The psychological relationship between lightness and saturation*

JULES DAVIDOFF

University College of Swansea, Swansea, Wales, U.K.

Lightness and saturation, though independent in a physical sense, have often been seen to be psychologically related. The first experiment uses a multidimensional scaling technique to again verify this nonindependence for untrained Os. A further experiment, using a matching technique, quantifies this relationship, giving the extent of the error in lightness for varying levels of saturation. This relationship is found to be independent of hue if saturation is measured by the colorimetric purity of the hue. Errors in value (lightness on the Munsell system) were found to be linearly related to colorimetric purity.

Color is normally specified by the three physical variables of intensity, dominant wavelength, and purity. These correspond to the psychological variables of lightness (brightness or value), hue, and saturation (chroma). However, it has been suggested from time to time that the psychological variables are neither independent nor valid. Note, for example, James's (1890) famous contention that "pink is surely not a portion of our feeling of scarlet." Newhall (1940), too, when discussing the spacing of the Munsell colors, says that untrained Os would have given slightly different settings.

In particular, there seems to be a confusion between lightness and saturation. Indeed, Warren (1967) points out that Richardson (1929) believed that by mixing white with red he was changing only saturation. However, the separation of saturation and lightness has been considered worthwhile. On a psychological basis, this could be justified from the finding of Wright and Rainwater (1962) that chroma (saturation) on the Munsell system, and not value (lightness), was the most important component when the emotional meaning of a color was considered on the Osgood semantic differential (Osgood, Suci, & Tannenbaum, 1957). More recently, though, Handel and Imai (1972) have again shown that a stimulus varying on the supposed independent dimensions of lightness and saturation is not in fact perceived as a two-dimensional stimulus, though, as they point out, this does not mean that with training the stimulus cannot be perceived dimensionally.

Confirmation of the need for three psychological variables in color analysis has come from the multidimensional scaling technique essentially based on interstimulus distance estimates found from the method of paired comparisons. If distances between stimuli vary in three physical dimensions, then Indow and Kanazawa (1960) and Ramsay (1966) have shown that three psychological dimensions are needed to fit the data. The programs used in these studies give an indication of the stress (goodness of fit) by fitting the distances given from the paired comparisons of all the stimuli. Using

two physical dimensions, Indow and Uchizono (1960) conducted a dimensional analysis of 21 colors from the Munsell ranges that were of constant lightness. A fair-to-good fit was found for two psychological dimensions. The same result was found for pigeons (Schneider, 1972), and the two dimensions could be safely labeled hue and saturation. The first method adopted here uses the multidimensional scaling technique, but asks Os to make judgments of stimuli from a hue with respect to only one physical dimension, i.e., lightness. The second dimension therefore acts only as a distraction. If the data can be fitted by only one dimension, then we know that lightness is psychologically independent of saturation. However, if another dimension is needed to fit the data, the most likely candidate must be saturation.

EXPERIMENT I

Method

Observers. Two female first-year undergraduate students (D.A. and H.F.), who had normal color vision but no experience with experiments in color vision and no knowledge of color theory, served as Os.

Stimulus Materials and Apparatus. The Munsell color system was devised in order to present colors which varied systematically on the dimensions upon which color is analyzed. The commercially available Munsell Color Book presents four subjectively equidistant hue levels (2.5, 5.0, 7.5, 10.0) for each of 10 hues, including red (R), green (G), yellow (Y), and purple-blue (PB). In this study, only 5.0R, 5.0G, 5.0Y, and 5.0PB were used. The hues can also vary according to their lightness or darkness. This is called value in the Munsell color system. There are 10 equal steps from Value 0/ (black) to Value 10/ (white). The number of subjective steps for the saturation of color (chroma) varies from hue to hue. However, the higher the chroma number, the purer the color. Thus, Munsell Color 5.0R7/4 would be a red, three steps of lightness down from white and four chromatic steps away from gray. This would normally be described as a dull pink. For a lengthier description of the Munsell system, see Newhall et al (1943).

Stimuli were presented on a white (70% reflectance) board subtending a visual angle of 24 x 30 deg and at right angles to the line of sight. In the middle, on the left of the board, was a fixed holder for a 21 x 12 mm Munsell chip. Sixteen thin wires ran horizontally across the board. On these were holders which could be moved independently and remotely by the O. The board was illuminated from 45 deg by two 150-W iodine quartz

*Part of this paper was given at a meeting of the Experimental Psychology Society, Amsterdam, April 1973.

