Quantitative Estimates of Saturation*

JUDITH WHEELER ONLEY, CARL L. KLINGBERG†, MARVIN J. DAINOFF, AND GARY B. ROLLMAN
University of Rochester, Rochester, New York
(Received 18 June 1962)

Scales relating saturation to colorimetric purity have been derived from magnitude estimations for red (W29) and green (W65) test stimuli of 4° and 1.5° subtense. Scales based on homochromatic comparisons at several luminance levels follow a power-law form, with exponents ranging from 1.80 to 2.30. There is evidence that saturation scales for red and green light differ in slope, with the red scale being of consistently higher slope than the green. This difference is demonstrated clearly by heterochromatic saturation matches for a large number of colorimetric purities, but is not substantiated by magnitude estimations utilizing heterochromatic comparisons. It is suggested that observers are not able to apply a ratio definition of saturation in making quantitative comparisons of heterochromatic stimuli by the method of magnitude estimation, and that real differences between saturation scales for differing hues may be masked by the adoption of median magnitude estimations as a quantitative index of perceived saturations.

INTRODUCTION

AMONG the basic psychophysical data relevant to a description of color are the functions which relate the psychological aspects of hue, brightness, and saturation to a corresponding aspect of the light stimulus (usually, dominant wavelength, luminance, and colorimetric purity, respectively). Although it has been shown that each psychological aspect of color may be influenced by interactions among the perceptions produced by all three major psychophysical correlates,[1] it is nonetheless possible to estimate how each aspect varies with its closest correlate by controlling the other two variables while manipulating the most relevant one. The function relating brightness to luminance has been extensively investigated in this way.[2] The relation between hue, expressed by color names descriptive of variations in color across the spectrum, and dominant wavelength is widely accepted. Saturation is, however, a much less familiar concept and has been given somewhat conflicting definitions.[3] It also is the least frequently abstracted of the three perceptual aspects mentioned. Given a constant luminance and dominant wavelength, what changes in saturation result when we vary the colorimetric purity of the stimulus over the full range from white to a spectrally pure light?

The subjective judgment of relative saturation for stimuli of the same dominant wavelength but differing colorimetric purities can be made with considerable accuracy. Most observers would agree on an ordinal scale of saturation with little or no difficulty: increases in the colorimetric purity of a given color are accompanied by systematic increases in its saturation. When we ask observers to make more quantitative judgments of saturation, however, the definition of saturation which has been adopted by each observer becomes critically important. The problem of the definition of saturation has been widely recognized.[4,5] It has been customary when dealing with empirical evaluations of the subjective magnitude of specific colorimetric purities to distinguish between what we might call a "chroma" definition and a "saturation" or "amount of color" definition.[6] In the former, saturation is defined as ". . . the attribute determining the degree of difference from the achromatic color perception most resembling it. . . ."[7] This definition corresponds to the definition of Munsell chroma, and would imply that equal Munsell chromas would appear equally saturated. The second definition states that saturation is ". . . the attribute of a visual sensation which permits a judgment to be made of the proportion of pure chromatic color in the total sensation. . . ."[8] Whereas the first definition implies the psychophysical operation of interval judgments of difference from gray, the second implies ratio judgments, abstracting the chromatic component from a complex perception. There is no reason to expect that these two classes of psychophysical operations would yield identical estimates of the relation between saturation and colorimetric purity, especially when applied to stimuli which differ in dominant wavelength. It has been demonstrated that the two definitions produce conflicting results when applied to stimuli of differing luminances.[6]

For a wide range of sensory continua, it has been empirically demonstrated that psychophysical judgments of intervals and of ratios result in differing

* These experiments were conducted in the Department of Psychology, University of Rochester, under the support of a grant from the National Institute of Neurological Diseases and Blindness, United States Public Health Service.

† Now at The Boeing Company, Seattle, Washington.

[1] Committee on Colorimetry, *The Science of Color* (Thomas Y. Crowell Company, New York, 1953), p. 252.

