Perceptual color spacing derived from Maximum Likelihood Multidimensional Scaling.

VALÉRIE BONNARDEL[1], SUCHARITA BENIWAL[2], NIJOO DUBEY[2], MAYUKHINI PANDE[2], KENNETH KNOBLAUCH[3], DAVID BIMLER[4*]

[1] *Department of Psychology, University of Winchester, Winchester SO22 4NR, United Kingdom*
[2] *National Institute of Design, Paldi, Ahmedabad, Gujarat 380007, India*
[3] *INSERM U846, Stem Cell and Brain Research Institute, Bron, France; Université de Lyon 1, Lyon France*
[4] *School of Psychology, Massey University, Private Bag 11-222, Palmerston North 4442, New Zealand.*

**Corresponding author: d.bimler@massey.ac.nz*

Received XX Month XXXX; revised XX Month, XXXX; accepted XX Month XXXX; posted XX Month XXXX (Doc. ID XXXXX); published XX Month XXXX

The canonical application of multidimensional scaling (MDS) methods has been to color dissimilarities, visualizing these as distances in a low-dimensional space. Some questions that remain are how well the locations of stimuli in color space can be recovered when data are sparse, and how well can systematic individual variations in perceptual scaling be distinguished from stochastic noise? We collected triadic comparisons for saturated and desaturated sets of Natural Colour System (NCS) samples, each set forming an approximate hue circle. Maximum-Likelihood MDS was used to reconstruct the configuration of stimuli more accurately than the standard 'vote-count' approach. Individual departures from the consensus response pattern were minor, but repeated across stimulus sets, and identifiable as variations in the salience of color-space axes. No gender differences could be discerned, contrary to earlier results. © 2015 Optical Society of America

OCIS codes: (330.1720) Color vision; (330.5020) Perception psychology**.**

http://dx.doi.org/10.1364/AO.99.099999

1. INTRODUCTION

In a half-century-long research program, color researchers have been reconstructing subjective color spaces, based on similarity judgments among color stimuli. Often the goal was to examine the adequacy of putatively uniform color specifying systems, most notably the Munsell system (e.g. [1]; reviewed in [2]), but also the OSA-UCS [3]. The Method of Triads plays a prominent role in this tradition [4,5], eliciting purely ordinal judgments without requiring observers to convert their perceptions of dissimilarity into numerical form. Applications of the method typically combine it with incomplete data designs, for the number of triadic comparisons in a complete design increases as the third power of the number of stimuli, soon exhausting the respondents' goodwill.

The resulting perceptual spaces show a degree of variation, even in the normal trichromat population, which can be expressed within the geometrical framework in terms of dimensional-weighting parameters. These parameters – obtained through multidimensional scaling analysis (MDS) – are stable across retests [4,6,7], and display a genetic component [7].

The individual variations also have been reported to show a significant gender difference [8], with males tending to place less weight on a red-green direction of color, represented in geometrical terms as a group mean compression of the color place along the red-green axis. In [7,8] the stimuli were 32 Munsell color samples distributed roughly around a hue circle. Triads were selected randomly, with each subject responding to 70 combinations. Using 21 Munsell color samples and a list of 70 pre-selected triads, [9] partially replicated this result in young and elderly adults. This was a Balanced Incomplete Design or BID [10], with each stimulus appearing in 10 different triads, known as a λ=1 design because a given pair of stimuli appears only once.

In another application of pre-selected triads to collect odd-one-out judgments [11], 75 hue triads were printed on cards. The MDS analysis was able to resolve female carriers of color-vision deficiency from non-carriers, with the carriers (heterozygous for the X-linked genetic trait, who express both normal and abnormal forms of photopigments in their cone cells) displaying a mild form of the color-space compression that characterizes color deficiency.

Early studies in this triad-method tradition typically applied conventional MDS algorithms. This approach begins by using the responses to estimate entries in a matrix of pairwise similarities (as is prescribed in, for instance, the widely-used ANTHROPAC suite [12]). Concerns arise, however, when the triads are spread thin to accommodate a meaningfully-sized stimulus set, as in (for instance) the λ=1 BID with 70 triads for 21 stimuli [13]. The similarity estimates are low-resolution and highly susceptible to the vagaries of triad composition. Two stimuli A and B could be relatively unlike each other, but if by chance they share their triad with a third stimulus C from the far side of color space, even more different from both of them, then C will be the odd-one-out while the pair {A, B} will be assigned a similarity estimate of 1. Conversely, the pair could be quite similar but the estimate similarity can still be zero if the third stimulus in that triad is even more similar to A, leaving B to be picked as odd-one-out. One cannot rely on averaging the data over a number of respondents to eliminate sampling bias of this kind, or any distortions it might create within the MDS solution, because all respondents in a BID see a given stimulus pair in the same context.