Fig. 1. The fit in 1-4 dimensions given by KYST for the scaling distances of 16 stimuli of constant hue. Data is from one O and for one color (H.F., red).

lamps (A1/216) in Gnome projectors with Wratten filters to obtain Illuminant C. The illuminance of the board was 0.6 log fL.

Within the four hues used, samples of the following Munsell notations were chosen: 2/1, 4/1, 6/1, 8/1, 2/2, 5/2, 8/2, 3/4, 5/4, 7/4, 5/6, 6/6, 7/6, 5/8, 6/8, and 7/8.

Procedure. The procedure adopted to obtain distances for scaling was that of Indow and Uchizono (1960). The paired comparisons were not obtained singly, but each stimulus acted in turn as a standard for the rest. One of the 16 stimuli of a given hue was placed in the holder on the left of the board, and the rest at random in the moveable holders. At the top of the board, three 1.5-cm gray squares of neutral Munsell Values 4/, 5/, 6/ were placed with their centers 3 cm apart to assist in the scaling procedures. The O viewed the display from 1.5 m. The following recorded instructions were then given.

"Before you on the board are some colors. Your task is to pay attention to the lightness and darkness of those colors and ignore anything else. This is rather like trying to predict how a color will turn out in a black-and-white photo or on black-and-white TV. For example, if you watch football on black-and-white TV, you know it is sometimes difficult to tell the teams apart because their shirts look similar, though you know in real life it

would be easy to distinguish them because of the color. What I want you to do is to tell me how far apart in lightness each of the colors on the wires is from the standard on the left. For example, if you thought the standard in the holder on the left would come out to be like the middle gray on the top of the board and one of the colors to be like one of the other grays, you would move that color to be a distance from the holder equal to the distance between those two grays. Colors will turn out to be both darker and lighter than the standards. Colors which are an equal number of steps of lightness or darkness away from the standard should be placed at the same distance. You have as long as you like to make any decision, and you are quite at liberty to go back to any color and move it again. Please do not squint when you make these judgments. If you have not understood any of these instructions, please ask."

When the O was satisfied with the distances the colors were set at, the distance of each color from the standard was measured. The E then removed the colors from the holders, placed a different stimulus as the standard, and the rest of the stimuli of that hue were replaced in the moveable holders in a different prearranged random order. The procedure was repeated until all the stimuli of a given hue had acted as standards. While the Os were able to take the stimuli for each presentation in any order, the usual procedure was to set the darkest ones first with reference to the neutral grays at the top, and then to set all the others relative to those, and finally to take an overall look at the colors and make final adjustments.

The first testing session consisted of a trial run lasting 1 h; the final testing of each hue took 90 min and was carried out over 2 days at two sessions per day. It was not possible to completely control for order effects, but the two Os used the four hues in reverse order to each other.

Results

Each paired comparison is represented twice, since each color acts as the standard. The means of these two readings gave the scaling distance for each paired comparison. An analysis of these distances was given by the multidimensional scaling program, KYST.¹ A typical fit for one hue is given in Fig. 1. This gives the stress (goodness of fit) values in fitting the data from one to four dimensions for one O. The other three graphs from that O and the four graphs from the other O are similar to Fig. 1 and have stress values within 2% of that shown in Fig. 1. From Kruskal (1964a), the fit in one dimension of Fig. 1 would be given to be poor-to-fair. The fit in two dimensions would be given as fair-to-good. The improvement of fit by fitting more than two dimensions is fairly small. The first dimension (see Table 1) is, as one would expect, obviously related to

Table 1
Mean Scores on the First Dimension Found from KYST for the Fit in One (A) and Two (B) Dimensions Grouped According to Value

		Value						
Hue		2	3	4	5	6	7	8
Red	A	-1.598	-1.087	-0.872	-0.371	0.357	0.800	1.602
	B	-1.567	-0.998	-0.870	-0.327	0.355	0.770	1.538
Yellow	A	-1.621	-1.321	-0.650	-0.399	0.429	0.771	1.533
	B	-1.503	-1.241	-0.584	-0.336	0.408	0.723	1.391
Green	A	-1.731	-1.200	-0.679	-0.329	0.464	0.710	1.487
	B	-1.559	-1.127	-0.577	-0.294	0.482	0.684	1.227
Purple-Blue	A	-1.599	-1.304	-0.787	-0.391	0.366	0.901	1.528
	B	-1.452	-1.303	-0.646	-0.397	0.366	0.820	1.441