[2] cf. R. M. Hanes, J. Exptl. Psychol. **39**, 438, 719 (1949); S. S. Stevens, Psychol. Rev. **64**, 153 (1957); R. G. Hopkinson, Nature **178**, 1065 (1956); J. W. Onley, Science **132**, 1668 (1960).

[3] Morton C. Bradley, Tech. Studies in the Field of the Fine Arts, Fogg Art Museum, Harvard University, April 1938, Vol. VI, No. 4.

[4] W. D. Wright, *The Measurement of Colour* (The Macmillan Company, New York, 1958), p. 153.

[5] A. Pope, *The Language of Drawing and Painting* (Harvard University Press, Cambridge, Massachusetts, 1949). See also footnote 3.

[6] cf. R. M. Evans, J. Opt. Soc. Am. **49**, 1049 (1959).

[7] D. B. Judd, *Color in Business, Science and Industry* (John Wiley & Sons, Inc., New York, 1952).

[8] British Standard 1611, 1953, quoted by W. D. Wright, reference 4, p. 154.

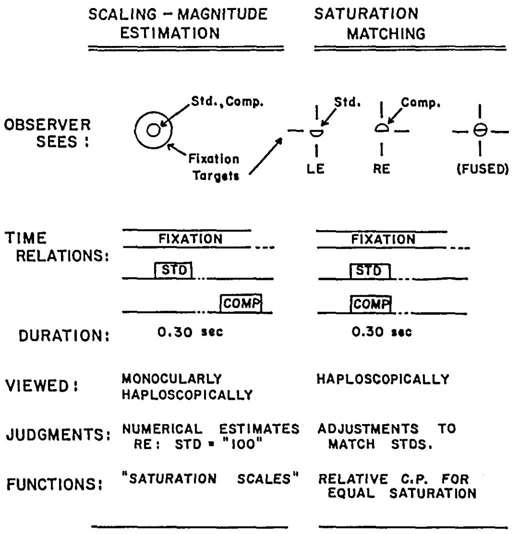

FIG. 1. General procedures.

estimates of the subjective magnitude scale.[9] The question of which scale is the more fundamental remains unanswered. It appears quite probable that the practical choice of scale will continue to be dictated by the conditions of each specific application.[10] At a more basic level however, there is mounting evidence that a number of sensory experiences evaluated in the ratio sense can be adequately described by a single general law whose parameters vary with the modality and conditions of stimulation.[9] Although this psychophysical power law has not as yet been demonstrated to have unequivocal physiological significance, it is tempting to attribute the generality of the power-law findings to some general underlying discriminative process, and to consider ratio judgments to be the more fundamental.

To derive a psychological ratio scale relating saturation to its psychophysical correlate colorimetric purity, we may adopt any one of several methods, all of which require the observer to abstract from a complex color perception that aspect which represents its saturation (where saturation is defined as the relative amount of chromatic color in the total sensation). Two such scaling methods are those of ratio production and magnitude estimation. Galifret has reported saturation scales for red surface colors obtained by each of these methods and compared to a scale synthesized from discriminability data.[11] While the scale obtained from magnitude estimations was found to approximate a power-law form (with slope, and thence exponent of the power law equal to 1.0), that derived from the production of ratios was found to be curvilinear on a log–log plot, and suggestive of a somewhat higher slope. Data of the magnitude-estimation procedure were found to agree with the scale obtained from discriminability data, a result accepted by the author as evidence for the tentative labeling of saturation as a metathetic sensory continuum.[9]

Scales derived by the method of magnitude estimation and the method of ratio production for the visual dimension of brightness have been shown to agree within the errors of measurement of either procedure.[12] The discrepancy in scales obtained by the two methods for the saturation dimension is disturbing. Although it is not clear from Galifret's report that stimuli for the two procedures were totally equivalent, his findings raise serious questions concerning the application of traditional scaling procedures to the saturation dimension. A complete empirical investigation of the saturation scale should include some procedure which provides an independent verification of its form and the values of its parameters.

