Notes and Comment

Research note: A multidimensional scaling comparison of color metrics for response times and rated dissimilarities

DAVID BIMLER

Massey University, Palmerston North, New Zealand

Individual-differences multidimensional scaling was applied to a set of proximity data for equiluminant lights (Paramei & Cavonius, 1999) to explore any differences between two data collection procedures (rated dissimilarities, and same/different response times [RTs]), as well as between color-normal and abnormal observers. Two conclusions emerged: (1) The pattern of similarities from observers with anomalous color vision can be understood in terms of a compressed color plane (the weighted Euclidean model of individual differences); and (2) there is evidence that the color "plane" is either curved or governed by a non-Euclidean distance function. When color-normal observers are examined in the weighted-Euclidean framework, minor differences emerge between RT and rating data. But the main distinguishing feature of RT data is a pattern of decreasing accuracy for larger color distances. This must be taken into account, since it can itself induce curvature.

In a recent study (Paramei & Cavonius, 1999), two different experimental paradigms were applied to a single set of color stimuli (15 equiluminant lights) in order to collect matrices of dissimilarities. First, response times (RTs) were measured as observers decided whether paired stimuli were the same or different; second, the observers rated the perceived dissimilarities between pairs of stimuli. Nosofsky (1992) has pointed out that these two tasks are not necessarily measuring the same thing. Each task mediates between the proximities perceived by the informants and the data collected; it may modulate the proximities in the process.

In addition, the Paramei-Cavonius (1999) data include variations between observers. Three were normal trichromats; three had abnormal color vision. An immediate reaction is to analyze the matrices using individual-differences multidimensional scaling (MDS), to take these variations into account. The result of individual-differences MDS is typically a single *group space*, $\mathbf{X}_0$, plus a set of subject-specific parameters that describe

a *personal space*, $\mathbf{X}_m$, for each data source, where $\mathbf{X}_m$ is a compressed or an elongated version of $\mathbf{X}_0$. Here $\mathbf{X}_0$ and $\mathbf{X}_m$ are $N \times P$ matrices, describing the locations of N points (where N is the number of stimuli) in a P -dimensional space.

Individual-differences MDS will also detect differences between the rating and the same/different task, if these can be represented as a compression of color space. There are precedents for this: The distinctness of the border between areas of color can be used as an index of dissimilarity, but because border distinctness is tritanopic (Tansley & Boynton, 1976), color space is a line rather than a plane when mapped in this way. Another example is the elegant *gestalt fusion* method used by Stalmeier and de Weert (1988) to obtain triadic data. The stimuli were presented as contiguous blocks of color that could be seen in two possible configurations, depending on which pair of colors was more similar. Mapped with this method, color space is compressed along tritanopic confusion lines, in comparison with conventional triadic comparisons from the same subjects for the same stimuli.

The reliability of RT data is also an issue. A number of studies in which RT data were collected as indicators of dissimilarity or discriminability and analyzed with MDS found a tendency for the number of dimensions in the MDS solutions to be inflated (Sergent & Takane, 1987, with human observers; Blough, 1988, with pigeons). For instance, in Blough's Experiments 1B and 1C, with stimuli varying on two known dimensions, three- and even four-dimensional solutions were required to recover the structure. These excess dimensions are possibly artifacts, generated by the relatively low reliability of the shorter RTs (corresponding to supraliminal dissimilarities). It is necessary to test this possibility as a preliminary step to analyzing the matrices.

REACTION TIME UNRELIABILITY

To a good approximation, RTs decline toward an asymptotic value B as dissimilarity increases, following a negative exponential trend. This implies that a small dissimilarity δ_{ij} can be inferred from the corresponding large RT_{*ij*}: Representing δ_{ij} by x , and RT_{*ij*} by y ,

$$x \sim -k^{-1} \log[(y - B)/(A - B)]$$

for observer-specific parameters k , A , and B (the inverse of the formula in Paramei & Cavonius, 1999). The slope of this curve means that, as y approaches B , random perturbations will have more impact on x : For large enough distances in color space, the RTs lose their predictive value as they taper off toward their asymptotic minimum and the rank order among them is lost in the noise. But in-

