Reprinted from C. P. Gibson "Colour coded vs Monochrome Electronic Displays",
Her Majesty's Stationery Office, London, 1984

REACTION TIME AS A MEASURE OF THE DISCRIMINABILITY OF LARGE COLOUR DIFFERENCES

C. R. Cavonius
Inst. f. Arbeitsphysiologie
University of Dortmund

J. D. Mollon
Dept. of Exper. Psychology
University of Cambridge

ABSTRACT

The discriminability of colours along two axes in chromaticity space has been studied by measuring the time that a normal observer requires to decide whether two colours are identical or different. By means of a multidimensional scaling program, these data were used to generate a space in which the distance between pairs of colours increases as the response time decreases. The results of this analysis suggest that colours that lie along straight lines in the C.I.E.-(u',v') diagram lie on curves in discriminability space; and that conventional representations of colour space overestimate the discriminability of red from green.

When targets are of small area and brief duration, reaction times are particularly long for colours that lie along a tritanopic confusion line (roughly, a line from yellow to violet). Combinations of these colours should therefore be avoided when choosing colours for colour-coding applications.

INTRODUCTION

It is clear that the use of colour displays will increase in the future. If colour is to be more than a novelty, and is to be used to improve the transfer of information, it is imperative to know how discriminable one colour is from another. Contrary to the impression one may obtain from the literature in this area, we know very little about the discriminability of dissimilar colours.

We are not concerned here with thresholds for the discrimination of colours: first, because these have been studied in detail by Wright (1) and MacAdam (2); and second, because these data are irrelevant, since no one would attempt to colour-code information by using colours that can barely be discriminated from one another. Attempts have been made to predict the perceived differences between pairs of colours that are more widely separated in colour space, by counting the number of steps of just-noticeable difference (*jnd's*) between two colours. Whether this is a defensible method is a question that has a long history in experimental psychology. A related approach is to select a series of colours so that they form a series of small, perceptually equal, steps along a dimension such as hue, saturation, or lightness: the Munsell colour system, for one example, was constructed in this way. Useful though such a system may be in practical colour specification, it shares several serious problems with the jnd method if it is used to predict the discriminability of colours that are very different from one another: we don't know whether steps of hue are equivalent to steps of lightness and steps of saturation; nor whether steps along, say, a red-green axis are equivalent to steps along a blue-yellow axis. And we do not know how steps on different dimensions are to be added.

A seemingly more direct method is that of scaling the perceived difference between pairs of clearly-discriminable colours. This can be done by presenting the observer with pairs of colours, and asking him to assign to each pair a number that describes their dissimilarity (3, 4); or by presenting the colours three at a time, and asking the observer to arrange them on a board in such a way that colours that are more dissimilar are farther apart (5). This begins to look more like what we want, since judgements may be made about colours that are spaced a useful distance apart in colour space, and since a general judgement of dissimilarity is asked for, rather than a judgement along a particular sensory dimension. However, even if we accept the validity of scaling methods, we do not know whether colours that are judged to be very dissimilar are also highly discriminable.

Since none of these earlier studies offers a direct measure of discriminability (as opposed to "dissimilarity"), we have initiated a series of experiments, using a method that appears to be more relevant to the problems of discriminability in visual displays. Our measure of discriminability is the length of time that an observer needs in order to decide whether two colours are the same or different. This is a more direct approach than the methods described above, since the observer is not asked about the way in which the colours differ; he need only decide whether they differ. It should be noted that this is a very different task from that of simply responding to the presence of a colour: there is ample evidence that in the latter case reaction time is the same for all colours, as long as they are of equal luminance (6,7,8).

Although the method is a neglected one, the use of disjunctive reaction time to estimate supra-threshold perceptual differences has a long history. It was proposed as early as 1902 by J. M. Cattell (9) and in 1906 his student, Henmon (10), showed that disjunctive reaction times to pairs of paint samples in an orange-red series increased as the proportions of red and orange in the mixtures became more similar. It is not, of course, supposed that reaction times decrease linearly with the perceptual difference between two stimuli, only that they decrease monotonically with the difference. For an engineer concerned with the design of a colour-coding system, data obtained from reaction times have an immediate validity; for the designer wishes to optimize the speed of correct response.

