sum of squared differences from data mean:

$$MSS = \sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

Sum of squared differences from predicted values:

$$PSS = \sum_{i=1}^{n}(X_i - P_i)^2 .$$

Percentage reduction in variance attributable to model:

$$= 100\left(\frac{MSS - PSS}{MSS}\right) .$$

For experiment 2 the model results in a 200% *increase* in the variance over that which would be obtained if a simple mean were used. For experiment 3 a mere 3% improvement is obtained through the use of the model.

Model 1 assumes human colour processing to be a homogeneous space. There are two strong reasons for thinking that this is an unwarranted assumption.

Firstly, there is the issue of form perception. How does the brain decide that a pattern of stimulation is an "object" rather than part of the general variation in light and colour? Clearly, gestalt factors such as proximity of data points and whether or not proximal points have a common colour are important. However, beyond the formulation of rather vague principles, vision science offers little help in modelling when the brain should perceive a cloud of coloured points as one object or two. The data from experiments 2 and 3 show that it is much harder to distinguish two clusters when the points are separated by colour but spatially intermingled and this is undoubtedly a major reason why Model 1 give such a poor account of the data. The brain seems to be able to make a synthesis of colour and spatial information, using colour information more effectively when it is correlated with a spatial separation than when there is no spatial separation.

Secondly, there is the issue of colour processing. Most modern theories of colour propose that colour information passes from the retina to the brain in three more-or-less independent channels [Hurvich1981]. More is known about colour processing as an isolated phenomenon than the interaction of colour and form. Therefore we attempted to apply a channel model of colour processing to give a better account of the data.

6.2. Model 2

For our second model we assume the opponent channels are independent. Thus the colour channel which has the highest signal-to-noise ratio for a particular vector in the colour space is the one that "detects" the cluster, determining the threshold.

There are a number of models of human colour channels available. We arbitrarily chose Guth et. al. [Guth1980]. Given the noise in the data and the essential similarity of the various opponent process models it is reasonable to assume that any opponent channel model would give similar results. The strategy used was to first attempt to model the data obtained in conditions where clusters were separated only by colour and then incorporate this colour processing model into a variation of the hyperellipsoid model given above.

Guth's model is expressed as a function of relative cone excitations.[12] Thus it is necessary to first transform the cluster centres from the RGB monitor coordinates into a coordinate system based on relative cone excitations. This is done in three stages.

First we scale the RGB values to reflect the relative luminances of the phosphors:

$$r = L_R R$$

$$g = L_G G$$

$$b = L_B B \ .$$

Second we transform into CIE XYZ coordinates using:

$$X = 0.62000r + 0.21000g + 0.15000b$$

$$Y = 0.33000r + 0.67500g + 0.06000b$$

$$Z = 0.05000r + 0.11500g + 0.79000b$$

Note that the coefficients in the above set of equations are the chromaticity coordinates of the phosphors as given in section 4.1.

Third we convert to excitations for the three cone types, C_1, C_2 and C_3.[13]

---

[12] Cones are the colour receptors in the eye. There are three classes of cones: short wavelength sensitive, medium wavelength and long wavelength sensitive.

[13] Guth's model is actually expressed in terms of Judd Chromaticity coordinates, not CIE chromaticity coordinates. However, the difference between the two only becomes pronounced for stimuli with a lot of energy at extremely short wavelengths. For the kind of stimuli used in the present experiment the differences are negligible.

$$C_1 = 0.2435X + 0.8524Y - 0.0516Z$$

$$C_2 = -0.3953X + 1.1642Y + 0.0837Z$$

$$C_3 = 0.6225Z$$

Now we have the stimuli expressed in tristimulus values which correspond to relative excitations of the three cone types.

Guth calls his three channels "A" for the achromatic channel, "T" for the red-green channel, and "D" for the yellow-blue channel. To convert to ATD values we use

$$A = 0.5967C_1 + 0.3654C_2$$

$$T = 0.9553C_1 - 1.2836C_2$$

$$D = 0.0483C_3 - 0.0248C_1 .$$

The set of equations given above can be collapsed to yield

$$r = L_R R$$

$$g = L_G G$$

$$b = L_B B$$

$$A = 0.30875r + 0.6306g + 0.056b$$

$$T = 0.2266r - 0.3217g - 0.05362b$$

$$D = -0.009153r - 0.01193g + 0.02259b.$$

The three values, L_R, L_G and L_B have been left separate for reasons which will become apparent later.