I am grateful to G. V. Paramei and C. R. Cavonius for generously giving me access to their data and for the invitation to extend the results of reanalyzing them in this research note. Correspondence concerning this article should be addressed to D. Bimler, Human Development Studies Section, Department of Educational Studies and Community Support, Massey University, Private Bag 11-222, Palmerston North, New Zealand (e-mail: d.bimler@massey.ac.nz).

cluding inaccurate matrix entries in an MDS analysis can be actively detrimental to its outcome; they are not simply ignored. The randomized rank order of the RTs is incompatible with a low-dimensional representation. "Very low RTs, corresponding to large distances in psychological space, may thus tend to be disordered by variability. In attempting to fit such disordered points, ALSCAL tends to equalize large distances. To do this while still providing good fits at shorter distances, the algorithm curls the representation of the points into new dimensions" (Blough, 1988, p. 71).

A noteworthy feature of the Paramei–Cavonius (1999) RT data, when each observer's matrix is analyzed in isolation, is the gap opening up in each solution (Figures 3A, 4A, and 8A in Paramei & Cavonius), separating R, O, Y, GY, and G from the other stimuli (including W). In comparison with the solutions for observers' ratings (Figures 3B, 4B, and 8B), the blue/purple stimuli are compressed into a narrow strip, with W displaced toward them. To summarize this difference, a pair of two-dimensional MDS solutions were derived, one for the combined RT matrices, and one for the combined ratings; these are su-

perimposed in Figure 1, with lines connecting the corresponding items. A similar distortion appears in MDS solutions for RT data presented by Sergent and Takane (1987) and by Blough (1988). Since this displacement has the effect of making larger distances more equal, one explanation is that, like extra dimensions, it is an artifact resulting from the disordered RT ranking.

An alternative explanation is that the RT procedure treats the blue and yellow halves of the color circle differently (compared with the rating procedure), but no sign of this kind of nonlinear distortion was found in earlier MDS studies of RT (Mollon & Cavonius, 1986). The only significant procedural difference between the Mollon–Cavonius and the Paramei–Cavonius (1999) studies is the number of trials (i.e., accuracy). RTs were averaged over 50–90 trials in the former and over only 10 in the latter—and over 16 trials in Sergent and Takane (1987).

Whether inaccuracy among the short RTs is affecting the MDS solutions can be tested by analyzing longer RTs in isolation. Fifty percent of the RTs were simply replaced by *missing data* values: A matrix entry was jettisoned if the corresponding entry in the rating matrix (for that sub-

Figure 1. Superimposed two-dimensional multidimensional scaling solutions for rated dissimilarity and reaction time data. Items in response time solution are represented by squares, connected by dotted lines to the corresponding items in the rating configuration.

ject) exceeded the median value. If, indeed, there is no systematic difference between the accuracies of long and short RTs, there is no danger of contaminating the modified RT matrices with information from the ratings by using the latter as the criteria for rejection.

MDS solutions in three dimensions were retained for the ratings, RT, and modified RT matrices, using the *weighted Euclidean* model of individual variation (described below). Solutions were compared in two ways: by using the product-moment correlation coefficient r^2 between interpoint distances and the least-squares Procrustes statistic g_I between point coordinates (Peay, 1988). The irreducible discrepancy (as a fraction of total variance) between configurations that have been rescaled and rotated to maximize the fit between them is indicated by g_I ; here, it was calculated with the program GPA. These statistics were $r^2 = .905$, $g_I = .017$ between the rating and RT solutions; $r^2 = .958$, $g_I = .008$ between ratings and modified RTs; $r^2 = .950$, $g_I = .011$ between RTs and modified RTs. A similar pattern emerges when comparing two-dimensional solutions and, again, with separate solutions for each observer's data. We conclude that the longer RTs (shorter distances) are indeed more compatible than the short RT values with the rating data. Since artifacts from the short RTs are a potentially confounding factor for an important aspect of the Paramei-Cavonius (1999) data—the existence of a third dimension—the subsequent analysis uses the modified RT matrices.