METHOD

Stimuli were generated on a Barco colour monitor (type HIREM). Two conditions were used, one of which was intended to be relatively easy for the observer, in that large colour targets were presented for an unlimited viewing time; in the other the targets were small, and flashed briefly. The large targets were 1 deg x 1 deg squares, whose nearest edges were 1 deg to the right and left of a fixation mark; they remained on the screen until the observer responded. The small targets were 10 min x 10 min squares, the nearest edges of which were 5 min from the fixation mark; they were presented during a single sweep of the monitor. If presented continuously, the stimuli had a luminance of 15 $cd.m^{-2}$; the luminance of the brief presentations could not be measured, but the peak luminance was presumably the same as that during continuous presentation. All stimuli were presented in a neutral 1.0 $cd.m^{-2}$, 9 deg x 12 deg, background, which resembled the average background one might find in a display. When presented, the stimuli replaced the background in the region that they covered.

Two sets of stimuli were used, in separate experiments. Both were selected to lie along straight lines in chromaticity space: one set (7 colours) along a tritanopic[1] confusion line; and the other (8 colours) along a protanopic[1] confusion line (Fig 1). Within each set, the stimuli were spaced at equal intervals in the CIE-(u',v') colour diagram, which has been proposed as an approximation to a uniform chromaticity space (UCS): i.e., a space within which equal displacements correspond to perceptually equal colour differences. During an experimental session, each colour within a set was paired with every other colour 3 times, in a random order. In addition, an equal number of stimuli in which both colours were identical were intermixed randomly, so that the total number of presentations in the tritan series was 126, and in the protan series, 168. The following results are based on the means of 5 sessions, or 15 responses to each pair.

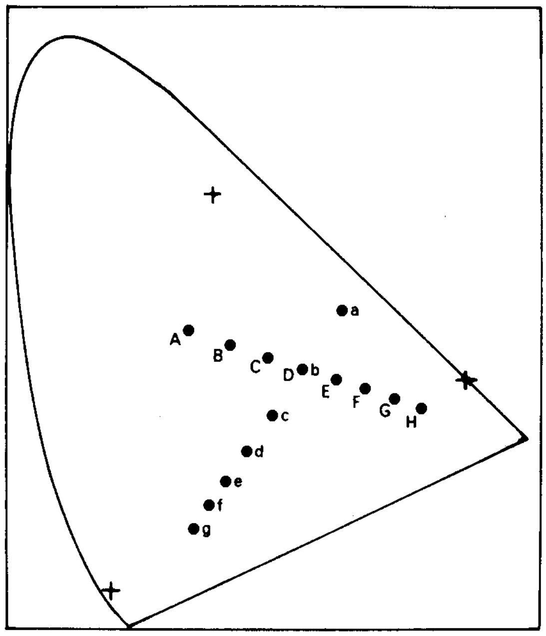

Fig 1. Location of the stimulus colours in the CIE-(x,y) diagram. Protan axis: A-H; Tritan axis: a-g. This diagram is shown because it is more familiar; in the (u',v') diagram the stimuli along each axis are equally spaced. Crosses: location of the CRT phosphors (manufacturer's specifications).

The observer had two keys, for "same" and "different" responses. His task was to respond as rapidly as possible, but to keep incorrect responses to a minimum. A new pair of colours appeared about 2 sec after his last response, and he was informed of incorrect responses by a buzzer. The identical pairs were included to force the observer to decide whether the stimuli were in fact different; in the following analyses only the responses to dissimilar pairs of colours have been used, the mean reaction times being based on both correct and incorrect responses.

The resulting data were analyzed in two ways. In order to get a general impression of whether the different stimulus conditions resulted in markedly different performance, we took the mean reaction time (RT) to all pairs of colours that were separated by one step in a given stimulus series (i.e. protan or tritan series); all responses separated by two steps; by three steps; and so on. In addition, the numbers of errors for pairs separated by 1,2,3...steps were counted.

The second method of analysis was intended to show whether the stimulus series form two sets of equally-discriminable steps of colours, as should be the case, since they were selected to lie at equal intervals in the UCS diagram. The basic assumption is that the discriminability of two colours is inversely related to the RT to that pair. No assumptions are made as to the nature of this relationship, other than that it is monotonic. The task is therefore to arrange all of the colours in a multidimensional space in such a way that the distance between two colours that give a relatively short RT is greater than the distance between two colours that give a relatively long RT; i.e., the rank-ordering of the distances between pairs of colours should be the inverse of the rank-ordering of their RT's.

