Perception, 1981, volume 10, pages 651-666

The aesthetics of colour†

I C McManus¶, Amanda L Jones§, Jill Cottrell#
The Psychological Laboratory, Downing Street, Cambridge, CB2 3EB
Received 27 October 1980, in revised form 16 July 1981

Abstract. A paired comparison experiment is described in which preferences for Munsell colour patches were elicited from 54 subjects. Value, chroma, and hue preferences were examined separately. Individuals differed in their preferences for value and chroma, and also for hue. There were also sex differences in preference. Overall the majority of subjects showed a preference for blue hues and a dislike for yellow hues, thus demonstrating a continuity between animal and human colour aesthetics.

> "All men, completely organised and justly tempered, enjoy colour; it is meant for the perpetual comfort and delight of the human heart ..."
>
> Ruskin, *Modern Painters*

1 Introduction

The vast literature on colour preferences is bewildering, confused, and contradictory. Eysenck, in 1941, summarised previous work which had analysed the preferences of 21060 subjects. Burnham et al (1963) and Osgood et al (1975) have also summarised previous work. Despite the heroic numbers of subjects involved in such work we would like to suggest that, with a few notable exceptions, the results are generally worthless, since adequate accounts of the colours are not given, and in general the number of colours used is too restricted to be of much generality. There are three clear exceptions to the above criticism, but these studies also have problems which complicate their interpretation. Granger (1955) reported the results of a study in which 50 subjects rank ordered a large number of Munsell colours for preference. Separate account was taken of the three colour dimensions of hue, value, and chroma, and differences between subjects were looked for. In general hues were preferred in the order blue, green, purple, red, yellow. Subjects showed significant degrees of similarity in their rankings. Guilford and Smith (1959) asked 40 subjects to make rating judgments of Munsell colour chips, the chips being representative of the three Munsell dimensions. Guilford and Smith found 'considerable consistency' between their subjects. Hues were preferred in the order blue, green, purple, red, yellow. Helson and Lansford (1970) also used a large array of Munsell colour chips, and took account, in some detail, of variations in preference due to background and illumination. However, they tested only 10 subjects, and the judgments were nine-point ratings of the colour patches. No mention was made of the agreement or otherwise of subjects. Hues were preferred in the order blue, green, red, purple, yellow.

†A preliminary version of this work was presented at a meeting of The Colour Group, at the Institute of Ophthalmology, on 9 January 1980, a brief report of which has been presented elsewhere (Mollon 1980).

¶Present addresses: Department of Psychology, Bedford College, Regent's Park, London, NW1 4NS, *and* Department of Psychiatry, St Mary's Hospital, Harrow Road, London, W9.
§Present address: 14 Montagu Drive, Leeds, L58 2PD, West Yorkshire.
#Present address: Willow Chase, Manor Way, Oxshott, Surrey, KT2 2OHU.

In the present paper we will describe an experiment in which we have attempted to improve upon previous experimental designs, in order to answer several specific questions about colour preferences.

1.1 *Are aesthetic responses to colour constant throughout an experimental session?*
Burnham et al (1963), referencing Crawford and Washburn (1911), claimed that "aesthetic responses to single colours in experimental situations are subject to fatigue or adaptation; significant losses in affective value occur within a few seconds or minutes". This conclusion is in distinct contrast to the work of Humphrey (1972) who examined the preferences of rhesus monkeys for visual stimuli over a period of time, and found that whilst naturalistic photographs did indeed show the kind of aesthetic fatigue postulated by Burnham et al, this was not the case for colours, preferences being maintained at a constant level throughout the experiment. Humphrey thus attributed colour preferences to 'pleasure' rather than 'interest'.

Clearly if preferences *do* change during the course of an experiment it is important that we should be aware of it, and such changes should be explicitly looked for.

1.2 *Are there consistent sex differences in colour preferences?*
Eysenck (1941) pointed out that Dorcus (1926), St. George (1938), and Jastrow (1897) had all found sex differences in colour preference. However, von Allesch (1924) and Garth (1931) found no such differences. Eysenck in his own experiment found only minimal sex differences, which he did not test for significance. In his analysis of the results of previous workers, Eysenck found only minimal sex differences involving the ordering of orange and yellow. Granger (1955) concluded that "there are no marked sex differences in colour preferences". Guilford and Smith (1959) felt that colour preferences differed between men and women, but they did not make it clear whether these differences were in hue, value, or chroma, and they did not test differences for statistical significance. Helson and Lansford (1970) found significant differences in colour preference between men and women, but these were not at constant value and chroma, and seemed to depend primarily on the effects of background and illumination, rather than of hue per se.

The question of sex differences in colour preference is therefore somewhat confused, although most modern workers have tended to conclude that there are no such effects. These conclusions usually confound hue with value and chroma, and are often not tested for statistical significance.

1.3 *Are there differences between individuals in their pattern of colour preferences, and, if so, are these differences a function of hue, value, and/or chroma?*
Historically there has been a large shift in the answer given to these questions. Cohn (1894), von Allesch (1924), and Chandler (1934) felt that colour preferences were almost totally idiosynchratic. Dorcus (1926) concluded: "we must be rather sceptical as to whether there is such a thing as colour preference". Eysenck (1941) examined the intercorrelations between subjects' judgments and concluded that they were significant. Guilford (1940) proclaimed in the title of one of his papers: "There *is* system in colour preferences" (our emphasis). Granger (1952) concluded that "colour preferences are objective in the sense that ... they are to a considerable extent independent of personal taste". Guilford and Smith (1959) tacitly assumed that there was hardly any degree of disagreement amongst their subjects, as similarly did Helson and Lansford (1970). Burnham et al (1963) were more cautious: "Preferences for ... single chromatic patches are masked by large variations among individuals in contradistinction to the rather small variation of judgments made at different times by any particular person".

1.4 *Even if there are no consistent preferences between individuals, is there evidence that an individual's preferences are internally consistent?*
Elsewhere, one of us (McManus 1980) has shown that for aesthetic judgments of simple geometric figures there is only minimal agreement between subjects when a whole population is considered, and yet each individual subject's judgments are internally consistent, and reliable over a period of several years. In order to be able to make such conclusions it is important that a large amount of information is collected from each subject. Neither rank ordering nor rating of stimuli gives sufficient information to enable one to make statements about internal consistency. The best method is undoubtedly the method of complete paired comparisons, a detailed statistical account of which has been given by David (1963).