To determine how a pair of clusters will affect each of the three opponent channels of this model it is necessary to calculate the cluster separation (we call the the *signal* in what follows) as a ratio with the width of the clusters as they appear to each of the three channel (we call this the *noise*). To obtain the signal produced by a given pair of clusters we pass the cluster centre coordinates through the above set of equations to obtain cluster centre values in terms of A, T and D tristimulus values. if the two clusters are denoted 1 and 2 we obtain a signal on each opponent channel by using

$$S_A = A_1 - A_2$$

$$S_T = T_1 - T_2$$

$$S_D = D_1 - D_2 \ .$$

Since variance is additive, *noise* on each of the opponent channels is given by

$$r = L_R R$$

$$g = L_G G$$

$$b = L_B B$$

$$A_n = 0.30875r + 0.6306g + 0.056b$$

$$T_n = 0.2266r + 0.3217g + 0.05362b$$

$$D_n = 0.009153r + 0.01193g + 0.02259b \ .$$

Recall that for experiments 2 and 3 noise was always given the value 1 on R, B and B.

To predict a threshold for perceiving clusters separated in colour space we select an arbitrary signal-to-noise ratio as the threshold. We define clusters separated by more that this signal-to-noise ratio as visible, and cluster separated by less than this ratio as invisible. Thus, given that the channel with the highest signal/noise ratio defines the threshold, we get

$$\Omega = \max\left\{\frac{S_A}{A_N}, \frac{S_T}{T_N}, \frac{S_D}{D_N}\right\} V_C \ .$$

V_C is the distance between cluster centres required to create discriminable clusters separated in the specified direction. Ω is the threshold signal-to-noise ratio.

To obtain V_C, having specified a threshold, we can use

$$V_C = \min\left\{\frac{A_N}{S_{A,}} \frac{T_N}{S_T}, \frac{D_N}{S_D}\right\} \Omega$$

The results for clusters separated only in colour space are given in Figure 20, compared to data from experiments 2 and 3, respectively. As can be seen, a very poor result is obtained. The model predicts far higher thresholds than were obtained.

To get a reasonable fit to the data some further assumptions are required. One possible reason for the poor data prediction is that the model assumes some specific state of colour adaptation in the subject which may have little bearing on his actual spate of adaptation.

The visual system comes to a state of equilibrium with respect to the overall balance of colour. For this reason the approximately 5:1 ratio between the maximum phosphor luminances $L_{R,}$ L_G and L_B (see section 4.1) may not reflect the relative effect these phospors will have on the visual system once a state of adaptation has been attained. A reasonable alternative assumption is that the human visual system adapts to the colour balance of the monitor. We therefore assume maximum gun luminances which are effectively equal for the purposes of modelling.

Thus, for the second attempt we change the constants to $L_R = L_G = L_B = 1.0$;

The modified model yields the results shown in Figure 21.