To address this concern, we have previously applied a Maximum-Likelihood form of MDS to triadic data [7-9,11]. This interprets each odd-one-out decision as comparisons between similarities, and fits the solution directly to those comparisons, obviating the intermediate stage of estimating similarities (the sampling bias arises because all triads contribute to similarity estimates in the same way, regardless of the magnitude or the geometry of the triangles they form in color space). If stimuli are located with spatial coordinates, so that the distances between them represent the dissimilarities, Thurstone's model of pairwise choice provides estimates of the probability that one stimulus pair will be perceived as more similar than another. The arrangement of points is then optimized to maximize the probability of producing the specific combination of judgments reported by the respondents [7].

Here we used two separate implementations of the ML-MDS algorithm. The MTRIAD program (as in [7]) is coded in MS-DOS, restricting its portability. For confirmation and for wider applicability, we replicated that analysis with a new implementation of the algorithm written in the OpenSource statistical computing and graphics environment R [14]. A similar algorithm drives Takane's MAXSCAL [15], while TRISOSCAL [16, pp. 155-158] follows the same principle but minimizes a least-squares function $Stress_1$ rather than maximizing likelihood. This approach should not be confused with Ramsey's application of Maximum-Likelihood estimation to dissimilarity judgments [17].

Data were collected for two hue circles at different levels of saturation, each consisting of 21 NCS color samples, to test how well hue intervals around the circles were recovered compared to vote-count method.

A second research question arises in the process. Ref. 1 [p. 146] observed that there is a ceiling level of "intuitively palpable" dissimilarity: "If two colors differ beyond this range, i.e., red and green or blue, they are simply 'entirely different' and subjects cannot easily differentiate sizes of such color differences". It is conceivable that when color pairs have perceptually little in common, judgments of relative similarity draw partly upon culturally-transmitted conventions about relationships within the familiar color wheel of aesthetic theory. Dichromat informants internalize these relationships well enough that when their judgments of similarity among color concepts [18,19] or in color categorization tasks [20] are used to reconstruct the hue circle, it is comparable to the MDS solution summarizing perceptual judgments from normal trichomats.

This raises the possibility that some of the gender differences reported in earlier studies could be grounded in male / female differences in acculturation, which in turn might vary across cultures. We will return to this possibility in the Discussion. Thus it becomes desirable to seek further replication with respondents from a different cultural background where the 'color wheel' may not permeate popular culture as deeply. Subjects here were 77 Indian students. The BID of 70 selected triads allows this question to be addressed by correlating the responses from any pair of subjects and probing the pattern of correlations for any group differences, using Principal Components Analysis (PCA) [21].

Stimulus saturation may play a role. Good, prototypal examples of basic color terms were present in the saturated but not in the desaturated set. Although pairwise color dissimilarities were well above the threshold of discrimination in the latter, it could be argued that any influence of cultural conventions upon judged similarity will be smaller for desaturated colors, partly because the pairs are more similar (so fewer pairs exceed the limits of "intuitive judgment"), and partly because they lack the clear examples of hue prototypes which a saturated stimulus set might contain and to which learned color relationships would apply.

2. METHODS

A. Stimuli

Since this study was conducted in partnership with a paint company (namely Asian Paint), the NCS was adopted, as its hue-mixture notation is readily meaningful in applied contexts. The NCS was constructed according to Hering's opponent-hue paradigm. It locates four primary hues (Red, Yellow, Green and Blue) at the cardinal positions of the hue circle, each at the highest level of saturation, and divides each quadrant between cardinal points into 10 steps. White and Black form the poles of a third orthogonal axis, and are the apices of two cones extending upwards and downwards from the hue circle, allowing hues that are lighter or darker than those of the hue circle to be specified by their components of Whiteness or Blackness. Colors with lower levels of saturation are specified by a smaller radial coordinate of Chromaticness.

Two sets of 21 stimuli were selected from the A6 NCS box, one saturated and one desaturated set, both lying on approximate circles in the CIE-Lab chromaticity diagram, with roughly equal angular spacing are plotted in Figure 1 with their notation listed in Table 1.

Table 1 about here

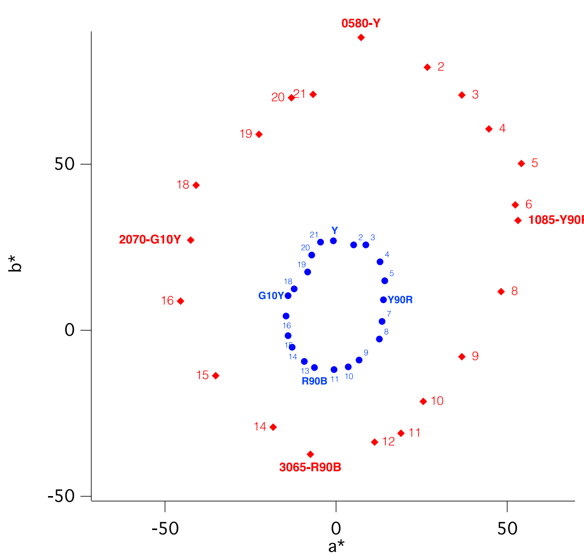

Figure 1: NCS samples for the saturated (red diamonds) and desatrurated (blue dots) sets plotted in the CIEL*a*b* diagram NCS notation of cardinal hues are indicated for each set.