The present study utilizes the methods of magnitude estimation and saturation matching to investigate the form and parameters of a ratio scale relating saturation to colorimetric purity of stimuli yielding color perceptions of red and green hues at several luminance levels. The magnitude estimation procedure is applicable to the derivation of scales for stimuli of each dominant wavelength and luminance level independently, but offers no information as to how the saturations of differing dominant wavelengths may be related. The psychophysical investigation of this question has been carried out both by the use of heterochromatic saturation estimations and by a direct saturation-matching procedure. While it is in no way implied that this study offers a complete picture of saturation relations, its limited data present information of serious methodological implications by providing an evaluation of the consistency of the scales and scale parameters predicted by these differing judgmental procedures.

PROCEDURES

Three independent experiments were performed. The general procedures employed in the study are summarized in Fig. 1, and their specific conditions will be described briefly.

Magnitude Estimates—4° Test Field

Each of the 18 color-normal observers was instructed to make numerical estimates of the saturation produced by test stimuli of three different dominant wavelengths (those of Wratten filters W29 red, W65 green and W48 blue, as illuminated by tungsten light, and of varying colorimetric purities). All judgments were made relative to a standard of the same dominant wavelength [13] and

[9] S. S. Stevens and E. H. Galanter, J. Exptl. Psychol. 54, 377 (1957).
[10] B. S. Rosner, "Psychophysics and Neurophysiology," in *Psychology: A Study of a Science*, edited by S. Koch (McGraw-Hill Book Company, Inc., New York, 1962), Vol. 4.
[11] Y. Galifret, L'Année Psychologique 59, 35 (1959).
[12] cf. J. W. Onley, Science 132, 1668 (1960).
[13] Note, however, that particularly for the blue stimulus (results of which are not reported here), there is a marked shift in hue with

were nominally based on a ratio definition of saturation. Observers were told to define saturation as the "relative amount of color" in each sample, and were shown color samples of varying purities to provide some degree of familiarization with the saturation dimension.

Observations were made monocularly, with the preferred eye. Each observer made one estimate for each of 10 colorimetric purities of the three dominant wavelengths, at each of two luminance levels (68 and 6.8 mL). Dominant wavelength, luminance level, and colorimetric purity were randomized in the design of this experiment.

Magnitude Estimates—1.5° Test Field

Each of 17 color-normal observers was instructed to make numerical estimates of the saturation of red and green (W29, W65) test stimuli of varying colorimetric purities. Judgments were made relative to both homochromatic and heterochromatic standards, and were as in the above procedure, based on a ratio definition of saturation. Familiarization with the saturation dimension was accomplished by the use of color chips and sample stimuli provided by the actual experimental apparatus.

Observations were made haploscopically; that is, the observer viewed the standard test stimulus with one eye, then viewed the comparison stimulus with the other eye. The haploscopic procedure was adopted because of the complex after-image problems encountered with the high luminance level in this experiment. For this study, each observer made one estimate for each of eight colorimetric purities of the two dominant wavelengths at each of two luminance levels (680 and 68 mL). Dominant wavelength, luminance level, colorimetric purity, and type of judgment (i.e., homochromatic or heterochromatic) were randomized in the design of this experiment, with specific restrictions being placed on successive stimulus presentations (e.g., a homochromatic series always began each experimental session; heterochromatic comparison of a given test stimulus never immediately followed its homochromatic comparison).

For both magnitude-estimation procedures, the fixation target was a dim (0.18 mL) tungsten-illuminated annulus, viewed for 5 min prior to a given stimulus series.

Saturation Matches—1.5° Test Field

Each of two color-normal observers adjusted a comparison stimulus to match in saturation each of a series of standard stimuli. Homochromatic judgments provide the baseline against which heterochromatic comparisons may be evaluated. The observers in this experiment were highly trained in making colorimetric matches and experienced in the complex judgments required by heterochromatic photometry.