MODELING THE INDIVIDUAL DIFFERENCES

The 12 matrices (six of rated dissimilarities and six of modified RTs) were analyzed with the MDS program ALSCAL, using the weighted Euclidean model (Carroll & Wish, 1974). This mode of MDS assumes that, as observers assess the interitem dissimilarity, they vary in the importance or salience they place on particular dimensions. In other words, the model allows $\mathbf{X}_m$ to be obtained from $\mathbf{X}_0$ by compressions and expansions along orthogonal axes (i.e., compression and stretching are the forms of intersubject variations it postulates). It is also known as the INDSCAL model, but it is convenient to reserve that label for metric MDS, whereas the treatment here is non-metric. Dimensional weights (salience) $w_{m,p}$ are derived as subject-specific parameters. In most cases, $\mathbf{X}_0$ itself is very similar to the MDS solution obtained by combining the matrices without allowing for individual differences—that is, by treating them as replications—but it is rotated to a set of axes that are no longer arbitrary.

A three-dimensional solution was optimal (Kruskal's stress₁ = 0.13, 0.11, and 0.10 in two, three, and four dimensions). The coordinates for each point in $\mathbf{X}_0$ are listed in Table 1, while Figure 2 plots the first two dimensions. In this case, subject variation in the first two dimensions is dominated by the difference between the deuteranopic

Table 1
Coordinates of Stimulus Points in
Three-Dimensional Weighted Euclidean Solution

Stimulus	D1	D2	D3
G	-0.92	1.70	0.12
GY	-1.27	0.98	0.31
Y	-1.41	0.32	0.40
O	-1.19	-0.64	1.36
B	1.14	0.74	-0.22
BP	0.55	-0.11	-1.24
P'	-0.04	-0.76	-1.18
PR	-0.18	-1.67	0.16
I	1.71	0.76	1.31
IP	1.50	0.16	-0.12
PI	1.01	-0.83	-0.32
P''	0.35	-1.47	-0.24
GB	0.08	1.50	-0.74
W	-0.30	0.29	-1.71
R	-1.02	-0.98	2.12

observer E.H.R. and the others. $\mathbf{X}_0$ is rotated so that the deuteranopic confusion lines are parallel to D2. Thus, D2 is a green-red axis, whereas D1 is a blue-yellow axis (in the weighted Euclidean model, the axes are the same for all the subjects). Table 2 lists dimensional saliences, or *weights*, w_{m1} , w_{m2} , w_{m3} , for each subject. It shows that the personal configuration for E.H.R. is elongated along D1 and compressed along D2, to the extent of being practically one-dimensional for both rating and RT data. Conversely, D2 has a relatively high salience for the normal trichromats C.R.C., E.L.E., and G.V.P.: No dimension is dominant in their respective $\mathbf{X}_m$ s.

Although emphasizing the perceptual differences of the color-vision deficient observers, these results obscure more subtle differences between rating and RT data. When the weighted Euclidean analysis was repeated using only the six matrices from the three normal trichromat observers, the configuration is similar to Figure 2, but rotated through 45° so that D1 is a red-indigo axis, whereas D2 is a G-W-PI axis. The $\mathbf{X}_m$ for the RT data are relatively compressed along this new D1, indicating that same-different judgments in this direction take longer than rated dissimilarities would suggest. This corresponds to the earlier observation (Mollon & Cavonius, 1986) that the mechanisms involved in RT judgments introduce tritan-like features to the color metric.

Applying the weighted Euclidean model to the complete RT matrices (either on their own or in combination with the ratings) produces a structure of *folded dimensions*. The blue-yellow axis is *folded* into two unipolar axes (white-blue, and white-yellow), reflecting the different behavior in individual solutions of the blue and yellow halves of the color plane.

A metric analysis of the dissimilarity ratings (treating them as interval, rather than as ordinal) leads to similar results. Indeed, when the ratings from the m th subject are

Figure 2. D1-D2 plane of three-dimensional group configuration for rated dissimilarities. Value of D3 is indicated by size of lettering.

plotted against the reconstructed distances D_m (the matrix of distances between points in X_m), the relationship between them turns out to be quite linear.

We turn now to the third dimension. This is not independent of the first two dimensions; the points do not fill the space. Remaining confined to a roughly spherical shell, they curve up from W at the lowest point to saturated hues, such as R, I, and O. For convenience this can be labeled

an *achromatic* dimension. This *achromatic axis* provides only a small reduction in stress₁ and accounts for only a small proportion of the model's variance accounted for: .13, as opposed to .49 and .32 for the first two dimensions. However, it fits the *spherical model of color perception* (Izmailov & Sokolov, 1991).