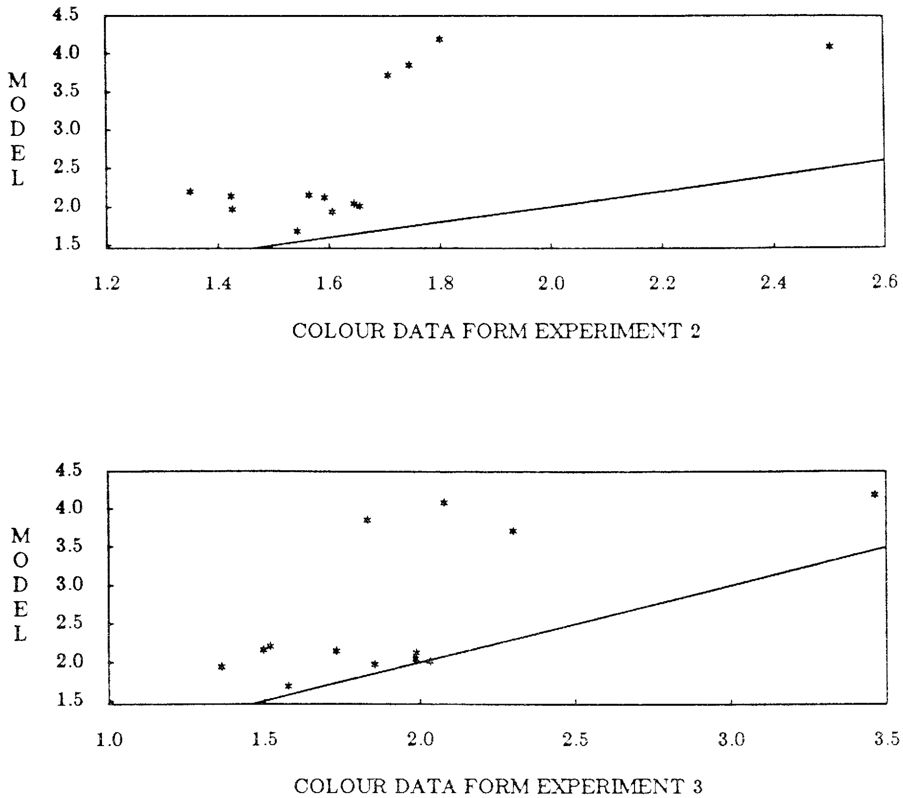


Figure 20. The predictions of the first version of Model 2 are plotted against the data from the conditions in which clusters were separated only in colour.

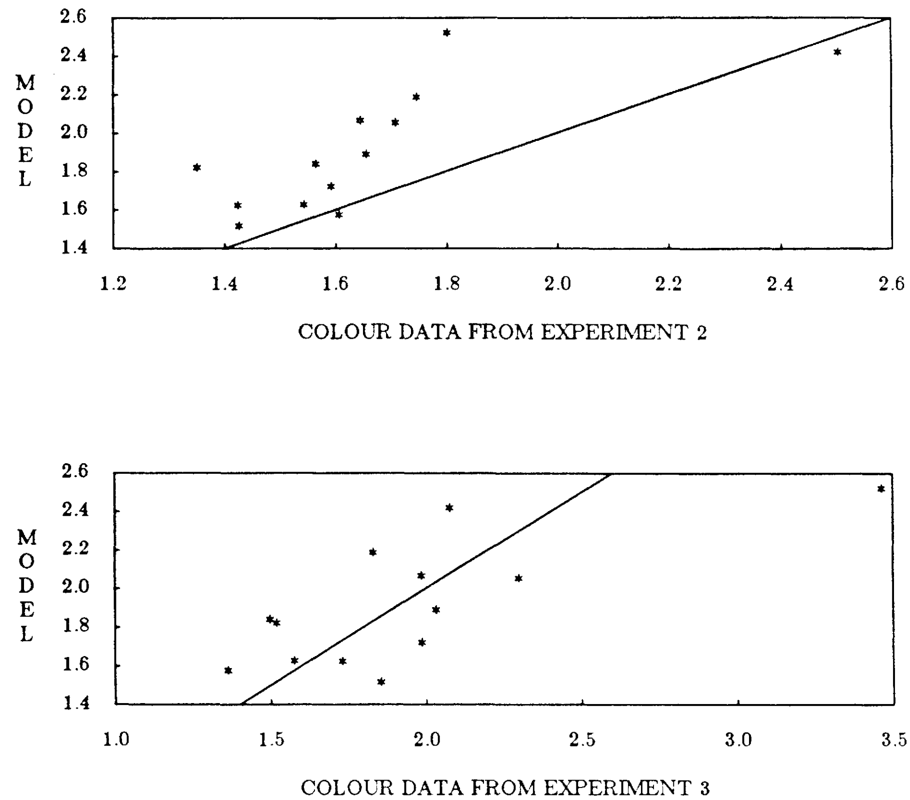


Figure 21. The predictions of the second version of Model 2 are plotted against the data from the conditions in which clusters were separated only in colour.

Now we incorporate this model of colour processing into the hyperellipsoidal model. This involves a two stage calculation. For the first stage we calculate the cluster separation distance (C) required to achieve a signal to noise ration of 1.0 ignoring spatial separation for the moment. This is given by

$$C = \max\left(\frac{N_A}{S_{A,}} \frac{N_T}{S_T}, \frac{N_D}{S_D}\right).$$

The threshold for detecting two clusters as distinct in the five-dimensional space is