Observations were made haploscopically, as shown in Fig. 1. Red and green (W29 and W65) test stimuli of varying colorimetric purities were studied at a luminance level of 74 mL. Hue, colorimetric purity, and type of judgment were randomized, and the experimental design included six judgments by each observer for each of the experimental conditions.

For the matching procedure, fixation targets were dim hairline reticles, illuminated by a tungsten source at a level which was just visible to the observer. Test stimuli were delivered in an otherwise darkened field, and were preceded by a 5-min period of adaptation to the surround prior to each series.

changes in colorimetric purity. See W. DeW. Abney, *Researches in Colour Vision and the Trichromatic Theory* (Longmans Green & Co., London, 1913) for data bearing on the Abney Effect.

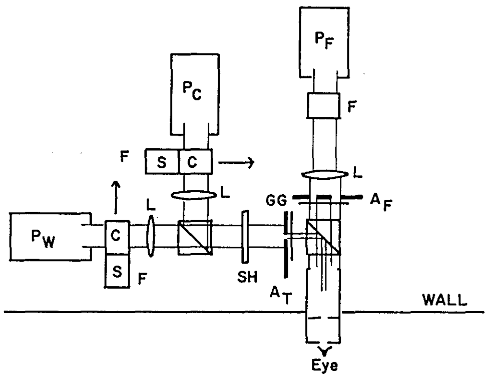

FIG. 2. Schematic diagram of the apparatus used in the scaling experiments. Projectors: P_f (fixation field), P_w (white component of test field), P_c (chromatic component of test field); F filter racks: S (standard), C (comparison); L lens; Sh shutter; GG ground glass slides; A_f fixation aperture; A_t test aperture.

APPARATUS

For the scaling experiments, stimulus fields were provided by three 35-mm slide projectors arranged as shown schematically in Fig. 2. Filter mountings for standard and variable stimuli were positioned to allow rapid substitution of one for the other. Test-flash duration was controlled by an electromechanical timing unit, activating a rotary solenoid shutter. The presentation of each test flash was under the control of the observer, who pressed a control button when he considered his fixation on the target to be adequate. A fine-ground glass slide positioned across each of the apertures assured that the normally viewed stimuli appeared uniformly bright.

For the matching experiment, stimulus fields were obtained by means of a multiple-channel Maxwellian-view optical system which has been described in detail elsewhere.[14]

Light Sources

The light sources used for the three experiments were tungsten-filament lamps, all operated at approximately

[14] J. W. Onley, J. Opt. Soc. Am. 51, 667 (1961).

the same color temperature.[15] For the 4° scaling study, projectors W and C contained 750-W 105-V spiral-filament projection bulbs. To obtain better field uniformity, these projectors were converted to utilize vertical-ribbon-filament microscope lamps (6-V, 18-A) for the 1.5° scaling study. Light sources in the matching apparatus were 12-V 50-cp auto bulbs with a V-shaped filament that is particularly adapted to the stimulus requirements of the Maxwellian view.

Calibrations

Luminance calibrations of all stimulus fields were obtained by direct-matching procedures, by using the Macbeth illuminometer and monocular viewing. For the chromatic components of each stimulus field, this necessitated heterochromatic brightness matching, which was carried out independently for each of the two observers in the matching experiment. Specifications of colorimetric purity throughout this report are based on the judgments of these observers, and probably depart at least slightly from measurements based on the standard observer.