A suspicion lingers that this curvature is an artifact of scaling a planar structure in three dimensions, in the same way that items with an underlying one-dimensional structure often arrange themselves in a parabola—the *horse-shoe* artifact—when two dimensions are available. Table 2 provides evidence against this. The salience of D3, w_{m3} , varies across observers and experimental procedures in a systematic way. Whatever the nature of the third dimension, the way it varies between observers within the rating task is duplicated within the RT task (with a superimposed between-task variation: for each observer, J.O.N. excepted, w_{m3} is higher in the RT procedure). For both tasks, the achromatic axis is most salient for C.R.C. and the two protanomalous observers, H.A.N. and J.O.N. It contributes relatively little to the private configurations of E.L.E. and G.V.P. Which is not to say that their rating matrices produce two-dimensional solutions when analyzed in isolation. However, the D3 in such analyses (Figures 4B and 5B

Table 2
Subject Parameters (Dimensional Weights)
for Weighted Euclidean Solution

Subject	Data type	w_{m1}	w_{m2}	w_{m3}
C.R.C.	Ratings	0.60	0.65	0.37
	RTs	0.41	0.77	0.47
E.L.E.	Ratings	0.69	0.64	0.18
	RTs	0.62	0.64	0.37
G.V.P.	Ratings	0.59	0.74	0.19
	RTs	0.42	0.79	0.35
H.A.N.	Ratings	0.73	0.56	0.31
	RTs	0.77	0.37	0.47
J.O.N.	Ratings	0.72	0.43	0.48
	RTs	0.78	0.35	0.45
E.H.R.	Ratings	0.95	0.10	0.18
	RTs	0.92	0.15	0.21

in Paramei & Cavonius, 1999) have little in common with D3 of the group analysis.

THE CURVATURE OF THE COLOR PLANE

A surface is positively curved (locally) if $C < 2\pi R$, where C and R are the circumference and radius of a circle. This appears to be the case for the color plane. It curves in three dimensions, so that the distances from one stimulus to the next around the perimeter of a circle (saturated stimuli) are smaller than one would expect, given the circle's radius. This assumes that distances are measured along arcs on the surface itself; they could, alternatively, be measured along chords through three-dimensional space, but this merely introduces a monotonic transformation, which cannot be detected with ordinal-level data. Note that curvature is not defined for a one-dimensional space (such as the personal configuration for E.H.R.): Nonmetric MDS cannot distinguish a straight line from a horseshoe.

A possible explanation is that perceived dissimilarity includes contributions from an additional quality of color, related to the distance between a stimulus and white (e.g., saturation or brightness). Naturally, MDS is not suitable for observing this quality directly. It can simply call our attention to the fact that white is more easily discriminated than it should be (less similar to chromatic stimuli), given the pattern of similarities between the chromatic stimuli—that is, that absence of color acts like a percept in its own right.

The MDS analysis can be tailored to this explanation by scaling X_0 in three dimensions, while imposing a constraint that the points must lie on the surface of a sphere—instead of asserting, as above, that the points are “confined to a roughly spherical shell.” A number of programs have incorporated this constraint (e.g., Cox & Cox, 1991). Alternatively, curvature can be accommodated without embedding the color plane in a higher dimensional space, by using a non-Euclidean formula for distance (Indow, 1982).

Models of individual differences become particularly complex when they must coexist with curvature. If the points of a three-dimensional X_0 are located on a spherical surface of constant curvature, the problem is that the X_m s are no longer uniformly curved: The sphere is compressed into spheroids. This is most clear for E.H.R., for whom GB and PR are desaturated and almost indistinguishable from W. In the RT task, E.H.R. identified GB–W pairs as *same* in 9 out of 10 trials and PR–W pairs as *same* in 10 trials. The three-dimensional weighted model cannot duplicate these confusions: Flattening X_0 along D2 does not affect the D3 values that separate W, on the one hand, from GB and PR, on the other.