Table 2
Mean Scores on the Second Dimension Found from KYST
Grouped According to Chroma Level

Hue	Chroma Level				
	1	2	4	6	8
Red	0.208	-0.018	0.0417	-0.117	-0.187
Yellow	0.231	0.048	0.175	-0.221	-0.311
Green	0.298	0.199	0.005	-0.179	-0.422
Purple-Blue	0.150	0.081	0.043	-0.031	-0.293

intensity. But the fit is really not quite good enough to allow us to say that the data are fitted by only that dimension. Inspection of the output in two dimensions for individual Os and for individual hues unfortunately does not allow us to clearly label the second dimension. However, contraction of the data by averaging over the two Os and value (lightness) at each chroma level produces an interesting result. Table 2 gives these mean scores for the second dimension of the KYST analysis for each of the hues. If a monotonic trend test is conducted on the data, there is found to be a significantly increasing negative score which goes along with increasing chroma ($z = 4.3$, $p < .001$). The change in sign from positive to negative is, however, quite arbitrary.

EXPERIMENT II

There does seem to be some necessity to think of the inexperienced O as regarding saturation and lightness as being related. As group data seems to be necessary to label the second dimension, it seemed unnecessary to use the exhausting procedure of Experiment I. Therefore, a simpler method (described below), adapted from Evans (1959), was adopted.

It was possible that the relationship found in Experiment I between lightness and saturation was brought about by the instructions which mentioned practical examples of these lightness judgments and therefore might pertain only, say, to judgments of photographs. Therefore, it was worthwhile to compare the general instructions with specific instructions which mentioned only lightness.

Method

Observers. Twenty (10 male, 10 female) first-year undergraduate students with normal color vision and no knowledge of color theory were given instructions similar to those of Experiment I. Sixteen similar Os were given instructions which mentioned only lightness.

Apparatus. A white board (24 x 30 deg visual angle) was presented at right angles to the line of sight of the O. Mounted upon it were nine grays from the neutral Munsell range. These were of 21 x 12 mm size and were placed at 5 cm distance from each other. They went from Value 1/ through to Value 9/. Beneath these, in a movable holder, was placed a color chip which could be moved along the line of the grays by the O. For each of the four Munsell ranges of 5.0R, 5.0PB, 5.0Y, and 5.0G, the following colors were used: Values 5/, 6/, 7/ for all the chroma levels of /1, /2, /4, /6, /8. Also, Munsell Chips 8/1, 4/1,

3/1, 2/1, 8/2, 4/2, 3/2, 2/2, 8/4, 4/4, 3/4, 8/6, and 6/10 were included to encourage full use of the gray scale. The stimuli were illuminated from 45 deg by Illuminant C at 0.6 log fL.

Procedure. The O sat 1.5 m from the white board. A sample color chip was placed in the movable holder. The O heard recorded instructions asking him to match the colored chip to the lightness of one of the grays. These instructions were either of a general nature, as in Experiment I, or of a specific nature, only mentioning the fact that the judgment of lightness was required.

The O moved the color chip back and forward along the line of the grays until he agreed on a match. It was then removed and replaced by another color from the same range. This was repeated until all 28 samples from the range had been judged. The order of the four hues was controlled by a Latin square. Testing took between 25 and 45 min.

Results

Performance in the task was measured by the error made in making the judgments. An error score of -1 was given if the match was one value unit too light and +1 if it was one value unit darker than the true match. The results of the Os were averaged over the three values of 5/, 6/, 7/ for all chroma levels. Figure 2 shows the result of plotting the average errors (measured in value units) against chroma levels. An analysis of variance (Kirk, 1968, p. 237) was carried out for both sets of instructions. Differences were highly significant. For the general instructions, the differences between the hues, $F(3,361) = 6.09$, $p < .001$, and between chroma, $F(4,361) = 11.27$, $p < .001$, were significant. A similar result was obtained for the specific lightness instructions, as can be seen in Fig. 2: for the Munsell hues, $F(3,319) = 13.9$, $p < .001$, and for differences

Fig. 2. Average error in value units for the lightness matches of four hues under two different sets of instructions. Positive errors represent matches against darker standards.