Colorimetric purity for spectral-plus-white combinations of lights is defined as follows:

$$Pc = B_\lambda / (B_\lambda + B_w), \qquad (1)$$

where B_λ is the luminance of the spectral component and B_w is the luminance of the white component. This definition applies not only to monochromatic-plus-white-light combinations, but also to stimuli found to yield a color match to them. Given the excitation purity of a filtered light of given color for the 1931 CIE standard observer, colorimetric purity for any actual observer with normal color vision may be estimated by the following formula[16]:

$$Pc = Pe y_\lambda / y_S, \qquad (2)$$

where Pc is the excitation purity for the illuminant used, y_S is the chromaticity coordinate of the filter for that illuminant, and y_λ is the corresponding chromaticity coordinate of its dominant wavelength. This formula was employed to estimate the colorimetric purities of all chromatic and chromatic-plus-white stimuli utilized in these experiments.

Under many viewing and judgment conditions, the light provided by the tungsten sources used here may be perceived as possessing some hue. In the context of the present experiments, it was in fact defined as "white," and in no instance did any observer assign a saturation estimate greater than zero to the white component presented alone. In fact, a number of observers failed to perceive a relative saturation greater than zero even for stimuli of colorimetric purities as high as 0.2 (note for example Fig. 3 in which lower limits of the ratio estimates in the log–log plot cannot in some cases be shown). It should be emphasized that in the scaling experiments, observers were required to make estimates of saturation *relative* to standard stimuli which were moderately chromatic in appearance. Compared to these standard stimuli, the tungsten source itself appeared quite achromatic.

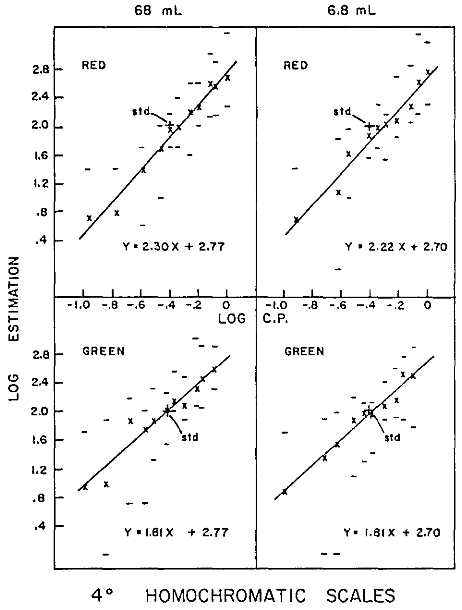

FIG. 3. Median and range of saturation estimates of 18 observers. The functions drawn are least-squares linear fits of the median estimates, and are characterized by the equations shown.

RESULTS AND DISCUSSION

Results of the magnitude estimation study for 4° fields are summarized in Fig. 3, which shows the median and range of estimates based on the group of 18 naive observers. All estimates were made by comparing the variable test stimuli with a homochromatic standard of colorimetric purity equal to 0.4, where this standard was assigned the number "100." Observers for both magnitude-estimation studies were allowed to report either numbers (e.g., "200") or ratios (e.g., "twice as saturated") to express the relative amount of chromatic color in each comparison stimulus.

A straight line on the log–log coordinates of Fig. 3 and of subsequent figures describing magnitude estimation data indicates agreement with a power-law prediction. The best-fitting straight line has been derived by the method of least squares, and its equation is shown for each condition. The slope constant in each equation represents the exponent of the power law for that case. Note that for all conditions of this experiment, median estimates of saturation approximate a power-law form. There also appears to be some evidence

[15] Lamps for the scaling experiment were operated at approximately 2800°K, those for the matching experiment at approximately 2900°K. See G. R. Harrison, R. C. Lord, and J. R. Loofbourow, *Practical Spectroscopy* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1948), p. 174.

[16] Y. LeGrand, *Light, Colour and Vision* (John Wiley & Sons, Inc., New York, 1957), p. 181.

that there is a systematic difference between saturation scales for this particular pair of red and green stimuli, with the green scale showing a consistently lower exponent (i.e., the saturation of the green stimulus increases less rapidly than that of the red, for similar variations in colorimetric purity of the two stimuli), as might be expected due to the higher saturation of spectrum red.