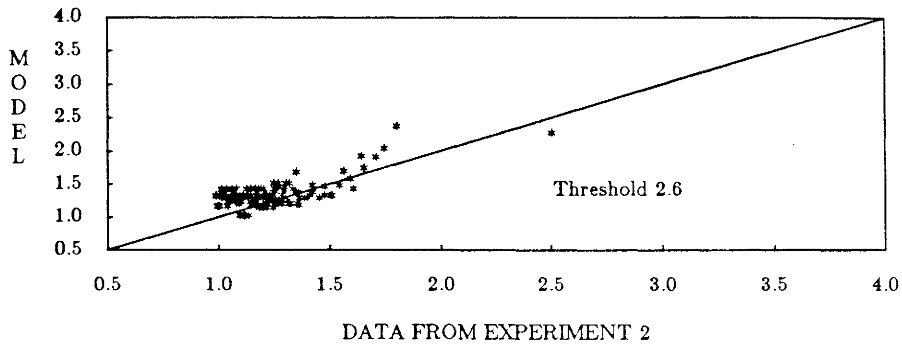
$$\Omega^2 = \left(X^2 + Y^2 + \frac{R^2}{C^2} + \frac{G^2}{C^2} + \frac{B^2}{C^2}\right).$$

Where Ω is a specified signal-to-noise ratio as before. Figure 22 shows the results for this hybrid model, comparing its predictions for the 121 conditions of experiment 2 and the 55 conditions of experiment 3. The value of Ω for experiment 2 was 2.5, while the value for experiment 3 was 3.1. This difference is close to the 22% difference in the thresholds shown in Figure 15. Given these values of Ω, MODEL 2 accounts for 40% of the data from experiment 2 and 57% of the data from experiment 3.

Model 2 is a considerable advance over model 1 , both because it successfully accounts for some of the varience and because it relies on a single variable to account for the differences between the two experiments. This signal to noise ratio is used to account for colour separation and for spatial separation and mixed data.

We could continue making refinements to the model and no doubt with enough tweaking produce an accurate fit to the data. However, because of the ad hoc nature of the exercise, we could have little confidence in the predictive power of the result.

The most important additional modelling factor would be the introduction of some kind of gestalt "grouping factor" which would reduce the threshold for clusters separated on both colour and space. However, to



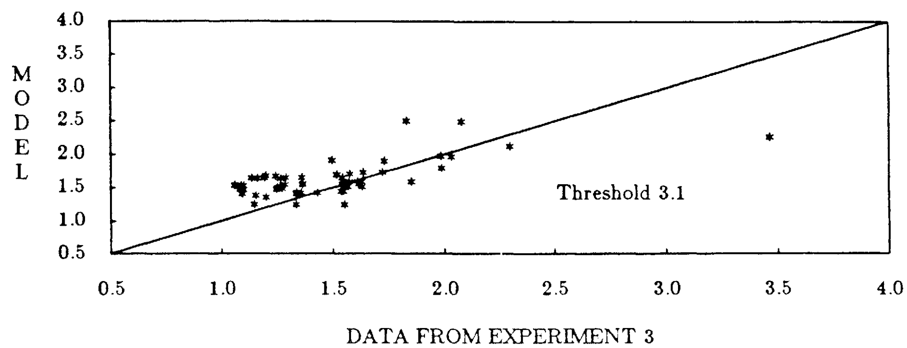


Figure 22. The Predictions of Model 2 are plotted against the 121 data points from Experiment 2 and the 55 data points from Experiment 3. The model threshold was set at 2.6 to fit the Experiment 2 data and 3.1 to fit the Experiment 3 data.

introduce such a factor in a meaningful way would require further empirical investigations of the variation in threshold as a function of the angle of cluster separation. What is required are measurements of the way the threshold changes with the angle between separation in terms of colour only and separation in terms of space only. There may be some regularity about this function; i.e. it may not be necessary to sample for all colours and all directions of spatial separation.

What can we conclude from this modelling exercise? How well have we met the stated goals of being able to predict thresholds for the detection of separate clusters? The answer is that the exercise has not been very satisfying. We are left with the feeling that too many unknown factors having to do with form perception are involved. Perhaps the only thing which has been gained is the feeling that we have some understanding of why certain directions in the RGB colour space result in better discrimination than others.

Overall we feel that the results of this study must stand as valuable mainly because they represent an extensive demonstration of the utility of colour in detecting clusters in a multidimensional space, not because they add anything significant to our theoretical understanding of how humans perceive such patterns.