2 Method

A total of 54 subjects, (27 male, 27 female) the vast majority of whom were undergraduate members of the University of Cambridge, took part in the main experiment. All subjects had normal colour vision as assessed by the Ishihara colour test, and the City University Colour Vision Test (Fletcher 1978). Each subject was tested individually over a period of about one hour, the subject being allowed to work at his own rate. As far as the subject was concerned the experiment consisted of a series of comparisons between the two items of a colour pair. The subject was told that he simply had to say which colour of the two he preferred more, making his judgment on a six-point scale. Subjects were informed that the six points could be regarded as representing a strong preference for the left-hand stimulus, a medium preference for the left-hand stimulus, and a weak preference for the left-hand stimulus, and similarly for the three degrees of preference for the right-hand stimulus. Each subject thus made a single judgment of each pair, each judgment being recorded on a line divided into six parts. Subjects were asked to try and use all three degrees of preference (ie weak, medium, and strong) about an equal number of times throughout the experiment, although it was emphasised that this was only to be a guide, and not to be interpreted strictly. Subjects were told that they must record a preference for each pair, there being no way in which complete indifference could be recorded. The instructions as to what was meant by 'preference' were purposefully kept vague, and no subjects seemed to experience difficulty with this. Subjects were asked to work as quickly as possible, first impressions being emphasised rather than considered, reasoned judgments. Subjects were also reassured that no pair of colours would actually be presented twice, although it might occasionally seem that that was the case. The 256 judgments were made in three consecutive blocks of judgments, a few minutes rest being allowed between the blocks.

The stimuli consisted of Munsell colour chips (matte finish), 23 mm high and 16 mm wide. Each chip was mounted in the middle of a piece of black card, 177 mm high and 113 mm wide, the card having an approximate Munsell specification of N2·6. Pairs of cards were placed on the base of a Macbeth daylight illuminator, and were thus at an angle of 45° to the horizontal. The cards were viewed at a comfortable reading distance in a darkened room. The Macbeth lamp simulates CIE Illuminant C, and its colour temperature is thus approximately 6800 K.

The particular stimuli were chosen subject to several constraints. It was felt desirable that the colours used should encompass most of the range of the colour solid, and also that they should be regularly related one to another so that interpretation of the data would be simplified, in particular it being desired that for some pairs value and chroma would be identical, so that effects of hue alone could be determined. These constraints are, however, mutually exclusive insomuch as the

Munsell colour solid is not symmetric in its three dimensions. Thus, whilst there is a yellow of value and chroma 8·5/12, there is no equivalent blue of the same value and chroma. And yet to use a yellow of the maximum value and chroma attainable for a blue is to present a colour (5Y 7/6) which is far from an 'ideal' yellow for most subjects. We therefore presented two separate groups of colours to each subject.

(i) A series of 'archetypal' colours. These were selected so that they represented an ideal, or typical, yellow, blue, etc, independent of constraints upon value and chroma, these usually being as extreme as possible. The actual Munsell colours used were 5R 5/12, 5YR 6/12, 5Y 8·5/12, 5GY 8/10, 5G 5/8, 5BG 6/8, 5B 7/6, 5PB 4/10, 5P 4/10, 5RP 6/10, N1·75 ('black'), and N 9·5 ('white').

(ii) A series of 'pastel' colours. These colours were selected so that for all the hues used a complete range of values and chromas was available. It was thus possible to look separately at hue, value, and chroma preferences. Owing to the shape of the colour solid, these pastel colours could not have such extreme values and chromas as the archetypal colours. The twenty colours chosen were the value/chroma combinations 9/2, 7/2, 7/6, and 5/6 for the hues 5R, 5Y, 5G, 5B, and 5P. Note that the colour 5B 7/6 thus occurred in both the pastel and the archetypal series. Figure 1 shows the approximate positions of the colours on a chromaticity diagram, based on the figures of Newhall et al (1943).

A total of 31 colours was presented. A problem with the method of paired comparisons is that for a complete design the number of pairs to be presented increases as the square of the number of stimuli; 31 stimuli would thus require 465 pairs to be presented. This was felt to be too many judgments for a subject to make, and thus the experiment was run by having a complete paired comparison design within the archetypal colour series, and within the pastel colour series, but showing no pairs between the series. A total of 256 pairs was therefore shown (66 archetypal, and 190 pastel).

The order in which the pairs were shown was randomised, and a single order of presentation generated within certain constraints:

(i) each colour would appear an equal number of times to right and left;

(ii) no single colour patch would appear in two consecutive pairs;

(iii) the two colour series, pastel and archetypal, were randomly interleaved throughout the experiment;

(iv) each colour would appear an equal number of times during each third of the experiment.

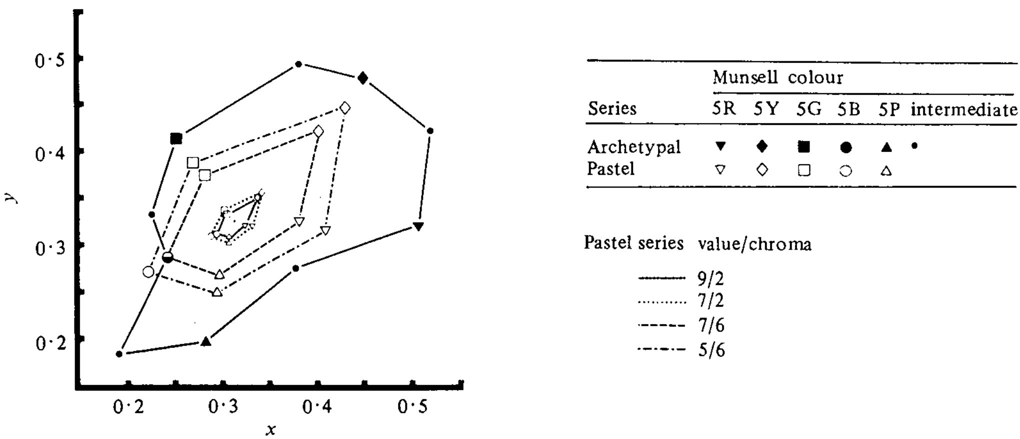

Figure 1. Approximate positions of the stimuli in the archetypal and pastel series on a standard chromaticity diagram.

With these constraints a single order of presentation was produced. This order was then divided into three blocks, containing respectively 85, 85, and 86 pairs, and labelled 1, 2, and 3. Each subject saw the stimuli within a block in the same order, but blocks were presented in different orders to different subjects. Thus in the first third of the experiment, a third of subjects saw block 1, a third saw block 2, and a third saw block 3. The six possible block orders (123, 132, 213, 231, 312, and 321) were each used nine times, thus entailing a total of 54 subjects. In each particular block order there were either 4 men and 5 women, or 5 men and 4 women, subjects being randomly allocated to particular block orders. The reason for using this block design is that during each third of the experiment every possible stimulus pair will be seen an equal number of times across subjects, and thus particular pairs will not be confounded with order of presentation.