Because of the quantity of data involved, this clearly cannot be done manually and we have turned to the class of techniques known as Multidimensional Scaling (MDS). (For a general introduction to MDS methods, see ref 11.) The conventional input data for an MDS program are subjective ratings of stimulus differences (see Introduction). Formally, however, the programs are independent of how the estimates of psychological differences are obtained and thus reaction times can legitimately be substituted for subjective ratings. An MDS analysis of reaction times to discriminate numbers has been reported by Shepard (12). The program that we used is Bell Laboratories' KYST (13). We used KYST because it can analyze non-metric data (i.e. data for which only the rank-order is strictly known). We are thus not obliged to assume a particular mathematical relationship between reaction time and some internal measure of difference; we need assume only that the relationship is monotonic. A second reason for using KYST is that it can be used for the analysis of the data of individual subjects, whereas some MDS methods can deal only with group data.

Fig 2. Results of a non-metric KYST ordering in two dimensions of driving distances between European cities. The outline map was fitted by eye.

Since MDS methods are not common in visual science, it may be useful to give an example of how KYST deals with data that are more familiar to the reader (Fig 2). From the Shell road atlas of Europe, we took the 231 pairs of driving distances between 22 large cities. Using driving, rather than airline, distances introduces a curious non-linearity: whereas the driving distance between nearby cities will generally be similar to the distance in a straight line, driving distances between widely separated cities tend to be exaggerated by the fact that oceans are in the way: the direct distance from Madrid to Athens, for example, is 2330 km, whereas the driving distance is 4030 km. In view of this complication, it is impressive that non-metric MDS produces a recognizable ordering of the cities (Fig. 2).

An unfortunate characteristic of MDS methods is that there exist no objective criteria that will specify with certainty the number of dimensions required in the similarity space that is constructed. One helpful guide is a measure known as 'stress', which represents the poorness of fit between the data and the results of the MDS. If adding a dimension gives a substantial reduction in stress, then it may be worth including the dimension: but if an added dimension brings little improvement, it is not apt to be worth keeping. Another guide is to consider whether the added dimension has a theoretical interpretation: if it does not, it is unlikely to contribute to our understanding of the process that we are studying. We shall see later that the interpretation of dimensions is not always straightforward.