The objective was to produce a hue circle that approximated perceptual uniformity. The selection process used Munsell color samples (2011 glossy edition) to guide the choice of equally-spaced samples, since perceptual uniformity is a design principle of the Munsell system but not the NCS. Conversion between systems used both chromaticity coordinates (CIEL*a*b*) computed from spectral reflectances measured with an Ocean Optic S2000 spectrometer, and visual matching by experimenters (VB, SB, ND and MP) under the D65 ceiling-panel illumination that was subsequently used in data collection.

The experimenters reached a consensus about four 'cardinal hue' Munsell chips exemplifying the four primary colors, first matching them to cardinal hues selected from the NCS gamut (marked in Table 1 and Figure 1) and then fine-tuning the choices. The remaining 17 stimuli were apportioned around the hue circle as follows: five samples each in the Y-R and R-B quadrants, three samples in the B-G quadrant, and four in the G-Y quadrant. This ensured roughly equal perceptual steps around the color wheel. These were converted back into the nearest NCS equivalent, with minor refinements for optimal visual continuity.

Stimuli were chosen with the highest available Chroma at each hue, while lightness varied from a minimum in the R-B quadrant (Value = 3, 4) through Value = 5 for cardinal Green up to 8.5 for cardinal Yellow. In NCS terms, samples had maximum Chromaticness, while Blackness peaked in the Blue sectors where the best examples are darker and was minimal in the naturally lighter colors of the Yellow sector. In CIE-La*b* terms, lightness L ranged from 33.1 to 83.6.

The desaturated set used the same hues, but Chromaticness and Blackness were constant at 10 and 20 respectively (in CIE-Lab, lightness ranged from 79.9 to 86.5). One stimulus was removed from the Y-R quadrant (Y80R) to improve the perceptual equality of hue spacing, and replaced by adding B60G in the B-G quadrant.

For three samples selected as closest to the primary hues (excepting Yellow which corresponded to the NCS primary hue), the chosen primary blue and red were shifted anticlockwise in the NCS polar diagram while the green was shifted clockwise. The same selections were made at both saturation levels. The distribution of the 17 samples in each quadrant tended to follow Sven Hesselgren's 1953 scaling (cited by [22], p. 97) with more numerous steps in the R-B quadrant and fewer in the B-G quadrant.

A list of 70 triads was prepared, following a Balanced Incomplete Design or BID [10] with λ=1: that is, each pair of stimuli appeared in one and only one triad. The 70 triads thus include all 210 possible pairs. Each stimulus appeared in 10 triads.

Each triad was assembled by cutting 20-mm squares from the appropriate sheets of NCS-A6 paper and pasting them in a triangular configuration on a grey card (matched to Munsell N5.5) measuring 167 x 193 mm. The center of the card was taken as the center of the triangle of samples. The 70 cards were bound into a black spiral box file (one each for the saturated and desaturated triads).

B. Procedure

Subjects were shown each triad plate in turn, on a grey table (matching Munsell N6.75), with D65 illumination from a X-rite D65 ceiling panel providing approximately 150 cd/m^2. For each plate they were asked to verbally indicate the most dissimilar stimulus, with the researcher recording the response, before proceeding to the next trial at their own pace. No time limit was imposed to complete the task, and the process took 15 to 20 minutes per set. Half the subjects were shown the saturated set first, and half saw the desaturated set first.

Using the index m to specify subjects ($1 \le m \le 77$), the m-th subject's responses were encoded as the lower triangle of a 21-by-21 matrix of estimated similarities, $\mathbf{S}_m$. A given matrix element s_{mij} was 1 if the i-th and j-th stimuli were most similar in the triad in which that pair appeared, and 0 if either stimulus i or j was odd-one-out in that triad (i and j specify stimuli: $1 \le i, j \le 21$). Thus each lower half-matrix contains 70 '1' elements and 140 '0' elements.

C. Subjects

Seventy-seven students (46 F) were recruited from the 1st-year undergraduate design course at the National Institute of Design (Gujarat, India), mean age 18 years (SD = 0.83). None had any special knowledge of color theory. Participation was voluntary, and part of the course design; they received a notebook as a token of appreciation for their participation. Their color vision was tested with the eight-plate shortened version of the Ishihara test [23], augmented with Plate #19, presented in the Macbeth lighting booth prior to the triadic task. The study was approved by the NID Research Ethics Committee.