Figures 4 and 5 summarize the findings of the magnitude estimation study for 1.5° test fields. All estimates were made by comparing the variable test stimuli with a standard of colorimetric purity equal to 0.5, where this standard was assigned the number "100." For the conditions of Fig. 4, comparison and standard were of the same dominant wavelength; for those shown in Fig. 5, heterochromatic comparisons were made as indicated. Best-fitting straight lines derived by the method of least squares are described by the equations shown for each function. Median estimates of saturation once again closely approximate power-law form: both homochromatic and heterochromatic comparisons yield a nearly linear result on the log–log plot. Note, however, from Fig. 4, that there is no evidence for difference in slope of the red and green scales at the 68-mL level which corresponds to that of the earlier study, and that the difference in slope for the 680-mL condition is in fact in the reverse direction.

Although heterochromatic comparisons, particularly at the lower luminance level, appear to yield a power-law result with an exponent which is systematically lower than that evidenced by homochromatic judgments, the results for red-vs-green comparisons and those for the reverse case are virtually identical; there

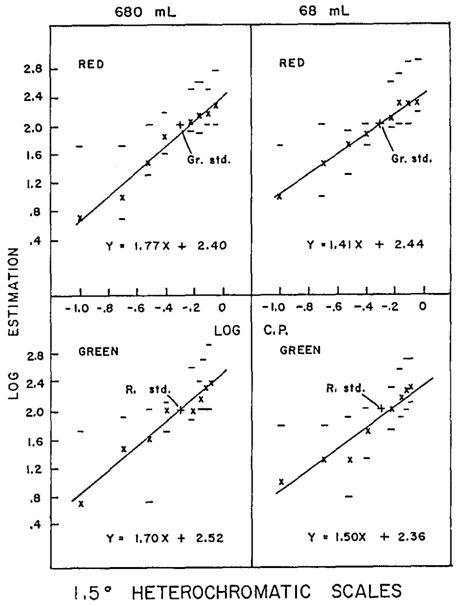

FIG. 5. Median and range of saturation estimates of 17 observers. The functions drawn are least-squares linear fits of the median estimates, and are characterized by the equations shown.

is no evidence from the data of heterochromatic estimates, Fig. 5, which would substantiate a systematic difference between scales for red and green lights. Although heterochromatic and homochromatic comparisons were made by the same observers within a single experimental session, most observers reported estimates much more conservatively in the heterochromatic series, and were reluctant to assign extreme values when making cross-color comparisons. It appears highly probable that few observers were able to adopt a stable ratio definition of saturation when forced to evaluate heterochromatic pairs.

It is clear from the complete ranges of estimates, shown in Figs. 3–5, that observers were generally able to reproduce an ordinal scale of saturations, even when stimuli were presented singly and could not be cross checked with stimuli of the continuum other than the standard. (Judgments greater than "100" were seldom given for stimuli below the standard; judgments less than "100" for stimuli of colorimetric purities higher than the standard were similarly infrequent, even for the heterochromatic comparisons.) Observers participating in the magnitude estimation studies reported very little difficulty in assigning numbers which they considered to be proportional to the relative amounts of color in comparison and standard stimuli when these were of the same dominant wavelength. Although it is unlikely that the specific instructions of the magnitude-estimation procedure would foster an interval definition of saturation by the observer, there remains no inherent means of verifying the adoption of a ratio definition. It may be that a saturation scale based on quantitative estimates by a naive observer possesses little more than

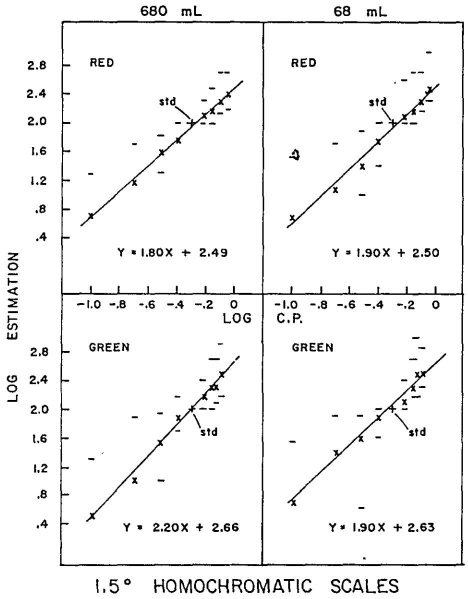

FIG. 4. Median and range of saturation estimates of 17 observers. The functions drawn are least-squares linear fits of the median estimates, and are characterized by the equations shown.