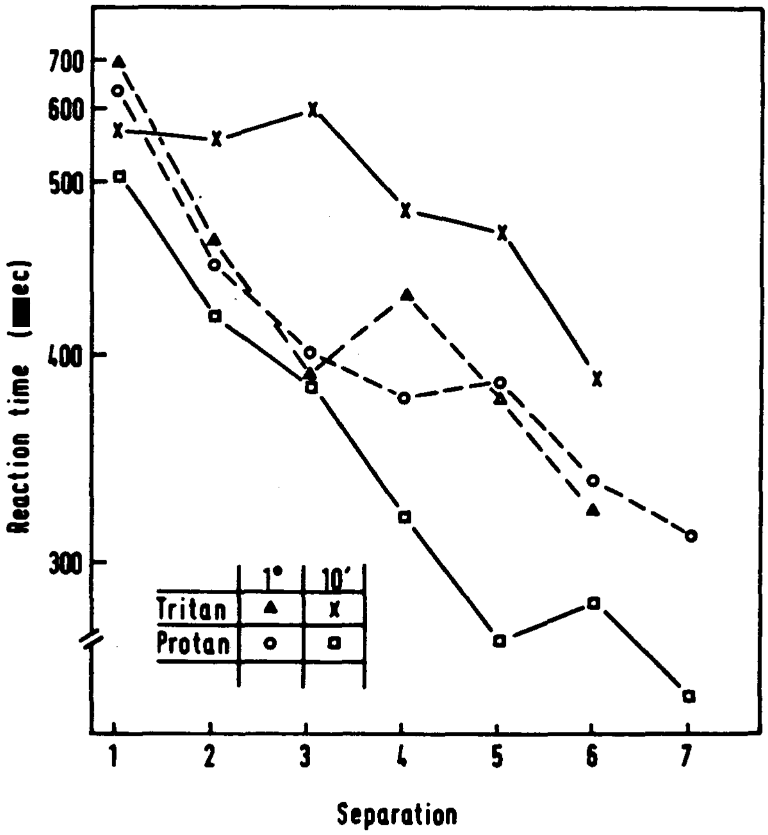

Fig 3. Mean RT's to pairs of colours separated by varying numbers of equal steps in a uniform chromaticity space (CIE-(u',v') diagram). The ordinate is 1/(RT-250 msec), an arbitrary scale that tends to give linear functions.

RESULTS

If we first consider RT's as a function of the separation of the stimuli in UCS space, we see that the protan and tritan series gave almost identical results when large targets were presented (Fig 3). For small, brief presentations the results are quite different: RT's to the protan series are shorter than those for any of the other conditions, whereas RT's for the tritan series are long, and begin to drop only when the separation in chromaticity is large.

Figure 4 shows that the error scores for the four conditions tend to parallel the RT results, again showing that performance along the tritan axis is degraded substantially by small, brief presentations.

However, when we lump together the responses to all stimuli separated by *n* steps, we assume that all such pairs are of similar discriminability and yield similar RT's. If this were true, the MDS analysis of each stimulus series should yield a straight line along which the colours lay at equal intervals. The results for the protan series (Fig 5) do not support this. In the case of both large and small targets, the one-dimensional MDS solution does not show the stimuli in the order in which they plot in the original chromaticity diagram: the two stimuli that are most remote in the chromaticity diagram (stimuli A and H) are not the most remote in the MDS diagram. A better description of the data may be provided by the two-dimensional MDS solution (Fig 5), where the stimuli in both cases lie on a horseshoe, with A and H curved in towards

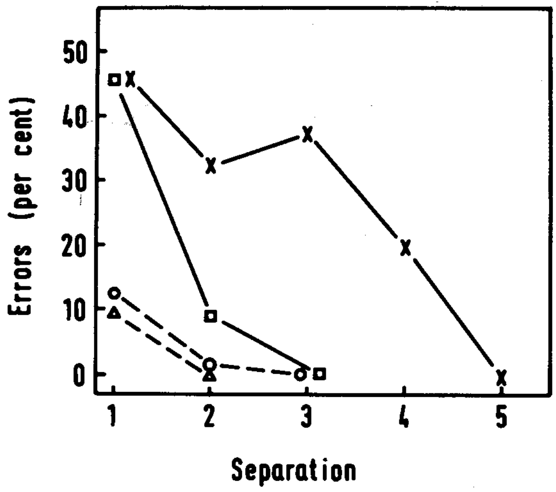

Fig 4. Error rate as a function of the separation of stimulus colours. Symbols as in Figure 3.

B A C D E H G F

B′ C′ A′ D′ E′ F′G′ H′

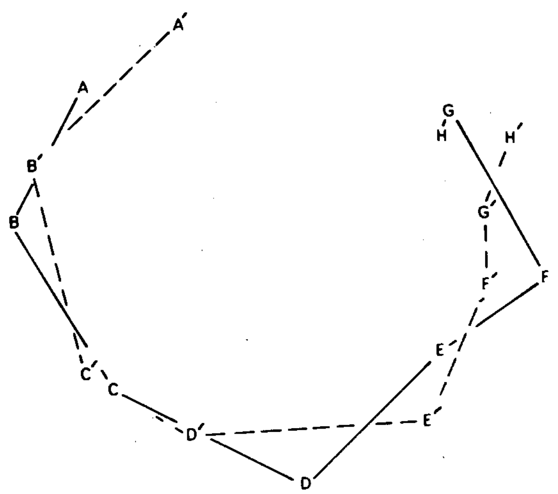

Fig 5. Above: one-dimensional MDS analysis of the large-field protan series (A-H) and the small-field protan series (A'-H').
Below: Two-dimensional analysis of the same data.