D. Data analysis

We followed two complementary approaches to summarizing the data. To examine the data for forms of systematic inter-subject variation (especially variations associated with gender as reported in previous studies), we wrote each subject's responses as a column of 1s and 0s, vectorizing the lower half of the $\mathbf{S}_m$ matrix. The 210-by-77 data table was subjected to Principal Components Analysis (PCA). Here we follow [21] who applied PCA as the first stage of Cultural Consensus Analysis (CCA) [24]. Note that this is the 'Q mode' of PCA, in which subjects rather than items are the unit of analysis, so each component is a prototypical pattern of responses from an idealized subject. If subjects converge upon a consensus about the perceptual structure (i.e. the relationships among the stimuli), so that any two subjects are closely correlated in their responses, this structure emerges from PCA as the first unrotated component PC1, the 'g-factor'. Each subject's loading on PC1 measures his or her 'competence' or access to the consensus.

Second, responses were analyzed with Multidimensional Scaling (MDS) to provide a geometrical framework of perceptual scaling. The outcomes of MDS are a collective spatial model, integrating similarity judgments from all participants, and models for any given subgroup of subjects. Points in a MDS solution, representing stimuli, are arranged in a specified number of dimensions so that the distances between them reflect the corresponding inter-stimulus similarities. If two hues are consistently chosen as the most similar pair in a triad, then those two points in the space should be relatively close, with the point for the odd-one-out located farther away.

As noted, we used a Maximum-Likelihood algorithm (in two implementations) to obtain a robust solution despite the sparse nature of triads in the BID. The algorithm adjusts distances within the solution so that they accommodate, as far as possible, the reported relationships among similarities.

A recurring result in MDS studies of color perception is that even normal observers differ in their relative weighting of the axes of color space [4-6,19]. Judgments of overall color dissimilarity from some individuals place more weight on differences along the green-red axis, while others attend more to blue-yellow differences. We applied the weighted-Euclidean model of individual difference, which fits dimensional-salience parameters $w_{m\text{g-r}}$ and $w_{m\text{,b-y}}$ to each subject: in effect compressing or stretching the consensus MDS solution along its axes to improve the accuracy with which the transformed inter-point distances predict that subject's triad judgments. The ratio of these parameters provides a more succinct index of color-plane compression, $\text{Comp}_m = w_{m\text{g-r}} / w_{m\text{,b-y}}$.

For comparison, we also scaled the data with a more traditional "vote-count" approach. That is, the s_{mij} for each stimulus pair were averaged across participants to obtain similarity estimates SIM_{ij}, entries in a matrix **SIM** which we analysed with non-metric MDS (Proxscal implementation within SPSS 20; tie-breaking option enabled). The binary nature of s_{mij} values (either 0 or 1) remained a feature of the SIM_{ij}.

3. RESULTS

A. Principal Component Analysis

As noted, the subjects' responses were written as a 210-by-77 table of coarse similarity estimates. Each row identifies a pair of stimuli. In the column corresponding to a given subject, a row contains 0 if either stimulus was the odd-one-out in the triad that contained that pair, and 1 if that pair was chosen as 'most similar' in that triad (i.e. the other stimulus was the odd-one-out).

Reducing the table for the saturated set with PCA, the solution was dominated by the first principal component PC1, which accounted for 56.25% of variance (eigenvalue = 43.31). All subjects' loadings on this consensus component were acceptably high, ranging from 0.89 down to 0.38. The next two PCs were bipolar in terms of loading structure, with loadings ranging from -0.25 to 0.66 for PC2 and from -0.33 to 0.35 for PC3. The Variance-accounted-for dropped to 4.34% for PC2 and then leveled off with 2.22% for PC3 (eigenvalues = 3.34, 1.71). For consistency with earlier work we retained all three components.

Similar results emerged from PCA of the responses to the desaturated set. The first three components accounted for 65.22%, 2.31 % and 2.21% of variance (eigenvalues = 50.22, 1.78, 1.70). That is, PC1 dominated the responses, with loadings ranging from 0.93 down to 0.57. Loadings ranged from -0.29 to 0.38 for PC2, and from -0.29 to 0.35 for PC3.

PC1 loadings were significantly correlated between the two stimulus sets (r = 0.454, p < 0.001). In other words, subjects who departed from the overall consensus for saturated triads and made 'noisier' judgments tended to respond similarly to desaturated triads. Despite earlier reports of noisier judgments from males [7,25], there was no significant difference in the loadings between males and females for either set.

Despite the low contribution to variance from PC2, the variations among subjects which it captured are systematic: there was a significant correlation between saturated and desaturated PC2 loadings (r = 0.546, p < 0.001). In the next section this mode of variation is interpreted as "salience of color-space dimensions". Subjects' PC2 loadings showed no gender difference. PC3 loadings were not replicated between sets and exhibited no gender difference.