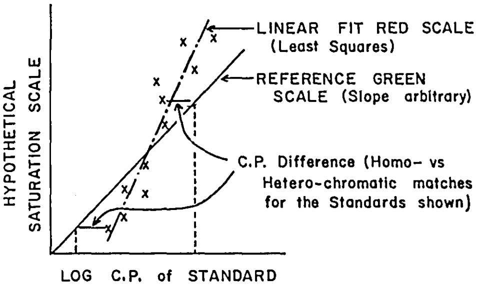

FIG. 6. Prediction of the relative slopes of saturation scales on the basis of heterochromatic vs homochromatic saturation matches.

ordinal properties, and this is most certainly true when the estimates require the abstraction of saturation from the complex perception of colors which also differ in dominant wavelength.

The results of the magnitude-estimation experiments differ considerably from those reported by Galifret.[11] While the scales obtained by Galifret for a red surface color exhibited power-law form, they were of considerably lower exponent than those reported here. It is by no means a novel finding that a subjective dimension of color varies not only with its physical correlate, but with other aspects, such as mode of appearance (e.g., surface or illuminant mode; while for the dark-adapted observer the brightness scale has an exponent of about 0.30, the similar scale for lightness exhibits an exponent near 1.20).[9] The practical definition of saturation may also be a critical factor contributing to lack of agreement in the two studies. Further, there is considerable variability between saturation scales obtained for individual observers in the present study, with the exponents of individual scales varying from as low as 0.90 to about 2.90, and clearly encompassing the range within which Galifret's median exponents lie. Although it is attempted by the use of a large group of observers to provide an adequate index of individual differences and a stable median estimate of the saturation scale, individual differences cannot be ignored, and comparison among conditions should ideally be made with the same group of observers as their own control. Familiarization and context effects, however, mitigate strongly against the use of the same observers for large numbers of judgments on the same sensory continuum, when the method of magnitude estimation is employed.

Due to the high degree of variability inherent in the scales derived from magnitude estimations, it is unlikely that one could reliably establish differences between scales for differing colors, even if these differences do in fact exist. Quantitative estimates such as those required by this procedure are relatively insensitive measures. Those who hold a conservative view of psychophysics would discount such measures in favor of the more critical judgments obtainable by procedures which require the observer to match stimuli, or to "null out" differences between classes of stimuli which vary along a specific physical continuum.[17] The latter type of judgment cannot directly yield a subjective scale for the dimension under consideration, but it may be applied in investigating or verifying the parameters of such a scale derived by ratio procedures.[14] In the present investigation, saturation matches for homochromatic and heterochromatic pairs of stimuli have been utilized to predict the relative exponents of saturation scales for red and green light.

Assuming an arbitrary reference slope of 1.00 for the subjective scale relating the log saturation of green (W65) light to log colorimetric purity, we may ask what slope will a scale obtained under the same conditions for red (W29) exhibit? The application of matching data to this question is illustrated schematically in Fig. 6. The colorimetric purity required for a homochromatic match in haploscopic view (see Fig. 1) is utilized as the reference baseline against which heterochromatic matches are evaluated. Plotted in this illustration are the differences in median colorimetric purity between homochromatic and heterochromatic matches for each of a series of standards of specific colorimetric purities. The use of the difference measures compensates for possible sensitivity differences in the two eyes. Figure 7 summarizes these difference measures for each of two trained observers, under conditions in which each color was varied in turn to yield a saturation match with itself and with the second color. Both observers required relatively less green light in the chromatic-plus-white stimulus to equal a given saturation of red light at low colorimetric purities of the standard, and relatively more green light at high