g f e d a b c

g′ b′ d′ f′ e′ c′ a′

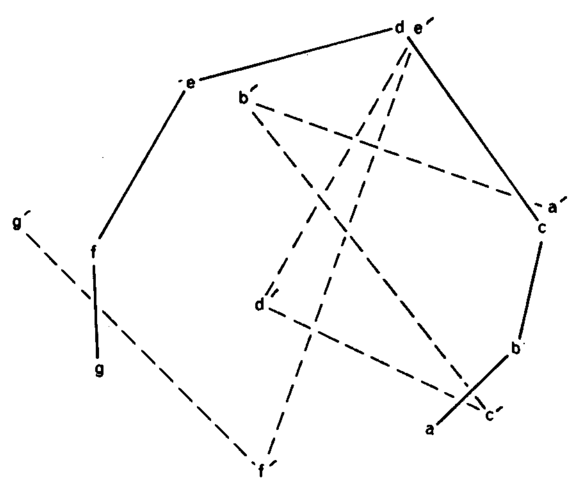

Fig 6. Above: one-dimensional MDS analysis of the large-field tritan series (a-g) and the small-field tritan series (a'-g').
Below: Two-dimensional analysis of the same data.

each other (see Discussion). The solutions for the large and small targets are strikingly similar. (There is no significance in the fact that the two plots are of the same size despite the difference in RT's; this arises because the analysis is in terms of rank order.) The analogous results for the tritan series are shown in Figure 6. The large-target results are similar to those of the protan series: in the one-dimensional solution, the greatest separation does not occur between the two stimuli most remote in the chromaticity diagram (a and g) and in the two-dimensional solution the stimuli lie on a horseshoe with its ends curved in. In the case of the small, brief targets, the MDS solutions for the tritan series dimensions are anarchic; this is probably a consequence of the high error rates that obtained under these conditions.

DISCUSSION

The time required to discriminate between two colours is a practical problem in the colour-coding of visual displays. From the present data, it appears that neither the separation in conventional uniform-colour diagrams nor the judged "difference" between colours is a reliable predictor of their discriminability. In both of the series of colours that we used (which covered a good deal of the range available with conventional colour CRT's), the colours at the far ends of the series are operationally more alike than either end is from the centre of the series. Thus red is more discriminable from yellow than it is from green; and green more discriminable from yellow than from red.

Number of dimensions required to represent the data. There is no absolute way of deciding how many dimensions are appropriate in representing the results of an MDS analysis. Sometimes data that are properly represented in just one dimension may yield semicircular shapes if a two-dimensional solution is allowed (14,15); the bending of the line permits the program to account for more of the random variation in the data. Nevertheless, we tentatively favour the two-dimensional descriptions of our data, as given in Figures 5 and 6. Our reasons are that the two-dimensional maps are similar in form for different conditions (as would be unlikely to be the case if they arose from random variation); that the one-dimensional solution requires an inversion of the physical ordering; and that passing from a one- to a two-dimensional solution reduces 'stress' by about 50% in each case.

Interestingly, the folding of the red-green and yellow-blue axes can also be found in the colour-scaling data of Chang and Carroll (4): although they found that their data could be arranged into something like a conventional colour circle if the solution was constrained to two dimensions, they felt that this gave an inadequate description of the data and they favoured a seven-dimensional solution. In particular, two of their extra dimensions were ones in which red is folded back on green, and blue is folded back on yellow. We have ourselves carried out preliminary experiments with a matrix of 16 colours drawn from a two-dimensional chromaticity space. In a two-dimensional solution, the MDS analysis revealed inversions analogous to those found in the one-dimensional solutions of the present paper: these could be removed if the stimulus surface were folded into a third dimension. However, we should mention that Chang and Carroll's treatment of colour space probably has an insecure foundation. They used an MDS program (INDSCAL) that exploits individual differences in order to construct a stimulus similarity space: and they deliberately included in their sample of subjects a ragbag of different types of colour observer - protanomalous, deuteranomalous, protanopic, deuteranopic, normal, and unclassified. It is most implausible to assume that the perceptual systems of these varied observers differ simply in the weight that they give to certain axes of the normal's colour space; and therefore it is likely that Chang and Carroll overestimate the number of dimensions that characterizes the perceptual colour space of any individual one of their observers. In particular, since each type of anomalous observer has a different set of colour-matching functions, the chromaticity space of the anomal will not coincide with that of the CIE standard observer. It is not that the anomal differs from the normal in giving different weightings to particular dimensions of colour space; rather, the anomal has different dimensions.