3 Results

The results of the experiment are best considered by looking carefully at the manner of representing the results. Let a_{ij} represent an individual's preference for stimulus i with respect to stimulus j. The preference scores can be in two forms. The six-point scale actually used by the subjects was coded as a score of 5 for a strong preference of i over j, 4 for a medium preference for i over j, and 3 for a weak preference for i over j, whilst 2, 1, and 0 represent respectively, a weak, medium, and strong preference for j with respect to i (or dislikes for i with respect to j). Hence, on this scale, $a_{ji} = 5 - a_{ij}$. Whilst there is much to be said for a six-point scale, in particular that subjects feel that it is easier to make difficult judgments if there are several possible degrees of preference, there are many statistical analyses which are more conveniently, or are necessarily, carried out on a binary scale, in which a 1 indicates a preference for i over j, and a 0 represents a preference for j over i. These binary scores can be obtained by recording scores of 0, 1, and 2 as 0, and scores of 3, 4, or 5 as 1. Each subject makes $\frac{1}{2}n(n-1)$ pair-comparison judgments, where n is the number of stimuli being shown. These judgments may be entered into an $n \times n$ matrix, each particular pair being entered twice, once as a_{ij} and once as a_{ji}, the results being 'reversed' for the appropriate one of the two entries. The leading diagonal (ie $i = j$) is filled with zeros. Thus a complete response matrix may be generated for each subject. The response matrices for subjects may be summed by simply adding together the appropriate cells of the individual matrices. Thus if there are N subjects, then cell s_{ij} of the summed binary matrix will represent the number of subjects who preferred stimulus i to stimulus j. Thus $s_{ji} = N - s_{ij}$.

In order to obtain scores for individual stimuli, either for a particular subject or across all subjects, we may sum across a row of the response matrix. Thus $a_{i.}$ or $s_{i.}$ will represent the relative preference for a particular stimulus with respect to all of the other stimuli. Figure 2(i)a shows the summed binary preferences for the archetypal stimuli, whilst figure 2(ii)a shows the summed binary preferences for the pastel series of stimuli. Figure 2(i)a shows that there is a general population preference for achetypal red and blue/purple-blue, with a dislike for green-yellow and for black. These results however confounded hue, value, and chroma. Figure 2(ii)a helps to resolve some aspects of the problem. It can be seen that, for the pastel series of stimuli, preferences and dislikes are greater in magnitude for stimuli of chroma 6 than for those of chroma 2. There is also a tendency for higher values to be preferred to lower values, particularly at chroma 2. There is also a clear liking for blue, and a dislike of red and yellow. Since a primary interest of the present study concerns preferences for hue independently of value and chroma, we can look at a subset of the 20 x 20 pastel matrices, ie the four 5 x 5 matrices in which value and chroma are held constant, (either at 9/2, 7/2, 7/6, or 5/6), and

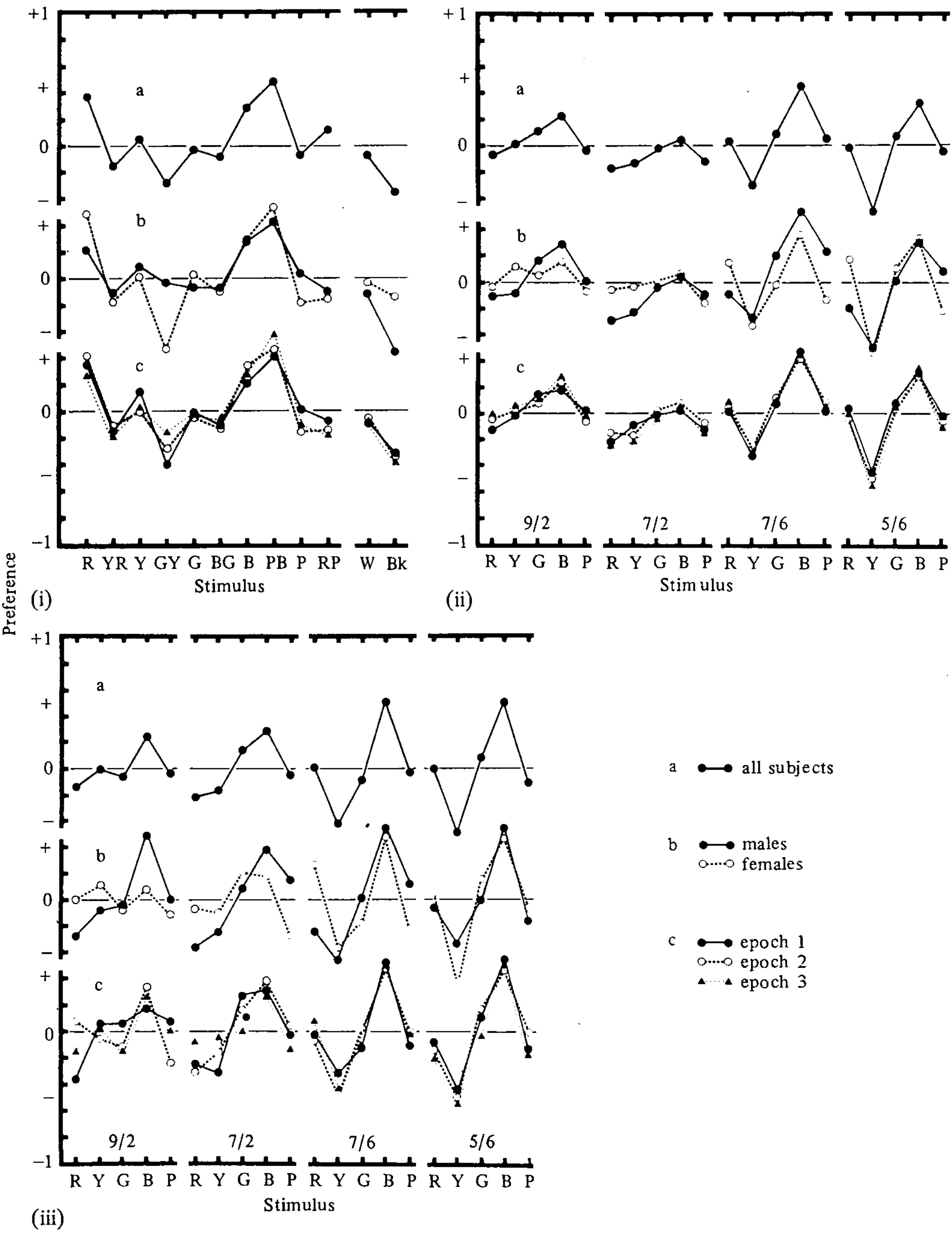

Figure 2. Summed binary responses of all 54 subjects to (i) the 12 archetypal colour stimuli, and (ii) and (iii) the 20 pastel stimuli. The responses summed across all subjects are shown in part a; the summed responses of males and females are shown separately in b; and the summed responses of subjects during the first, second and third epochs (thirds) of the experiment are shown in c. The stimuli are shown on the abscissa, R, Y, G, B, P denoting red, yellow, green, blue, purple respectively, and referring to the Munsell colours described in detail in the text. W refers to Munsell colour N 9·5 ('white') and Bk to Munsell colour N 1·75 ('black'). The ordinate shows the overall degree of preference for a stimulus on a scale from −1 (complete dislike) to +1 (complete agreement and complete preference for a stimulus). All preferences are relative only to the other stimuli shown in the figure: to the other 11 archetypal stimuli in (i), the other 19 pastel stimuli in (ii), and to the other 4 colours within each value/chroma combination in (iii).