Table 3
Average Error Rate ($\bar{X}$) from Fig. 2 With the Associated Colorimetric Purity (P_c)

Red			Yellow			Green			Purple-Blue		
$\bar{X}$	σ	P_c	$\bar{X}$	σ	P_c	$\bar{X}$	σ	P_c	$\bar{X}$	σ	P_c
0.11	0.61	5.3	0.26	0.63	18.2	0.32	0.69	9.5	0.08	0.48	3.1
0.16	0.80	9.3	0.12	0.87	25.0	0.24	0.75	11.5	0.12	0.90	4.2
0.23	0.85	18.8	0.54	1.19	49.2	0.35	0.87	26.9	0.22	0.81	8.2
0.39	1.06	26.2	1.06	0.99	72.1	0.53	0.96	39.0	0.27	1.21	11.0
0.61	0.98	33.7	1.21	1.28	78.4	0.58	0.90	44.8	0.38	1.22	13.4

between chroma levels, $F(4,319) = 12.1$, $p < .001$. The interaction between chroma and hue also reached significance, $F(12,319) = 2.9$, $p < .05$. Tests between the alternative instructions for each hue revealed no differences between the two sets of instructions.

It might, therefore, appear that not only are saturation and lightness related for the untrained O, but the relationship differs according to hue. However, chroma is a subjective measure of saturation, and chroma levels for different colors represent different purity levels. Indow and Stevens (1966) found that the exponents (slopes) of the functions relating subjective to objective purity (colorimetric purity) differ according to hue. These exponents were yellow (2.9), red and green (1.7), and blue (1.4). The exponents found by Indow and Stevens seem, then, to be related to the error rates found here. It would, therefore, make sense to convert the chroma values to colorimetric purity. Hardy (1936) gives the relationship between colorimetric purity (p_c) and excitation purity (p_e) to be $p_c = (p_e y_1)/y_s$, where y_1, y_s are the trichromatic coefficients of the points on the spectral locus and the sample point, respectively. Excitation purities for the Munsell samples are available from the Munsell Color Company or are given by Glenn and Killian (1940) and Granville et al (1943).²

Table 3 shows that when the average error rates of Fig. 2 are reconsidered in this manner hue no longer seems to be an important factor. The error rates for the four different hues now seems to be the same.³ A test of the equality of the slopes of the regression lines (β) is given by Kendall (1948). Using this procedure, the difference between β values is given by $F(3,12) = 0.41$. Though one is attempting to prove the null hypothesis here, the low F ratio does give us confidence in believing that the slopes of the four regression lines are the same.

DISCUSSION

One interesting point concerning the results of the second experiment is the direction of the errors. With only very few exceptions, the Os matched samples which increased in purity to standards that were darker than the true match. This seems to be in the opposite direction to that usually found with untrained Os. Newhall (1940) states quite clearly that for untrained Os the tendency is "to see chromatic colors as lighter than does the more experienced observer." For aperture colors, Boynton and Kaiser (1968) and Padgham (1971)

obtained a similar relationship. The present experiment differs from all these in that it provides the O with a scale upon which to make his lightness judgment, but why this procedure should make him give more darker judgments is not clear. Despite this, it seems clear that for the naive O lightness and saturation are not independent.

The saturation of a color can also be measured by the excitation purity (p_e) which is based on the chromaticity diagram of the XYZ system. p_c and p_e are not linearly related, and the simple relationship found here for p_c does not hold for p_e . It is worthwhile noting that Indow and Stevens (1966) found that their scaling data was also better fitted by p_c .

It would appear that the errors made in judging lightness come about solely from the distraction provided by the colorimetric purity of the color. The colorimetric purity of a sample is related to the amount of white in a mixture necessary to combine with a spectral color in order to obtain a match to that sample. In any instance, increasing the proportion of the spectral color influences the judgments of the apparent lightness of that color despite the fact that there is no change in physical intensity. To a first approximation, at least, there seems to be a simple linear relationship between errors in value and colorimetric purity.

Warren (1967, 1969) also points to other simple relationships in the judgment of colors. These involve the square root law which relates subjective intensities to the square root of a variety of physical measures. For example, in such a way, value is related to the square root of physical lightness (Godlove, 1933). Extrapolating from this, it would seem that the relationship between errors in physical lightness and colorimetric purity would also obey the square root law.