Tritanopia for small, brief targets. The results for the small, brief targets of the tritan series suggest that this is a very poor axis along which to select display colours if short presentations are to be used, since these conditions produced slow reactions, a high error rate, and an anarchic similarity space. We had suspected that discrimination would break down under these conditions, since other lines of evidence suggest that the normal observer tends toward tritanopia when the area of the target is reduced ('small-field tritanopia') or when the target becomes very brief ('tachistoscopic tritanopia') (16,17). These phenomena have been reviewed by Mollon (18). When a target is large and of long duration, it seems that integration in space and time compensates for, and conceals, the intrinsic insensitivity of the short-wave mechanism of the eye.

ACKNOWLEDGEMENTS

This work was supported in part by a twinning grant from the European Training Programme in Brain and Behaviour Research. We are grateful to J. J. Chang and Bell Laboratories for making available MDS programs; to H. Dembon and H.-M. Hattenhorst for developing the program for data collection; and to W. Brockmann for adapting the KYST program for this application.

NOTE

1 Colours that lie along a protanopic confusion line are colours that would be indistinguishable by a protanope, that is, a dichromatic observer who lacks the long-wave cones. Colours that lie along a tritanopic line are ones that would be indistinguishable by a tritanope, a dichromatic observer who lacks the short-wave cones. In the case of a normal observer, displacements along a tritan line (provided the colours are of equal luminance) should in theory modulate only the absorptions in the short-wave cones, leaving constant the absorptions in the other two classes of cone.

REFERENCES

1 Wright, W. D. The sensitivity of the eye to small colour differences. *Proc. Phys. Soc. (Lond.)*, **53**, 93-112, 1941

2 MacAdam, D. L. Visual sensitivities to color differences in daylight. *J. opt. Soc. Amer.*, **32**, 247-274, 1942

3 Ekman, G. Dimensions of color vision. *J. Psychol.*, **38**, 467-474, 1954

4 Chang, J. J. & Carroll, J. D. Three are not enough: an INDSCAL analysis suggesting that color space has seven (±1) dimensions. *Color Res. Appl.*, **5**, 193-206, 1980

5 Helm, C. E. A multidimensional ratio scaling analysis of perceived color relations. *J. opt. Soc. Amer.*, **54**, 256-261, 1964

6 Holmes, J. L. Reaction time to photometrically equal chromatic stimuli. *Amer. J. Psychol.*, **37**, 414-417, 1926

7 Pollack, J. D. Reaction time to different wavelengths at various luminances. *Perception & Psychophysics*, **3**, 17-24, 1968

8 Lit, A., Young, R. H., and Shaffer, M. Simple reaction time as a function of luminance for various wavelengths. *Perception & Psychophysics*, **10**, 398-399, 1971

9 Cattell, J. M. The time of perception as a measure of differences in intensity. *Philos. Studien*, **19**, 63-68, 1902

10 Henmon, V. A. C. The time of perception as a measure of differences in sensation. *Arch. Philos. Psych. sci. Meth*, No. 8, 1906.

11 Schiffman, S. S., Reynolds, M. L. and Young, F. W. *Introduction to Multidimensional Scaling*, New York: Academic Press, 1981

12 Shepard, R. N., Kilpatrick, D. W. & Cunningham, J. P. The internal representation of numbers. *Cognitive Psychol.*, **7**, 82-138, 1975

13 Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27, 1964

14 Shepard, R. N. Representation of structure in similarity data: problems and prospects. *Psychometrika*, **39**, 373-421, 1974

15 Rodieck, R. W. Metric of color borders. *Science*, **197**, 1195-1196, 1977

16 Brindley, G. S. The summation areas of human colour-receptive mechanisms at increment threshold. *J. Physiol.*, Lond., 124, 400-408, 1954

17 Farnsworth, D. Tritanomalous vision as a threshold function. *Die Farbe*, 4, 185-197, 1955

18 Mollon, J. D. A taxonomy of tritanopias. *Doc. Ophthal. Proc. Ser.*, 33, (*Colour Vision Deficiencies VI*, ed. G. Verriest, The Hague: W. Junk), pp 87-101, 1982.