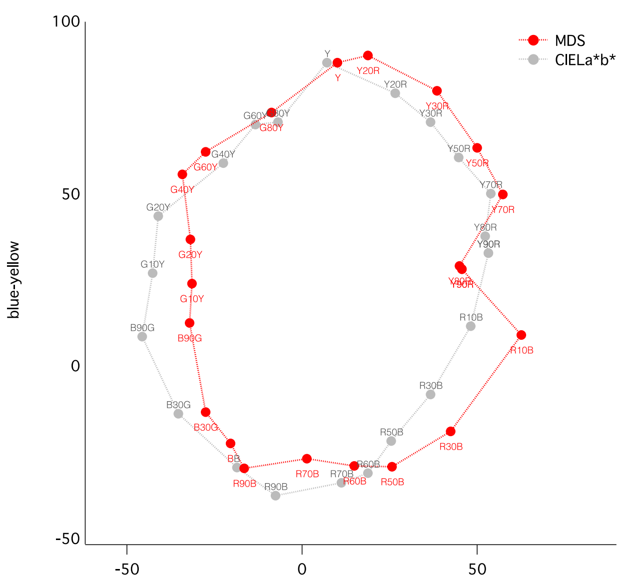

Figure 2: Saturated set : two-dimensional ML-MDS solution of dissimilarity judgments inter-point distances (red dots) plotted with the CIEL*a*b* stimuli coordinates (grey dots) for comparison.

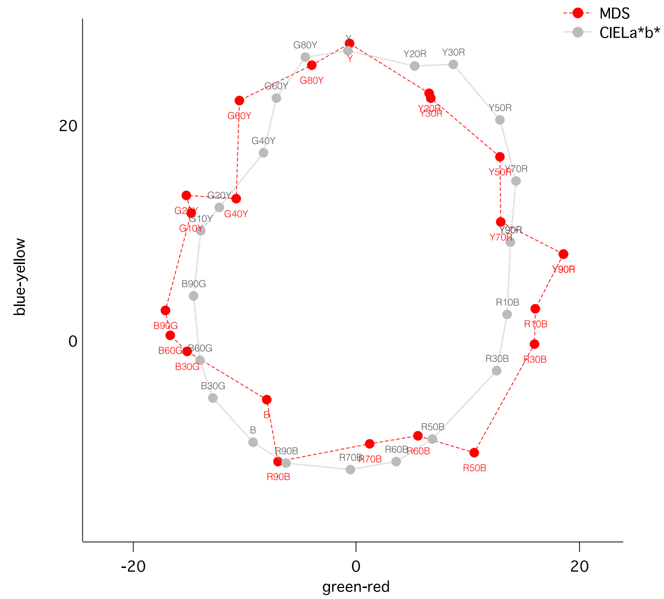

Figure 3: Desaturated set: two-dimensional ML-MDS solution of dissimilarity judgments inter-point distances (red dots) plotted with the CIEL*a*b* stimuli coordinates (grey dots) for comparison.

B. Maximum Likelihood Multidimensional Scaling

ML-MDS solutions for the saturated and desaturated stimuli are shown in Figures 2 and 3, each superimposed on the locations of the stimuli in the (a*, b*) plane. The superimposition was performed with Procrustes analysis. The normalized likelihood was better for desaturated triads, 0.773, than for saturated triads, 0.723 (where likelihood can range from 0.5 – meaning that the solution predicts the subjects' decisions no better than chance – up to 1 if its predictions are infallible).

Procrustes Analysis compares two configurations of points by measuring the mismatch between them, i.e. the sum of distances between each pair of corresponding points, after minimizing this mismatch by rotating and re-scaling the configurations. The normalised irreducible sum is the 'Procrustes distance' R^2 (where two exactly-congruent configurations have $R^2 = 0$). Here R^2 is very low, with values 0.0094 and 0.0096 for the saturated and desaturated stimuli.

The configurations are not exactly circular, but oval in shape. This matches the elongation displayed when the NCS stimuli are displayed in CIEL*a*b* space. In addition, the subjects' triadic responses locate the hues in the correct sequence around this oval (with the exception of the Y20R -Y30R inversion in the desaturated set, see Figure 3). A test of the accuracy of the reproduction at the local scale is possible because the 21 color-space distances $\Delta_{ab}(i,j)$ around the oval, between successive hues, are not identical. The correlation between the $\Delta_{ab}(i,j)$ and corresponding MDS-recovered distances reached significance for the desaturated solution (r = 0.67, p = 0.001) but not for the saturated solution (r = 0.40, p = 0.07).

A close comparison between predicted and observed inter-point distances reveals for the saturated set that Y and Y20R, B and B30G are perceived as more similar than the CIEL*a*b* model predicts, while G60Y and G80Y are perceived as more dissimilar. For the desaturated set, B60G and B90G are perceived as more similar than the CIEL*a*b* model's predictions, while G20Y and G40Y are perceived as more dissimilar and Y90R as more saturated.