REFERENCES

- Boynton, R. M., & Kaiser, P. Vision: The additivity law made to work for heterochromatic photometry with bipartite fields. *Science*, 1968, 161, 366-368.
- Cook, W. M. Ability of children in color discriminations. *Child Development*, 1931, 2, 303-320.
- Evans, R. M. Fluorescent and gray content of surface colors. *Journal of the Optical Society of America*, 1959, 49, 1049-1059.
- Gibson, K. S., & Nickerson, D. An analysis of the Munsell color system based on measurements made in 1919 and 1926. *Journal of the Optical Society of America*, 1940, 30, 591-608.
- Glenn, J. J., & Killian, J. T. Trichromatic analysis of the Munsell book of color. *Journal of the Optical Society of America*, 1940, 30, 609-616.
- Godlove, I. H. Neutral value scales, II. A comparison of results

- and equations describing value scales. *Journal of the Optical Society of America*, 1933, 23, 419-425.
- Granville, W. C., Nickerson, D., & Foss, C. E. Trichromatic specifications for intermediate and special colors of the Munsell system. *Journal of the Optical Society of America*, 1943, 33, 376-385.
- Handel, S., & Imai, S. The free classification of analyzable and unanalyzable stimuli. *Perception & Psychophysics*, 1972, 12, 108-116.
- Hardy, A. C. *Handbook of colorimetry*. Cambridge, Mass: M.I.T. Press, 1936.
- Indow, T., & Kanazawa, K. Multidimensional mapping of colors varying in hue and chroma. *Journal of Experimental Psychology*, 1960, 59, 330-336.
- Indow, T., & Stevens, S. S. Scaling of saturation and hue. *Perception & Psychophysics*, 1966, 1, 253-271.
- Indow, T., & Uchizono, T. Multidimensional mapping of Munsell colors varying in hue and chroma. *Journal of Experimental Psychology*, 1960, 59, 321-329.
- James, W. *Principles of psychology*. New York: Holt, 1890.
- Kendall, M. G. *Advanced theory of statistics*. Vol. 2. London: Griffin, 1948.
- Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif: Brooks/Cole, 1968.
- Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964a, 29, 1-28.
- Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964b, 29, 115-130.
- Newhall, S. M. Preliminary report of the O.S.A. subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America*, 1940, 30, 617-645.
- Newhall, S. M., Nickerson, D., & Judd, D. B. Final report of the O.S.A. subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America*, 1943, 33, 385-418.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- Padgham, C. A. The direct estimation of the luminosity of colored light sources. *Vision Research*, 1971, 11, 577-590.
- Ramsay, J. O. Some statistical considerations in multidimensional scaling. Educational Testing Service, Princeton, New Jersey, 1966.
- Richardson, L. F. Quantitative mental estimates of light and colour. *British Journal of Psychology*, 1929, 20, 27-37.
- Schneider, B. Multidimensional scaling of color difference in the pigeon. *Perception & Psychophysics*, 1972, 12, 373-378.
- Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962a, 27, 125-140.
- Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 1962b, 27, 219-246.
- Warren, R. M. Quantitative judgments of color: The square root rule. *Perception & Psychophysics*, 1967, 2, 9, 448-452.
- Warren, R. M. Visual intensity judgments: An empirical rule and a theory. *Psychological Review*, 1969, 76, 16-30.
- Wright, B., & Rainwater, L. The meaning of color. *Journal of General Psychology*, 1962, 67, 89-99.
- Young, F. W. A FORTRAN IV program for nonmetric multidimensional scaling. The L. L. Thurstone Laboratory, University of North Carolina, Report No. 56, 1968.

NOTES

1. The program, KYST (Kruskal-Young-Shepard-Torgerson), used here was supplied by J. B. Kruskal of Bell Laboratories. This program incorporates the basic MDSCAL approach to scaling developed by Shepard (1962a, b), then refined by Kruskal (1964a, b), and TORSCA (a preliminary factor analysis to determine the initial configuration to start computation), which was developed by Young (1968) from the work of Torgerson.

2. As an approximation, an alternative method for obtaining P_c is to use Eq. 10 of Gibson and Nickerson (1940), i.e.,

$$P_c(V/C) = \frac{CP}{V} c(5/5),$$

in which C, V are the Munsell chroma and value, respectively. Colorimetric purity can then be worked out from the data they supply.

3. Regression lines were calculated from the data given in Table 3, though one could argue that all four lines must also pass through the point (0,0). This is because it must be taken for granted that an O can match one gray to another. Indeed, Cook (1931) shows that even 6-year-old children can do this without error.

(Received for publication December 24, 1973;
revision received February 22, 1974.)