7 DISCUSSION

What do the results tell us about the utility of using colour to view clusters in a five-dimensional space? In general the results are favourable; in most cases adding colour was like adding three extra spatial dimensions to the display space. Thus colour does give the possibility of perceiving 5 data dimensions simultaneously. However, it is also clear that the perceptual space is not uniform and resolution in some directions is worse than others. Resolution is worst when clusters are separated in colour alone. When clusters are separated by colour and space, each extra colour dimension added almost as much as an extra spatial dimension. Observers were truly perceiving five dimensions.

Concerning the problem of poor resolution for certain directions, a relevant issue is the likelihood of clusters being separated only by colour. Assuming that the direction of cluster separation is equally probable for all directions in the 5D space, in the sampling of the space used in experiment 2 clusters were separated by colour on only 11% of the trials. Separation by both colour and space or by space alone accounted for the remaining 89%.[14] Thus it is only in 11% of the instances where resolution is moderately impaired and only in 2% (separation only on the blue gun or on the red and blue guns) where it is seriously impaired.

Fortunately, there is a simple practical solution to the problem of "bad" directions for cluster separation, which is to use multiple views of the data. With multiple views, clusters which are separated only by colour in a given view will be separated only by space in some other view. Since it not possible in advance to know which views are likely to be most

[14] The conditions in which clusters were separated by colour alone were over-represented in experiment 3 because of the decision to reduce the number of conditions in which clusters were separated by colour and space. This collapsing of conditions meant that the sampling was not uniform as it was in experiment 2.

revealing, the best solution is to make a colour enhanced generalised draftsman's plot. This is a plot of all pairwise combinations of the five data dimensions with some arbitrary mapping of the remaining three dimensions to colour dimensions (an example is given in Figure 3). The generalised draftsman's plot will ensure that any clusters which are separated only by colour in a given view will be separated only by space in some other view. The utility of the generalised draftsman's plot is increased as the number of clusters increases, since with more clusters the possibility of overlap is increased. At this point it may occur to the reader to question the need for colour enhancement. Since the draftsman's plot contains all the information about the data, why bother with colour enhancement? The answer lies in the difficulty of integrating information from the different views of the conventional generalised draftsman's plot. With the colour enhanced generalised draftsman's plot we are not interested in integrating across views - each separate view contains all the information - we are simply increasing our chances of the best possible view. It is not even necessary to present all the views of the generalised draftsman's plot in a matrix. An alternative would be to allow the user to step through them in sequence.

7.1. Generality of the Results:

All of the experimentation of this paper has been limited to examining perception of a single type of data structure, namely the cluster with a hyperellipsoidal probability density distribution. We can only speculate about the extent to which colour will be useful in detecting other data structures. However, our informal observations suggest that colour is likely to be effective in assisting in the perception of correlations in multidimensional space.

7.2. Practical Considerations

Beyond the conclusions which we derived from the empirical results, the time spent playing with various representation schemes using colour gave us a number of insights about what features are desirable in an interactive plotting package using colour.

1) It is essential to have control over the size of the plotted points. As a rule of thumb, if there are a large number of points they should be plotted smaller than a small number of points.

2) Control over the background colour is important. The background forms a visual reference by which all coloured points are judged.

3) The ability to arbitrarily map data dimensions to display dimensions is useful, since certain mappings will enhance certain data features.

4) Colour provides a valuable enhancement to the generalised draftsman's plot.

5) Colour coding enables the complete display of five-dimensional space. However, it can also be used as an enhancement to other schemes to examine more or fewer dimensions. For example, it is well known that simulating the rotation of a three-dimensional cloud of points about an axis which is perpendicular to the line of sight conveys information about the three-dimensional shape of the cloud. This technique could be enhanced by the use of colour to effectively display six data dimensions. For data of a higher dimensionality these techniques can be used to provide a six-dimensional window onto that data. If fewer than five dimensions are required the colour technique can be simply restricted. For example, four-dimensional data can be displayed by holding the signal to the blue gun constant and varying only the four remaining display variables.

7.3. Summary and Conclusions

Five-dimensional data can be usefully displayed on a colour monitor by mapping data dimensions to the following five display dimensions:

Position relative to X-axis.

Position relative to Y-axis.

Amount of light emitted by the red phosphor.

Amount of light emitted by the green phosphor.

Amount of light emitted by the blue phosphor.