Individual-difference MDS fits an index $Comp_m$ to each subject's dissimilarity comparisons, accounting for any variations among them in terms of color-plane axial compression. Values of $Comp_m$ were replicated across the two stimulus sets (p = 0.49, r < 0.001). $Comp_m$ was correlated with PC2 loadings for both the saturated (r = 0.72, p = 0.001) and desaturated set (r = 0.34, p = 0.002), leading to the interpretation of PC2 as 'relative dimensional salience'.

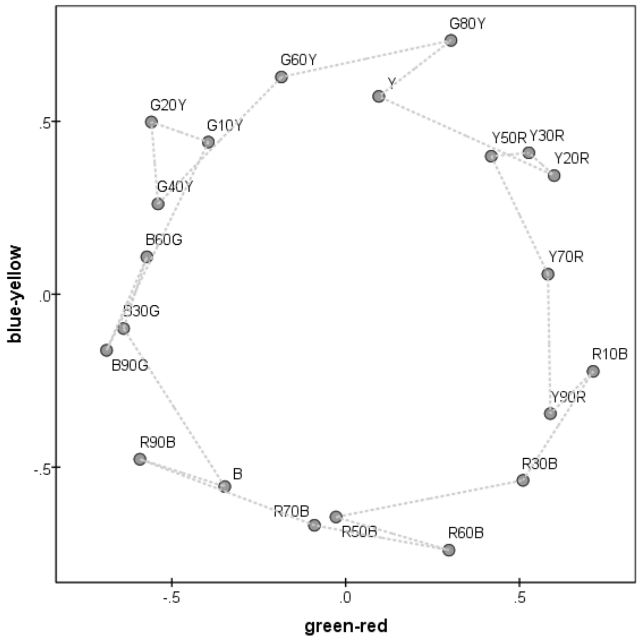

Figure 4: Desaturated data: solution from standard MDS, applied to 'vote-counted' matrix of estimated similarities.

A "vote-count" solution for the desaturated stimuli is shown in Figure 4; the result for saturated stimuli is similar. As well as the estimated similarities SIM_{ij} being polarized (as noted in the Introduction), they are only loosely related to the underlying distances, making them incompatible with a geometrical representation and resulting in high badness-of-fit values ($Stress_1$ was 0.203 for saturated and 0.198 for desaturated sets). The arrangement of the hues is circular, but with frequent back-tracking departures from the true hue sequence, with larger Procrustes distances of 0.022 (saturated) and 0.031 (desaturated) when they are compared with CIEL*a*b* locations. In particular, the 21 distances between successive hues were not correlated with the corresponding $\Delta_{ab}(i,j)$ distances (r = 0.39, 0.06 for the desaturated and saturated set respectively).

4. DISCUSSION AND CONCLUSIONS

The most immediate feature of our data (and perhaps the least surprising) is that subjects displayed a high level of agreement about the structure of color space. Their unanimity was not complete, but the second and third Principal Components – capturing departures among the odd-one-out judgments from the overall consensus, the first component – were much smaller than that consensus itself, accounting for only about 3.5% as much variance. Despite the small size of the second component, it appears to measure a consistent pattern of deviation from the consensus judgments: specifically, differences in the relative salience of the green-red and blue-yellow axes of the color plane. Subjects who displayed a high positive (or negative) loading on PC2 for saturated triads tended to vary in the same way for desaturated triads. Crucially, however, male and female students did not differ on these components, in contrast to the subtle gender difference in dimensional salience reported previously using a similar methodology [7,9].

We began by raising the possibility that when triads consist of sufficiently dissimilar pairs, there is a conceptual contribution to the odd-one-out judgments, with participants drawing on their knowledge of color relationships within the conventional hue circle. This would provide one explanation for reports of gender differences, if there is also a tendency for women to internalize these relationships better than men, or a cultural expectation that women should be more aware of or attuned to colors [26]. One robust finding in the color literature is that females acquire superior color lexicons, at least in English-speaking groups [27].

Ref. [21] collected responses to all 84 combinations of nine color concepts. Using PCA to compare subjects, they concluded that males and females were tapping into a single 'cultural consensus' about the structural represention of color, with the gender-specific departures from this implicit structure being non-zero but minor. In [25], males were less accurate than females when making tetradic judgments of conceptual color similarity, in the sense that their responses showed more internal inconsistency, but their decisions could be derived from the same conceptual structure: in effect they were 'noisy females'. But in contrast, [28] derived a spatial model for color terms by applying multidimensional scaling (MDS) to triadic or odd-one-out data, and found that female subjects placed more emphasis on a red-to-yellow-to-green dimension.

This explanation leads to the possibility that the earlier gender differences – found in New Zealand and UK participants – may not be universal, if we further allow for differences between cultures in how deeply the conventions of color relationships are internalized. It worth noting that the absence of gender difference cannot be attributed to the procedure; gender differences were obtained in a color preference task with the saturated set, in the same experimental conditions, albeit with a different sample of Indian participants. Here, in a color similarity task, individual differences were robust but were not associated with gender.