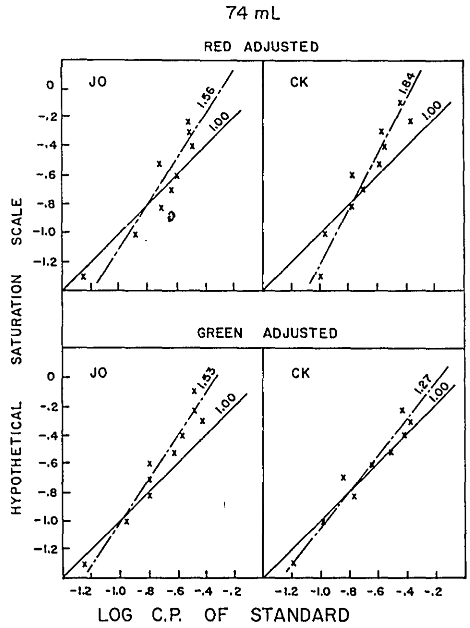

FIG. 7. Relative slopes of saturation scales for red and green stimuli of 1.5° subtense.

[17] G. S. Brindley, *Physiology of the Retina and Visual Pathway* (Edward Arnold, Ltd., London, 1960).

colorimetric purities of the standard. Stated another way, low-purity greens would appear more saturated than reds of the same purity; at an intermediate level (about 0.15 colorimetric purity) equal purity of the two colors implies equal saturation; high-purity greens would appear less saturated than reds of the same purity.

The experimental data of Fig. 7 have been fitted with straight lines by using a least-squares technique which minimizes horizontal deviations (since difference measures are plotted as deviations in x, rather than y) from the arbitrary reference curve for green light. In all cases, the slope for the red function is greater than that for green, and the average ratio of red to green slopes is 1.55, with observer CLK showing some systematic differences between adjustments of red and of green. Although the absolute values of these slopes are arbitrary, their ratios may be compared directly to the ratios of the power-law exponents for red and green saturation scales obtained in the two earlier experiments. The ratio of power-law exponents for median estimates under corresponding conditions in the 4° magnitude estimation study was 1.28; for the 1.5° data, the ratio was 1.00.

One of the observers of the matching study (the senior author) also participated in both scaling studies. Obtained ratios of red to green exponents for this single observer were: 1.20 for scales obtained by magnitude estimation, 4° stimuli; 1.44 for scales obtained by magnitude estimation, 1.5° stimuli; 1.56 and 1.53 estimated from red and green judgments, respectively, in the matching study. For this observer, all three procedures revealed consistent differences between saturation functions for red and green light. If any conclusion at all may be drawn from the data of a single observer, it is, in this case, that the observer may best be used as his own control, and that the existence of measurable differences between saturation magnitudes may be obscured by the selection of median magnitude estimates as the quantitative measure of perceived saturation.

The problem of definition of saturation remains central to the problem of scaling in this dimension. While observers may be capable of making ratio judgments when comparing equiluminant stimuli of the same dominant wavelength, they may not be able to adhere to a ratio criterion when making comparisons between differing hues. While the heterochromatic matching procedure offers a more simple judgmental situation, it does not directly evaluate the form of the saturation scale, and the application of heterochromatic matching data involves the assumption of a particular scale form (in this case, power-law form). The data of magnitude estimations of saturation tend to verify a power-law form for this subjective scale without establishing a definitive value for its exponent, or differentiating between scales for various dominant wavelengths. It is suggested by the data of the present three experiments that validation of scale information based on magnitude estimations requires additional procedures, such as saturation matching, and that reliance on median magnitude estimates of saturation as indicators of the form and parameters of the scale may be misleading.