An individual-differences model would be more in keeping with the spherical model of color vision if it represented stimuli on a plane and projected the private configurations onto a sphere *after* they underwent compression (desaturating some stimuli)—thus producing uni-

Table 3
Subject Parameters for Individual-Differences
Model Incorporating Weights and Curvature

Subject	Data type	w_{m1}	w_{m2}	R_m
C.R.C.	Ratings	0.85	1.13	7.7
	RTs	0.79	1.18	9.2
E.L.E.	Ratings	0.87	1.12	11.7
	RTs	0.86	1.12	16.8
G.V.P.	Ratings	0.79	1.17	10.2
	RTs	0.77	1.19	10.8
H.A.N.	Ratings	0.97	1.03	9.3
	RTs	1.08	0.92	9.9
J.O.N.	Ratings	1.03	0.97	9.2
	RTs	1.06	0.93	10.6
E.H.R.	Ratings	1.37	0.37	13.9
	RTs	1.34	0.47	90.0

Note— w_{m1} and w_{m2} have been normalized so that $w_{m1}^2 + w_{m2}^2 = 2$.

formly curved X_m . A program was written to follow this procedure, implementing Kruskal's algorithm for non-metric MDS. The extent of curvature (i.e., the salience of the achromatic axis) appears to vary between observers. To allow for this, the radius of the sphere on which each X_m is projected was made into another subject parameter, R_m , to be optimized along with the w_m . With only two parameters per stimulus, instead of three, this remains simpler than the three-dimensional weighted Euclidean model.

Table 3 lists the subject parameters resulting from applying this program to the rating and modified RT matrices (X_0 is hard to distinguish from Figure 2). To allow comparisons between R_m values, the personal configurations were rescaled to make the average d_{ij} equal to 10 in each case. High values correspond to low curvature. Contrary to Table 2, the RT data show less curvature than the ratings, a puzzling discrepancy that was, perhaps, produced by applying different individual-difference models. By comparing observers in each experimental procedure separately, it can be seen that curvature is greatest in the X_m for C.R.C. and least in those for E.L.E. and G.V.P. (it is not possible to interpret R_m for the one-dimensional X_m of E.H.R.). In effect, this is restating the conclusions reached from the three-dimensional weighted Euclidean model: that there is consistency for each observer, across the forms of data. If the three-dimensionality or curvature of the color plane is a numerical artifact, for some reason its salience varies systematically between observers.

REFERENCES

- BLOUGH, D. S. (1988). Quantitative relations between visual search speed and target-distractor similarity. *Perception & Psychophysics*, 43, 57-71.
- CARROLL, J. D., & WISH, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. II, pp. 57-105). San Francisco: Freeman.
- COX, T. F., & COX, M. A. A. (1991). Multidimensional scaling on a sphere. *Communications in Statistics*, 20, 2943-2953.
- INDOW, T. (1982). An approach to geometry of metric space with no a

- priori* mapping function: Multidimensional mapping according to Riemannian metrics. *Journal of Mathematical Psychology*, **26**, 204-236.
- IZMAILOV, C. A., & SOKOLOV, E. N. (1991). Spherical model of color and brightness discrimination. *Psychological Science*, **2**, 249-259.
- MOLLON, J. D., & CAVONIUS, C. R. (1986). The discriminability of colours on c.r.t. displays. *Journal of the Institution of Electronic & Radio Engineers*, **56**, 107-110.
- NOSOFSKY, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, **43**, 25-53.
- PARAMEI, G. V., & CAVONIUS, C. R. (1999). Color spaces of color-normal and color-abnormal observers reconstructed from response times and dissimilarity ratings. *Perception & Psychophysics*, **61**, 1662-1674.
- PEAY, E. R. (1988). Multidimensional rotation and scaling of configurations to optimal agreement. *Psychometrika*, **53**, 199-208.
- SERGEANT, J., & TAKANE, Y. (1987). Structures in two-choice reaction-time data. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 300-315.
- STALMEIER, P. F. M., & DE WEERT, C. M. M. (1988). Large color differences measured by spontaneous gestalt formation. *Color Research & Application*, **13**, 209-218.
- TANSLEY, B. W., & BOYNTON, R. M. (1976). A line, not a space, represents visual distinctness of borders formed by different colors. *Science*, **191**, 954-957.

(Manuscript received September 2, 1998;
revision accepted for publication March 18, 1999.)