compare the five hues (5R, 5Y, 5G, 5B, and 5P). These submatrices may each be analysed separately. Figure 2(iii)a shows the preference score for the hues at each value/chroma combination. As with figure 2(ii)a it is apparent that there is a preference for blue, and a dislike of red-yellow. There are many statistical analyses that can be carried out with these summed preference matrices. The simplest and most important is to ask whether the results could have arisen by chance; that is, in general are all $a_{i.}$ equal (or, to put it a slightly different way, within the summed preference matrices are all a_{ij} equal to all a_{ji}, within the limits of sampling error). The best statistic for testing this hypothesis is the coefficient of agreement, u, as defined by Kendall and Babington Smith (1940), and analysed in more detail by David (1963, p 25). u can take a maximum value of 1, when all subjects are in total agreement, and may have a minimum value of $-1/N$ if N is odd, or $-1/(N-1)$ if N is even. Clearly, whilst all subjects may agree completely, they are unable all to disagree completely, and hence $u_{\min} \neq -1$ (except when $N = 2$). For the archetypal colours of figure 2(i)a, $u = 0{\cdot}105$ ($\chi^2 = 451{\cdot}0$, 69 df, $p < 0{\cdot}001$), and for the pastel colours of figure 2(ii)a, $u = 0{\cdot}072$ ($\chi^2 = 958{\cdot}4$, 201 df, $p < 0{\cdot}001$). For the constant value/chroma results of figure 2(iii)a, the u values for 9/2, 7/2, 7/6, and 5/6 are respectively 0·018, 0·043, 0·143, and 0·153, all being statistically significant ($\chi^2 = 20{\cdot}32$, 10 df, $p = 0{\cdot}026$; $\chi^2 = 34{\cdot}08$, 10 df, $p = 0{\cdot}0002$; $\chi^2 = 89{\cdot}23$, 10 df, $p < 0{\cdot}001$; $\chi^2 = 95{\cdot}08$, 10 df, $p < 0{\cdot}001$). It seems clear from these values of u that high-chroma colours are easier to judge than low-chroma colours; nevertheless at all value/chroma levels there are consistent hue preferences.

In the introduction it was suggested that an important question concerns whether colour preferences change throughout the course of an experiment. Figure 2(i)c shows the summed binary matrices obtained during the first, second, and third epoch (third) of the experiment. For each of these binary matrices, $s_{ij}+s_{ji} = 18$, although, unlike the total matrices discussed earlier, the 18 subjects contributing to one particular cell will not necessarily be the same 18 as those contributing to another cell. Nevertheless the results may be treated as such for our present purpose. For convenience only epochs 1 and 3 will be analysed, since it is there that we would expect to find the greatest differences. Let s_{ji1} be the number of subjects who preferred i to j during epoch 1; hence $s_{ji1} = 18 - s_{ij1}$. If preferences change with time we would expect that, on average, s_{ij1} will not equal s_{ij3}. This possibility may readily be tested since all s_{ij} (for $i > j$) are independent, and all s_{ij1} are independent of all s_{ij3}. Hence for a particular value of i and j ($i > j$) we may carry out a chi-squared test for s_{ij1} and s_{ji1} being different from s_{ij3} and s_{ji3}, the data being arranged in a 2 x 2 matrix. These $\frac{1}{2}n(n-1)$ values of χ^2 may be summed, since they are independent, to give an overall estimate of whether two preference matrices differ. For the comparison of epochs 1 and 3 of archetypal colours shown in figure 2(i)c we obtain $\chi^2 = 78{\cdot}87$, which with 66 df, is not significantly different from chance expectations. Similarly for the difference between epochs 1 and 3 of figure 2(ii)c we obtain $\chi^2 = 203{\cdot}8$, which with 190 df is not significantly different from chance. These two values for archetypals and pastels may also be combined to give an overall test of the difference between epochs; this gives $\chi^2 = 282{\cdot}69$, which with 256 df is not significant. For completeness, the respective values for the four value/chroma combinations of figure 2(iii)c are $\chi^2 = 19{\cdot}54$, 10 df, $p < 0{\cdot}05$; $\chi^2 = 6{\cdot}61$, 10 df, ns; $\chi^2 = 5{\cdot}78$, 10 df, ns; and $\chi^2 = 6{\cdot}19$, 10 df, ns. The just significant value for the 9/2 hues may be regarded as a type I error, since the combined results for all four combinations give a value of 38·03, which with 40 df is not significant.

In summary there is no evidence that aesthetic responses to colour change throughout the course of an hour-long experiment, and this factor may be disregarded as a source of error in such experiments.

In the introduction it was suggested that previous work on sex differences in colour preferences was somewhat equivocal. Figure 2(i)b shows the preference scores for the archetypal stimuli by males and females separately. There is a tendency for females, but not males, to dislike strongly green-yellow, whilst males, but not females, strongly disliked black. These differences may be tested for significance in exactly the same way as the epochal differences described in the previous paragraphs; that is, by comparing each pair of cells in the male summed matrix with the corresponding pair of cells in the female summed matrix. For the data of figure 2(i)b this gives a χ^2 of 128·33, which with 66 df is highly significant ($p < 0{\cdot}001$). The coefficients of agreement, u, for males and females are 0·088 and 0·148 respectively, both of which are also highly significantly different from chance. Figure 2(ii)b shows data for pastel colours divided by subjects' sex: overall the preference matrices are significantly different ($\chi^2 = 277{\cdot}8$, 190 df, $p < 0{\cdot}001$). The coefficients of agreement for males and females are 0·099 and 0·063 respectively, both of which are highly significantly different from chance expectations. Scrutiny of figure 2(ii)b suggests that hue preferences in females are far less obvious than for males, for the 9/2 and 7/2 groups, and that there is less of an effect of high value. Figure 2(iii)b examines the sex differences in hue preference for hues at constant value and chroma. For 9/2, 7/2, 7/6, and 5/6 the χ^2 values for differences between the sexes are 16·01 ($p < 0{\cdot}10$), 18·24 ($p < 0{\cdot}10$), 22·50 ($p < 0{\cdot}020$), and 24·29 ($p < 0{\cdot}01$) respectively (each with 10 df). The combined χ^2 for the 5 x 5 matrices is 81·05, which with 40 df is highly significant ($p < 0{\cdot}001$). The coefficients of agreement also show sex differences. Thus at the 9/2 level males have a u of 0·077 ($p = 0{\cdot}146$), whilst females have a u of −0·019 ($p = 0{\cdot}991$, $u_{\min} = -0{\cdot}037$). At the 7/2 level the u values are 0·094 ($p = 0{\cdot}077$) and 0·021 ($p = 0{\cdot}68$), at the 7/6 level the u values are 0·183 ($p = 0{\cdot}0015$) and 0·149 ($p = 0{\cdot}0077$), and at the 5/6 level the u values are 0·126 ($p = 0{\cdot}021$) and 0·183 ($p = 0{\cdot}0016$) for males and females respectively. There is thus some evidence that males have relatively stronger (or at least more consistent) preferences at low chroma, whilst females have relatively stronger preferences at high chromas. These sex differences will be returned to later. With the present method of analysis it seems indisputable that there are sex differences; thus combining the data from pastels and from archetypal colours we obtain a χ^2 of 406·20, which with 256 df is equivalent to a z-score of 5·89 units ($p \ll 0{\cdot}001$).