Color differences $\Delta_{ab}(i,j)$ among the desaturated pairs were smaller than their saturated counterparts by a factor of 3 (Figure 1), and thus closer to the threshold of discrimination, so one might expect comparisons among them to be noisier. If anything, however, the saturated pairs were harder to compare: their lower likelihood value implies that saturated triads were less consistent across subjects, less compatible with a geometrical solution, and provided inferior reconstructions of successive distances around the hue sequence. This is what one would expect from the suggestion that a ceiling level of dissimilarity exists, such that dissimilarity comparisons become difficult and non-intuitive above the ceiling [1].

In addition, the desaturated stimuli were not clear, prototypical examples of color categories. This would have minimized any contribution of color-conceptual knowledge to assist the perception of dissimilarities. The overall parallelism of desaturated and saturated results suggests that this lack of conceptual, conventional influence also applies to the saturated judgments, in the present study at least.

It remains to return to the form of MDS used to convert odd-one-out judgments into color spaces. A very sparse selection of triads was used here: each stimulus pair was examined only once, in a similarity comparison with two other pairs. The standard 'vote-count' analysis applies MDS to the s_{mij} values, treating each 1 or 0 as a low-resolution estimate of inter-stimulus similarity, but this turns out to be problematic. The result is to displace some stimuli towards the interior of the circular MDS solution, when they should be confined to an elliptical perimeter (as in Figure 1), and some back-tracking is revealed when points are linked in the correct hue sequence (see Figure 4: some stimuli are adjacent in CIE-LAB space but they swap their locations in the solution).

These small vagaries would be reduced in a λ=2 design, involving twice the number of triad cards (to sample each pair twice, in different contexts) and twice the testing time. However, there are inevitable trade-offs and the added reliability of a λ=2 design was not worth the substantial increase in the procedure. Crucially, the present study has indicated that acceptable outcomes can be obtained from a λ=1 design when analyzed with ML-MDS.

The reconstruction of the arrangement of stimuli within the color plane by a Maximum-Likelihood algorithm is reassuringly accurate. The triadic methodology proved to be a valid tool for assessing whether an ensemble of stimuli meets desired criteria: in the present exercise, the goal was to select hues from the available NCS range to form an approximately uniformly-spaced hue circle.

The bare minimum expectation is that the MDS solution should recover the original sequence of hues. More stringently, because the intervals between hues along the sequence do vary in CIEL*a*b* space, we expect the recovered distances to correlate with the intervals $\Delta_{ab}(i,j)$. Maximum-Likelihood MDS successfully passes these tests. At a larger scale, it also recovers a global feature of the stimulus sets, i.e. the elongation of the oval they form in the color plane (Figure 1).

Funding. Asian Paint – National Institute of Design Research Chair funding to VB.