1) For the task of resolving clusters, a display of the type defined above gives the effect of having a five-dimensional display space. Adding colour is effectively like adding three additional spatial dimensions. In general, for two clusters to be perceived as distinct they have to be separated by between 3 and 5 standard deviations along most of the possible vectors. However, when clusters are separated on a few specific colour vectors, much greater cluster separation is necessary before two clusters can be resolved. Thus the perceptual space defined by two spatial dimensions and three colour dimensions is highly nonuniform.

2) Little or no training is required to enable subjects to utilise information which is conveyed through the medium of colour. The main effect of experience is to reduce the tendency of observers to overinterpret data, and find structures that do not exist.

3) The benefits of colour increase with the number of clusters present in the data. This can be attributed to the increasing probability of cluster overlap with increasing numbers of clusters.

4) The use of colour to express data dimensions should be combined with other techniques for viewing multidimensional data, such as rotation in real time, or generalised draftsman's plots. This will ensure views of the data which avoid those conditions when cluster resolution is poor, namely when data are separated only by colour.

5) Except for those instances where clusters are separated only by colour, colour enhances the immediate visual impression of a cluster as a visual object.

APPENDIX 1

The C code used to convert from coordinates in opponent colour space to coordinates in RGB colour space is shown below.

```
OPPtoRGB( L, yb, rg, R, G, B)
/* take in OPP 0.0 to 1.0,
   spit out RGB in 0.0 to 1.0
*/

float L, rg, yb;
float *R, *G, *B;
{
     float YB;

     YB = yb*L;
     *B = (1.0 - yb)*L;
     *R = rg*YB;
     *G = (1.0 - rg)*YB;

}
```

APPENDIX 2

The instructions given to the subjects during Experiment 1 are given below and on the following pages.

```
THANK YOU FOR VOLUNTEERING YOUR TIME
FOR THIS EXPERIMENT


Please enter your name and hit RETURN to continue


 : cware
```

page 1

With this experiment we wish to investigate the perception of discrete groups or "clusters" of data points in graphs. The experiment is divided into two discrete phases. In the first you will be shown a number of examples intended to help you identify clusters. In the second you will be asked to determine how many clusters there are in a series of test displays.

The clusters which you will be attempting to identify are present in a multidimensional space which is mapped onto the x and y dimensions you see in front of you. If you find the concept of a higher dimensional space confusing you can regard the monitor as a "window" into a space. Because the space has more than two dimensions clusters may sometime lie in front of one another and therefore be hard to distinguish.

For the first example, the display you see on the monitor contains two very distinct clusters, one at the bottom and one at the right edge, and two not so well separated clusters in the upper left corner, making a total of four clusters.

PRESS 'f' on the keyboard to move FOREWARD to the next page of instructions

page 2

In this example there are two clusters but they are not so discrete. The one in the upper left overlaps with the one in the lower right.

PRESS 'f' for the next page of instructions
PRESS 'b' for the previous page of instructions

page 3

There can be as many as six, or as few as 1 clusters. In this example there are 5 clusters.

PRESS 'f' for the next page of instructions
PRESS 'b' for the previous page of instructions

page 4

In this example there is only one cluster.

PRESS 'f' for the next page of instructions
PRESS 'b' for the previous page of instructions

page 5

Colour may help us to distinguish clusters where they are adjacent. Thus in this example there are two clusters. There is one at left of centre made up of pastel blues, pinks and violets, and one at right of centre made up of greens and browns.

PRESS 'f' for the next page of instructions
PRESS 'b' for the previous page of instructions

page 6

Colour can also enable us to distinguish clusters when one lies on top of another. In this example there are 4 clusters, a blue one, a yellow one, a pink one and a light blue-green one. The pink cluster and the light blue-green are both at the lower right. They can only be distinguished by their colour.

PRESS 'f' for the next page of instructions
PRESS 'b' for the previous page of instructions

page 7

In this example there are 5 clusters, a blue one, a green one, a hot pink one, a beige one and one made up of dark colours.

In the following test phase of this experiment you will be shown a sequence of graphs containing between 1 and 6 clusters. Your task is to try to estimate how many clusters there are in each graph and enter that number as a single integer followed by RETURN to move on to the next display. If you wish you can turn back the pages and review the examples by hitting 'b' before continuing with the experiment.

PRESS 'f' for the next page of instructions
PRESS 'b' for the previous page of instructions

```
page 8

There are a total of 40 graphs.
First you will get 20 trial graphs with colour then
20 trials without colour.
You should try to estimate the number of clusters in
each graph, enter that integer and hit RETURN


Press RETURN to begin
```

On all successive pages the subject received the prompt:

```
How Many:
```

References

Beatty1983.