Thus far the analysis has considered only population preferences, and these have been shown to be significantly different from chance expectations. However, population results often conceal many individual variations, and, in particular in aesthetics, may result from a minority of individuals with clear, strong, and consistent preferences, and a majority with either no preferences or inconsistent preferences. It is thus necessary to test each individual subject's results for consistency. This is most easily done by examining triadic interrelations. Let A *pr* B indicate that stimulus A is preferred to stimulus B. If the same subject also states that B *pr* C, then we would hope that if he were being consistent then also A *pr* C. If, however, C *pr* A, then we have what is known as a circular, or intransitive, triad (Kendall and Babington Smith 1940). If we have n stimuli, then there is a total of $\frac{1}{6}n(n-1)(n-2)$ possible triads. Kendall and Babington Smith (1940) have shown that the maximum possible number of circular triads is $\frac{1}{24}(n^3-n)$ if n is odd, or $\frac{1}{24}(n^3-4n)$ if n is even, whilst the minimum number is, of course, zero. In general, in randomly generated matrices there will be a mean of $\frac{1}{24}n(n-1)(n-2)$ circular triads, the distribution having a variance of $\frac{3}{16}n(n-1)(n-2)$ (Moran 1947). Thus for a 20 x 20 matrix there will be a total of 3420 triads, of which between 0 and 330 may be circular; on average there will be 285 (standard deviation 35·8) circular triads in random matrices. Figure 3a shows the actual and expected number of circular triads for the 20 x 20 matrix of pastel colours. It is readily obvious that all except one subject produced far less triads than would be expected by chance alone, the distribution having a mean of 83·3 and a standard deviation of 12·9. In general the results for the archetypal colours were similar and will not be given in detail; the distribution has a mean of 12·09 triads, and a standard deviation of 7·9. Only 3 subjects produced chance levels of circular triads in the archetypal series (one of whom was subject 8, who produced the very high number of triads in the pastel series). The correlation between the number of triads produced in the pastel and in the archetypal series was 0·51 ($n = 54$, $p < 0 \cdot 001$) suggesting individual consistency in the degree of consistency. Figure 3b shows the number of circular triads produced by the subjects for the constant value/chroma pastel colours. In a 5 x 5 matrix, even zero circular triads is not significantly different from chance expectations at the 0·05 level, and therefore individuals cannot be assessed individually. Figure 3b shows the expected proportions of the circular triads for random matrices: it is clear that the observed distribution is significantly different from the expected distribution, there being a tendency for preferences to be clearer and more consistent in high-chroma colours than low-chroma colours. It is worth remembering in looking at these results that the

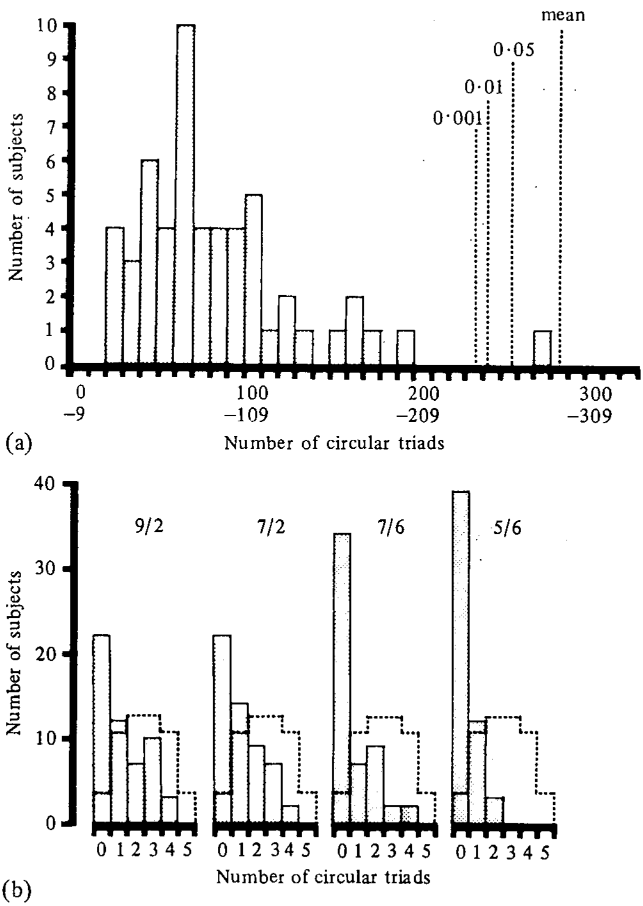

Figure 3. The number of circular triads produced by each subject when making preferences (a) for the complete 20 x 20 matrix of pastel shades and (b) for 5 x 5 matrix of pastel colours at constant levels of value/chroma. The ordinate shows the number of subjects, whilst the abscissa shows the number of circular triads produced by those subjects. The vertical dashed lines in (a) represent the critical number of triads for statistical significance from chance expectations. The dotted line in (b) indicates the expected proportions of subjects under a null hypothesis of random responding.

10 pair comparisons on which each graph is based were scattered randomly throughout an hour-long session containing 256 pairs, and thus memory is unlikely to play an important role in the consistency. That three-quarters of subjects produced no inconsistencies with the 7/6 or the 5/6 colours is therefore a remarkable finding.

Thus far in the analysis we have only considered the subjects' judgments as if they were on a two-point, binary scale (ie A *pr* B or B *pr* A). But, of course, the subjects actually used a six-point scale, and it is therefore of interest to ask whether subjects used the extra scale-points in a meaningful way. Consider a judgment a_{ij}. Let this be scored 1 if the subject responded with a weak preference (from either A or B), a 2 if the subject used a medium preference, and a 3 if the subject used a strong preference (in each case independently of direction of preference). Consider the scores as such which are produced in every possible triad (each judgment therefore being used several times in different triads). We may now average the scores produced if the triad is a part of a circular triad and compare them with the scores produced when the score is a part of a noncircular triad: these are shown in figure 4. Of the 54 subjects, 52 produced weaker preferences when the judgment was a part of a circular triad than if it were a part of a noncircular triad. ($\chi^2 = 46 \cdot 29$, 1 df, $p < 0 \cdot 001$). Most subjects are therefore using the rating scale in a consistent and meaningful manner.

With the exception of subject 8, all subjects individually have significant and consistent colour preferences. It is however possible that subjects have *different* preferences, and yet there would be still an overall population preference. This possibility may be studied by examining, by factor analysis, the relationship between the preferences of different subjects. To do this a 54 x 54 intersubject Pearsonian intercorrelation matrix was constructed, all of the 256 judgments (on a six-point scale) of each subject being correlated with the 256 judgments of every other subject. The principal components of this matrix were found using the PA2 option of the FACTOR program of the SPSS statistical package (Nie et al 1975). Figure 5 shows a graph of the eigenvalues against factor numbers (roots). If all subjects are essentially similar in their preferences we would expect a single significant factor. By using the scree-slope criterion (Child 1970), it would seem that there are at least three independent dimensions in the data; the hypothesised scree-slope has been indicated in figure 5 with a dotted line. The first three factors together accounted for 41·3% of the total variance. (20·7%, 11·6%, and 9·0% respectively). It might be argued that from figure 5 a more satisfactory scree-slope might be constructed starting at factor 8. This point is not disputed, but in the interests of simplicity only the first three factors will be considered in the rest of this paper.