References

1. T. Indow and N. Aoki, “Multidimensional mapping of 178 Munsell colors,” Col. Res & Appl. **8**, 145-152 (1983).
2. T. Indow, "Multidimensional studies of Munsell color solid," Psychological Review 95, 456-470 (1988).
3. J.M. Taylor and F.W. Billmeyer, “Multidimensional scaling of selected samples from the Optical Society of America Uniform Color Scales,” Col. Res. & App. **13**, 85-98 (1988).
4. C. E. Helm and L.R. Tucker, “Individual differences in the structure of color-perception,” Am. J. Psychol. **75**, 437–444 (1962).
5. P F.M. Stalmeier and C.M.M. de Weert, “Large color differences and selective attention,” J. Opt. Soc.Am. A **8**, 237-247 (1991).
6. J. D. Carroll and J.-J. Chang, “Reanalysis of some color data of Helm’s by INDSCAL procedure for individual differences multidimensional scaling,” Proceedings of the 78th Annual Convention, APA, 137-138 (1970).
7. D.L. Bimler, J. Kirkland, “Twins and odd-ones-out: A twin study of genetic contributions to variability in personal colour space,” Clin. Exp. Optom. **87**, 313-321 (2004).
8. D.L. Bimler and J. Kirkland and K. Jameson, “Quantifying variations in personal color spaces: Are there sex differences in color vision?” Col. Res. & Appl. **29**, 128-134 (2004)
9. V. Bonnardel, L. Harper, F. Dufie and D.L. Bimler, “Gender differences in weighting the dimensions of colour space: Do males neglect red-green differences?” Perception **35**, 187a (2006).
10. M.L. Burton and S.B. Nerlove, “Balanced designs for triads tests: Two examples from English,” Soc. Sci. Res. **5**, 247-267 (1976).
11. D.L. Bimler and J. Kirkland, “Colour-space distortion in women who are heterozygous for colour deficiency,” Vision Research **29**, 536-543 (2009).
12. S.P. Borgatti, “ANTHROPAC 4.0 Reference Manual,” Natick, MA: Analytic Technologies (1996).
13. A.K. Romney, D.D. Brewer and W.H. Batchelder, “Predicting clustering from semantic structure,” Psychol. Sci. **4**, 28-34 (1993).
14. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (2015)
15. Y. Takane, “A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent – Theory,” Jap. Psychol. Res. **20**, 7-17 (1978).
16. A.P.M. Coxon and C.L. Jones, “Measurement and Meanings,” Macmillan, London (1979).
17. J.O. Ramsay, “Maximum likelihood estimation in multidimensional scaling,” Psychometrika **42**, 241-266.
18. D. Jameson and L.M. Hurvich, "Dichromatic color language: 'reds' and 'greens' don't look alike but their colors do," Sens Processes **2**, 146-155 (1978).
19. R. N. Shepard and L.A. Cooper, “Representation of colors in the blind, colorblind, and normally sighted,” Psychol. Sci. **3**, 97-104 (1992).
20. V. Bonnardel, “Color naming and categorization in inherited color vision deficiencies,” Vis. Neurosci. **23**, 637-634 (2006).
21. C.C. Moore, A.K. Romney and T.-L. Hsia, “Cultural, gender, and individual differences in perceptual and semantic structures of basic colors in Chinese and English,” J. Cogn. Culture **2**, 1-28 (2002).
22. R.G. Kuehni and A. Schwarz, “Color ordered”, Oxford University Press (2008).
23. J. Birch, Diagnosis of Defective Colour Vision; Oxford University Press; 1993.
24. A.K. Romney, C.C. Moore, W.H. Batchelder and T.-L. Hsia, “Statistical methods for characterizing similarities and differences between semantic structures,” Proc. Nat. Acad. Sci. **97**, 518-523 (2000).
25. L.D. Griffin, “Males are ‘noisy females’ when it comes to reporting the psychological structure of the basic colours**,”** Perception **32**, 387 (2002).
26. D.L. Bimler, J. Brunt, L. Lanning and V. Bonnardel, Personality and gender-schemata contributions to colour preferences. In ‘Colour studies: A broad spectrum’, pp. 240-257, Amsterdam: John Benjamins (2014).
27. Mylonas, D., Paramei, G.V. and MacDonald, L. (2014). Gender differences in colour naming. In ‘Colour studies: A broad spectrum’, pp. 225-239, Amsterdam: John Benjamins.
28. N.L. Furbee, K. Maynard, J.J. Smith, R.A. Benfer, S. Quick and L. Ross, “The emergence of color cognition from color perception,” J. Ling. Anthrop. **6**, 223-240 (1997).

Table 1: NCS reference and coordinates in the CIEL*a*b* color space for the 21 saturated and desaturated samples. Labels for the saturated color set indicate their blackness, chromacticness and hue. Labels for the desaturated color set only indicate their hue as blackness and chromaticness were kept constant at 1020.

	Saturated	L*	a*	b*	Desaturated	L*	a*	b*
1	**0580-Y**	83.57	7.28	88.08	**Y**	86.50	-0.75	27.00
2	0585-Y20R	75.50	26.65	79.21	Y20R	83.84	5.20	25.62
3	0585-Y30R	68.74	36.74	70.87	Y30R	83.31	8.73	25.72
4	0585-Y50R	63.23	44.76	60.71	Y50R	82.03	12.91	20.58
5	0585-Y70R	56.16	54.09	50.24	Y70R	80.44	14.31	14.98
6	0585-Y80R	48.57	52.44	37.85	**Y90R**	80.51	13.84	9.26
7	**1085-Y90R**	44.56	53.27	32.99	R10B	79.93	13.51	2.60
8	1575-R10B	42.82	48.20	11.65	R30B	80.80	12.62	-2.67
9	3055-R30B	39.65	36.74	-7.99	R50B	81.21	6.80	-8.97
10	4050-R50B	33.65	25.55	-21.51	R60B	82.11	3.58	-11.06
11	3555-R60B	36.46	18.91	-31.01	R70B	81.81	-0.53	-11.82
12	4055-R70B	33.15	11.23	-33.69	**R90B**	81.84	-6.33	-11.23
13	**3065-R90B**	40.72	7.47	-37.34	B	81.95	-9.24	-9.29

14	3060-B	45.6	-18.49	-29.24	B30G	82.23	-12.91	-5.18
15	3060-B30G	49.67	-35.17	-13.72	B60G	83.51	-14.01	-1.64
16	3060-B90G	48.86	-45.52	8.75	B90G	83.37	-14.65	4.26
17	**2070-G10Y**	53.86	-42.57	27.05	**G10Y**	83.77	-13.99	10.36
18	1075-G20Y	62.95	-40.88	43.58	G20Y	83.91	-12.34	12.52
19	1075-G40Y	67.99	-22.50	59.02	G40Y	83.20	-8.37	17.55
20	1075-G60Y	73.05	-13.13	70.09	G60Y	85.30	-7.15	22.65
21	1075-G80Y	77.99	-6.76	70.96	G80Y	86.06	-4.59	26.45

Cardinal hue selections are indicated in bold.