Beatty, J. C., Raster Graphics and Color, *American Statistician* **37** pp. 60-75 (1983).

Box1958.

Box, G. E. P. and Muller, M. E., A Note on the Generation of Random Normal Deviates, *Ann. Math. Stat.* **29** pp. 610-611 (1958).

Boynton1979.

Boynton, R. M., *Human Color Vision,* Holt, Rinehart, and Winston (1979).

Chambers1983.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P.A., *Graphical Methods for Data Analysis,* Duxbury Press: Boston. (1983).

Chernoff1973.

Chernoff, H., The use of Faces to Represent Points in k-Dimensional Space Graphically, *Journal of the American Statistical Association* **68** pp. 361-368 (1973).

Cleavland1983.

Cleavland, W. S. and McGill, R., A Color-Caused Optical Illusion on a Statistical Graph, *American Statistician* **37** pp. 101-105 (1983).

Cowan1983.

Cowan, W. B., An Inexpensive Scheme for Calibration of a Colour Monitor in Terms of CIE Standard Coordinates, *Computer Grahics* **17** pp. 315-321 (1983).

Cowan1984.

Cowan, W. B. and Ware, C., *Colour Perception*, Notes for SIGGRAPH'84 tutorial. 1984.

Gentleman1983.

Gentleman, J. F., Graphical Representation, Computer Aided, in *Kotz-Johnson: Encyclopedia of Statistical Sciences, Volume 3.*, Wiley, New York. (1983).

Guth1980.

Guth, S. I., Massof, R. W., and Benzschawel, T., A Vector model for normal and dichromatic color vision, *J. Opt. Soc. Amer.* **70** pp. 197-212. (1980).

Hake1966.

Hake,H. W. and Rodwan, A. S., Perception and Recognition, in *Experimental Methods and Instrumentation in Psychology*, ed. J. B. Sidowski, McGraw Hill (1966).

Hurvich1981.

Hurvich, L. M., *Color Vision,* Sinaur, Sunderland, Mass (1981).

Ichikawa1978.

Ichikawa, H., Hakami, K., Tanabe, S., and Kawakami, G., *Standard Pseudoisochromatic Plates,* Igaku-Shoin (1978).

Kahneman1981.

Kahneman,D. and Henik, A., Perceptual Organization and Attention, in *Perceptual Organization*, ed. J. R. Pomeranz, Erlbaum (1981).

Meyer1981.

Meyer, M. A., Human Understanding of Two-Variable Color Maps, *American Statististician* **35** pp. 56-57 (1981).

Meyers1981.

Meyers, G. W. and Greenberg, D. P., Perceptual Color Spaces for Computer Graphics, *Computer Graphics* **14** pp. 254-261 (1981).

Naiman1985.

Naiman, A., Colour Spaces and Colour Contrast, *Graphics Interface '85*, pp. 313-320 (1985).

Pike1965.

Pike, M. C. and Hill, I. D., Pseudo Random Numbers, *Communications of the ACM* **8** pp. 605-606 (1965).

Sibert1980.

Sibert, J. L., Continuous-Color Cloropleth Maps, *Geo-Processing* **1** pp. 207-219 (1980).

Smith1978.

Smith, A. R., Color Gamut Transform Pairs, *Computer Graphics* **12** pp. 12-19 (1978).

Stevens1946.

Stevens, S. S., On the Theory of Scales of Measurement, *Science* **103** pp. 677-680 (1946).

Trumbo1981.

Trumbo, B. E., A theory for colouring bivariate statistical maps, *American Statistician* **35** pp. 220-226 (1981).

Wainer1980.

Wainer, H. and Francolini, C. M., An empirical enquiry concerning human understanding of two variable color maps, *American Statistician* **34** pp. 81-93 (1980).

Ware1983.

Ware, C. and Cowan, W. B., Changes in Perceive Color due to Chromatic Interactions in Striped Test Fields, *Vision Research* **11** pp. 1353-1363 (1983).

Williams1966.

Williams, L. G., The effect of Target Specification on Objects Fixated During Visual Search, *Perception and Psychophysics* **1** pp. 315-318 (1966).

Wyszecki1982.

Wyszecki, G. and Stiles, W. S., *Color Science: Concepts and Methods, Quantitative Data and Formulae,* Wiley (1982).