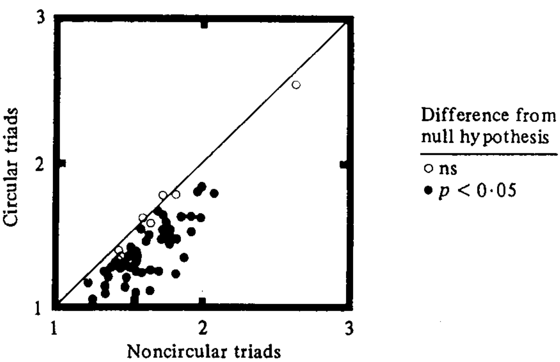

Figure 4. The mean absolute strength of preference made by each of the 54 subjects to pastel stimuli when the responses were part of either a circular or a noncircular triad. The diagonal line represents the null hypothesis that response strengths are similar in the two types of triad.

The first three factors of figure 5 may be reconstructed by multiplying each element of each subject's preference matrix by his loading on a particular factor, and summing across all subjects. Figure 6(i) shows the first three principal factors for archetypal colours. Factor P1 shows clear preferences at red/yellow and purple-blue, with strong dislikes of white and particularly of black. The second factor, P2, shows a preference at red, but not at yellow, and clear preferences at blue and purple-blue, with white and black also being preferred. Yellow-green is particularly disliked, as are purple and red-purple. Factor P3 shows a strong preference at blue and at white, with strong disliking for red, and with moderate disliking for green and black. These factors are difficult to interpret further since they confound hue, value, and chroma, but they probably represent the degree of individual variation for 'typical' colours. Hue, value, and chroma may be disentangled by examining the preferences for pastel colours. Figure 6(ii) shows the three factors P1, P2, and P3, plotted so as to show as clearly as possible the differences between values and chromas. First, consider the differences between 9/2 and 7/2, and between 7/6 and 5/6 ie a two-step difference in value. In all factors, at all hue values, these lines are sloping down from left to right. It thus seems fairly clear that for most subjects high value is preferred to low value. In its extreme form this would mean that white should be preferred to black, as indeed is the case in figure 6(ii) for factors P1 and P3; indeed, overall, 42 of the 54 subjects (77·8%) said that they preferred N 9·5 (white) to N 1·75 (black). Consider next the differences between 7/2 and 7/6 ie a four-step chroma difference. For factor P2 it is clear that for all hues these lines are running steeply up from left to right, suggesting that for this factor high chroma is preferred to low chroma. For factor P3 the converse is the case; the lines joining 7/2 and 7/6 are all running down from left to right, implying that low chroma is preferred to high chroma. Factor P1, the largest factor, is the most difficult to interpret. For blue there is a clear preference for 7/6 over 7/2, whilst for yellow there is a clear preference for 7/2 over 7/6. This must therefore be construed as an interaction between hue and chroma. A question of major importance is whether there are differences between individuals in *hue* preference. Figure 6(ii) suggests that there is in general a dimension of preference from blue-purple down to yellow-red. But, of course, figure 6(ii) inevitably has difficulties of interpretation since value and chroma differences are still present.

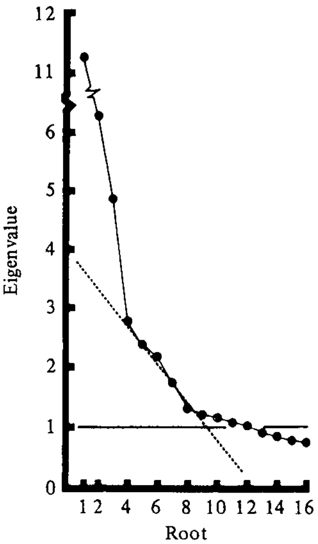

Figure 5. Eigenvalues plotted against roots for all judgments made by subjects. The dotted line indicates the hypothesised 'scree-slope'—see text for details.

Figure 6(iii) shows the analogous reconstruction to figure 2(iii)a, in which only hues at constant value and chroma have been considered. Consider factor P1 in figure 6(iii). At all value/chroma levels it is clear that blue is most preferred, with purple second most preferred, and yellow least preferred, there being no difference between the value/chroma levels. Factor P3 is also relatively simple: at all levels blue and then purple are the most preferred hues. Overall for P3 yellow is probably the most disliked hue, although there are some alterations in order between red and yellow at different levels; there are not, however, clear interactions. For factor P2 there is a clear and important interaction. Once more, blue is overall the most preferred hue, but for this factor purple is particularly disliked at all levels, whilst red shows a strong interaction with chroma, being strongly disliked at chroma 2 and strongly liked at chroma 6. These results suggest that there are probably differences between individuals in their hue preferences as well as in value and chroma preferences. One way of testing this possibility is to factor-analyse the intersubject correlation matrix based upon only those forty comparisons which are made at constant value and chroma. When this is done, by the same method as before, the first ten eigenvalues are 11·42, 7·03, 4·33, 3·51, 2·45, 2·23, 1·70, 1·62, and 1·53. It is clear from this series that certainly the first two factors, and probably the first three factors, are significant by a scree-slope criterion. We must conclude that differences in pure hue preference between subjects are probably multidimensional.

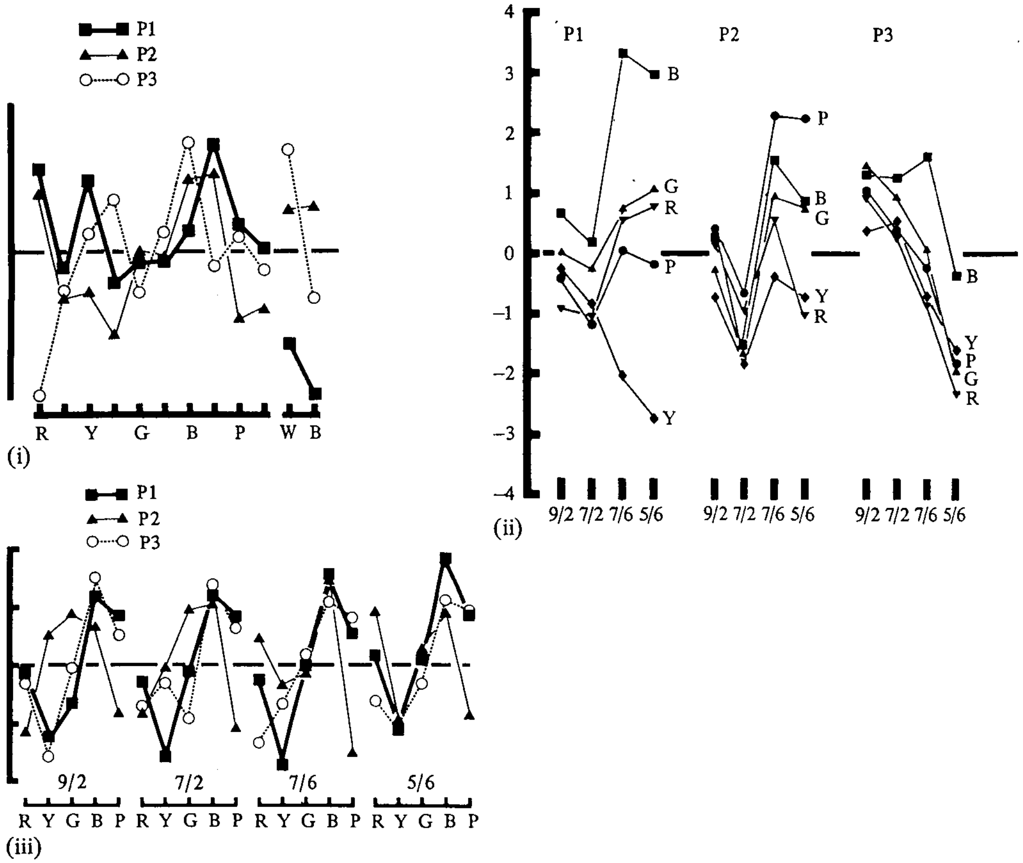

Figure 6. The reconstructed factors P1, P2, and P3 for (i) the archetypal colours shown in figure 2(i)a, (ii) for the pastel colours considered overall [analogous to figure 2(ii)a], and (iii) for the pastel colours at constant levels of value/chroma [analogous to figure 2(iii)a].

It must be realised that in describing three factors we are not, of course, describing three different types of individual, for any individual may load significantly on two or even three of the factors, so that subject's results would actually be a composite of the factors as described. It is also possible, of course, to have *negative* loadings on one or all of the factors. Figure 7 shows the loadings of each of the 54 subjects on the three principal factors. Subject 5 has a high positive loading on factor P1, and no loadings on factors P2 or P3; this is therefore a fairly pure result. Similarly, subject 18 loads only on factor P2, and subject 12 loads only on factor P3. There are no subjects who load negatively on just one factor, with the possible exceptions of subjects 42, 33, and 20. Subject 8, who has been mentioned earlier as having no preferences, does not load on any of the factors. It appears from figure 7 that pure negative loadings (ie the inverse preference functions of figure 6) are probably rare, and are not convincingly shown in the present study. Thus far only principal factors have been described; we have also carried out a Varimax rotation of the factors, but since this did not clarify the results in any obvious way, we will not report the results further.

The lack of any pure negative loadings in figure 7 is of relevance to the question of pure hue preferences. Consider figure 6(iii). At all levels, for all factors, blue is the most preferred colour. In the absence of significant negative factor loadings we might be tempted to conclude that in general all subjects will have either a blue preference or will have no hue preferences as such. However, this result is not secure without examination of the individual preference functions for hues at constant value and chroma. Since we are not interested in value/chroma–hue interactions, we may combine the hue preferences at each of the four constant value/chroma levels, testing these combined 5 × 5 matrices for significance (ie for consistency *within* a single subject) by means of the u statistic (David 1963, p 38). If we combine hue preferences in this way, 47 of the 54 subjects show hue preferences which are significantly different from chance at the 0·05 level or better (14 of these lying between the 0·01 and 0·001 levels, and 22 being at better than the 0·001 level). Only 7 subjects showed nonsignificant hue preferences, and only 2 subjects showed negative values of u for consistency. Figure 8 shows the averaged hue preference (across the four constant value/chroma levels) for each of the subjects. These have

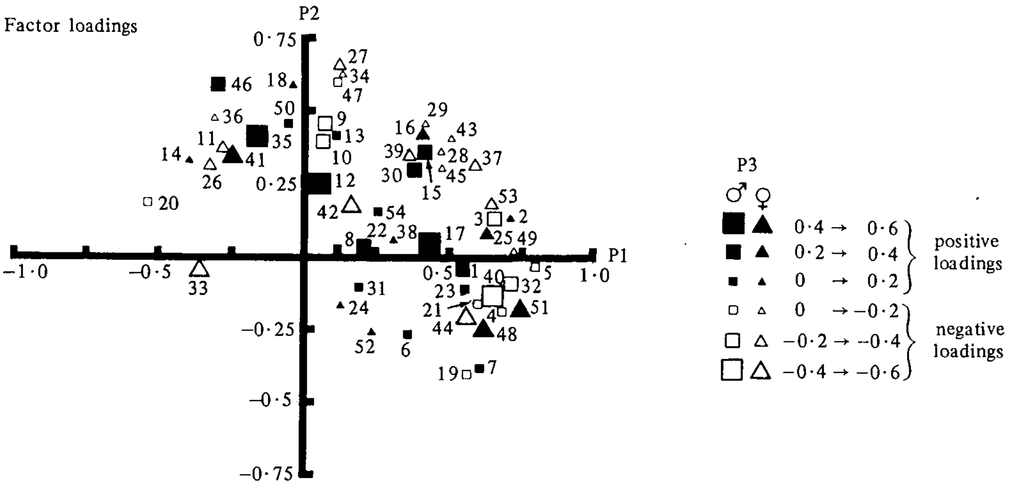

Figure 7. Loading of each subject on factors P1, P2, and P3. Loading on factor P1 is shown along the abscissa, that on factor P2 along the ordinate, and that on factor P3 is shown by the size of the symbol. The numbers alongside the points are the numbers of the subject used in the study (and also shown in figure 8).

been classified, broadly, and by eye, into four groups. The largest group (I: $n = 30$) comprises those individuals who appear to prefer blue and to dislike red-yellow. These individuals are thus similar to the overall population preference of figure 2(ii). A second group (II: $n = 11$) is of those individuals who have a preference for red or for yellow. These individuals are thus the opposite of those in group I, and in the case of subjects such as number 33 are the exact inversion of the type I response. A third group (III: $n = 9$) may broadly be categorised as having preferences for *both* blue *and* yellow, with broad dislikes for red, purple, and green. Finally a miscellaneous group (IV: $n = 4$) consists of individuals with almost flat or nonsignificant preferences, or, in the case of subject 40, a preference function which looks dissimilar to all of the others.

The classification given here is essentially carried out by eye. A more formal cluster analysis could have been carried out, but it was felt that the data were not strong enough to merit it. Nevertheless it seems clear that subjects *do* differ in their hue preferences, a minority having preferences for red and/or yellow. There was a significant tendency for the sexes to be arranged differently in the two groups ($\chi^2 = 6 \cdot 51$, 2 df, $p < 0 \cdot 05$), males representing 16 of the 30 subjects in group I, 2 of the 11 subjects in group II, and 9 of the 13 subjects in groups III and IV combined.

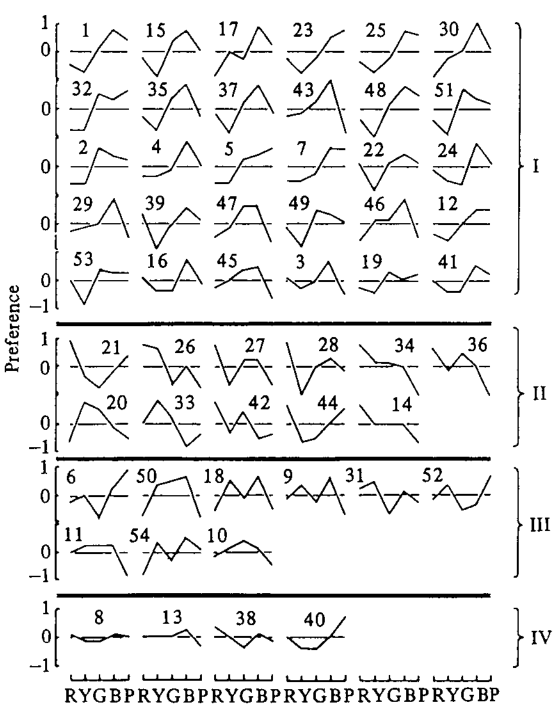

Figure 8. Combined hue preferences of each subject at constant levels of value/chroma (ie based on a total of forty preferences per subject). The subjects have been arbitrarily grouped into four groups (see text for details). Reading from left to right and top to bottom, subjects 1 to 51, 21 to 33, 6, and 50 are significantly different from chance expectations with $p < 0 \cdot 001$. Subjects 2 to 46, 42, 44, and 18 are significant with $p < 0 \cdot 01$, and subjects 12 to 45, 9 to 10, and 40 are significant with $p < 0 \cdot 05$. Only subjects 3, 19, 41, 10, 8, 13, and 38 showed no overall preference different from chance expectations.

4 Discussion

Subjects differ in their preferences for hue, value, and chroma. Of particular interest are the hue preferences. The majority of subjects prefer blue and dislike yellow, whilst a minority have other preference patterns. In studies of rhesus monkeys (Humphrey 1972; Sahgal et al 1975) and of pigeons (Sahgal and Iversen 1975) it was generally found that blue was preferred to other colours although not all animals showed identical results (Karr and Carter 1970). Despite the discrepancy between earlier studies of human preferences and animal preferences, due mainly to inadequate control of colour stimuli, it now seems apparent that once value and chroma are controlled then animal and human preferences are generally similar. It is thus tempting to suggest that there may be some underlying biological origins to such generalised hue preferences, and ethological arguments may be found to support such claims (Humphrey 1976). It is also tempting to suggest that it is not a coincidence that the dimension blue-yellow represents a dimension both of aesthetic preferences and also of the opponent colour system; but such speculation is clearly premature.

Acknowledgements. We are very grateful to Dr J D Mollon for his advice and assistance, and to Dr G W Granger and Dr N K Humphrey for their comments on the manuscript.

References

Allesch G J von, 1924 "Die aesthetische Erscheinungsweise der Farben" *Psychologische Forschung* **6** 1-91

Burnham R W, Hanes R M, Bartleson C J, 1963 *Colour: A Guide to Basic Facts and Concepts* (New York: John Wiley)

Chandler A R, 1934 *Beauty and Human Nature* (New York: Appleton-Century-Crofts)

Child D, 1970 *The Essentials of Factor Analysis* (London: Holt, Rinehart and Winston)

Cohn J, 1894 "Experimentelle Untersuchungen über die Gefühlsbetonung der Farben" *Philosophische Studien* **10** 562-603

Crawford D, Washburn M F, 1911 "Fluctuations in the affective value of colour during fixation for one minute" *American Journal of Psychology* **22** 579-582

David H A, 1963 *The Method of Paired Comparisons* (High Wycombe, Bucks: Charles Griffin)

Dorcus R N, 1926 "Colour preferences and colour associations" *Journal of Genetic Psychology* **33** 399-434

Eysenck H J, 1941 "A critical and experimental study of colour preferences" *American Journal of Psychology* **54** 385-394

Fletcher R J, 1978 "Recent experiences with the City spot test" *Modern Problems in Opthalmology* **19** 142-143

Garth T R, 1931 *Race Psychology* Cited by Eysenck (1941)

Granger G W, 1952 "Objectivity of colour preferences" *Nature* **170** 778-780

Granger G W, 1955 "An experimental study of colour preferences" *Journal of General Psychology* **52** 3-20

Guilford J P, 1940 "There is system in colour preferences" *Journal of the Optical Society of America* **30** 455-459

Guilford J P, Smith P C, 1959 "A system of colour preferences" *American Journal of Psychology* **72** 487-502

Helson H, Lansford T, 1970 "The role of spectral energy of source and background color in the pleasantness of object colours" *Applied Optics* **9** 1513-1562

Humphrey N K, 1972 "Interest and pleasure: two determinants of a monkey's visual preferences" *Perception* **1** 395-416

Humphrey N K, 1976 "The colour currency of nature" in *Colour in Architecture* Eds T Porter, B Mikellides (London: Studio Vista) pp 95-98

Jastrow J, 1897 "Popular esthetics of colour" *Popular Science Monthly* **50** 361-368

Kendall M G, Babington Smith B, 1940 "On the method of paired comparisons" *Biometrics* **11** 43-62

Karr A E, Carter D E, 1970 "Colour preferences in the pigeon: measurement during extinction and during variable-interval reinforcement" *Proceedings of the 78th Annual Convention of the American Psychological Association* pp 771-772

McManus I C, 1980 "The aesthetics of simple figures" *British Journal of Psychology* **71** 505-524

Mollon J D, 1980 "Post-receptoral processes in colour vision" *Nature* **283** 623-624

Moran P A P, 1947 "On the method of paired comparisons" *Biometrika* **34** 363–365
Newhall S M, Nickerson D, Judd D B, 1943 "Final report of the OSA Subcommittee on the spacing of the Munsell Colours" *Journal of the Optical Society of America* **33** 385–418
Nie N H, Hull C H, Jenkins J G, Steinbrenner K, Bent D H, 1975 *Statistical Package for the Social Sciences* 2nd edition (New York: McGraw-Hill)
Osgood G E, May W H, Miron M S, 1975 *Cross-cultural Universals of Affective Meaning* (Urbana, IL: University of Illinois Press)
Sahgal A, Iverson S D, 1975 "Colour preferences in the pigeon: a behavioural and psychopharmacological study" *Psychopharmacologia* (Berlin) **43** 175–179
Sahgal A, Pratt S R, Iverson S D, 1975 "The responsiveness of rhesus monkeys to visual stimuli differing in wavelength and angular orientation" *Journal of the Experimental Analysis of Behaviour* 24 377–381
St. George M V, 1938 "Colour preferences of college students with reference to chromatic pull, learning and association" *American Journal of Psychology* **51** 714–716