Vision Research 151 (2018) 88–98

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Color weight photometry

Jan Koenderink[a,b,c], Andrea van Doorn[a,b,c], Karl Gegenfurtner[a,*]

[a] *Abteilung Allgemeine Psychologie, Justus-Liebig-Universität Giessen, Germany*
[b] *Experimental Psychology, Leuven University, KU Leuven, Belgium*
[c] *Experimental Psychology, Utrecht University, The Netherlands*

ARTICLE INFO

Article history:
Received 18 December 2016
Received in revised form 27 May 2017
Accepted 1 June 2017
Available online 19 July 2017

Number of Reviews = 2

Keywords:
Luminance
Brightness
Heterochromatic brightness matching
Color vision

ABSTRACT

We study the "color weight" for a number of rather different paradigms. In well researched heterochromatic photometry methods we find that the "weights" determined by settings of naive observers are closely determined by the CIE luminance functional. This is very different for tasks that involve mid- and high-level aspects of perception. In several cases we find equipollence for the display red, green and blue channels. Moreover, in such cases the very nonlinear maximum-rule fits the data rather better than a linear functional. These findings are of interest when photometry needs to be applied for stimuli that are different from the high temporal and low spatial frequency gratings typical for flicker photometry. These results are relevant for science, ergonomics and art.

1. Introduction

"Luminance" has been one of the great success stories of the 20th century. It denotes the effect that lights of different wavelengths have on the visual system. A standardized version of luminance was established by the CIE in 1924, when it was realized that there was a good method that led to reliable results across different observers. This method was heterochromatic flicker photometry, in which observers minimize the visible flicker of two alternating lights of different wavelengths by adjusting their relative intensity. The CIE 1924 V(λ) function represents the average empirical functions of many observers, amended by other measurements mainly at the short-wavelength end of the spectrum. The major advantage of heterochromatic flicker photometry is the additivity of the resulting luminance values, termed Abney's Law. The luminance of two mixed lights is the sum of its individual luminances. This way, the luminance of any broadband distribution of light can be determined by multiplying it with the V(λ) function and adding over wavelength. The standard observer can thus be incorporated into a physical measuring device, which sometimes makes people forget that the device is measuring a photometric quantity that is based on human visual sensitivity.

Luminance also seems to have a physiological counterpart. The V(λ) function can be described as a weighted addition of the L- and M-cone spectral absorption functions, with the L-cones being weighted by a factor close to 2. There does not seem to be input from the S-cones to luminance. This weighting is similar to the way magno-cellular retinal ganglion cells and geniculate cells add their cone-inputs (Derrington, Krauskopf, & Lennie, 1984; Lee, Martin, & Valberg, 1988). These early-stage neurons primarily underlie the visual system's response at high temporal frequencies, and are now thought to be the physiological substrate for photometric luminance, as determined by flicker photometry. Notably, when visual stimulation equates the luminance of its components to achieve "iso-luminance" or "equi-luminance", some aspects of visual perception behave qualitatively different, in particular motion perception. All of this supports the notion that photometric luminance is firmly grounded in the human visual system.

Despite these overwhelming advantages of using photometric luminance as a way to measure the effectiveness of lights to stimulate the visual system, there are some drawbacks as well. Flicker photometry by definition uses relatively high temporal frequencies, and numerous experiments have shown that the precise way cone inputs are weighted does depend on the temporal and spatial aspects of the stimuli (e.g., Gegenfurtner & Hawken, 1995; Stromeyer, Chaparro, Tolias, & Kronauer, 1997). Even more important is the obvious shortcoming that luminance covers only one – arguably important – aspect of our perceptual experience. It is in strong disagreement with our experienced subjective weighting of different lights. Therefore a similar functional was established based on heterochromatic brightness matching (Commission Internationale de l'Eclairage, 1988). The resulting

[*] Corresponding author.
E-mail address: gegenfurtner@uni-giessen.de (K. Gegenfurtner).

http://dx.doi.org/10.1016/j.visres.2017.06.006
0042-6989/© 2017 Elsevier Ltd. All rights reserved.

spectral sensitivity curve is distinctly different from V(λ), but due to failures of additivity it is not widely used.

On a generic display unit the red, green and blue channels have luminances in (roughly) the ratios R:G:B = 3:6:1 (CIE Proceedings, 1988; Eisner & MacLeod, 1980; Smith & Pokorny, 1987). Since luminance is a linear functional of spectral radiant power density, this implies that a bright blue $r = g = 0$, $b = 1$ will be equiluminant with a "yellow" $r = g = 0.1$, $b = 0$ (where we assume that the maximum of any RGB-channel is 1). Unless in a completely dark room with no other lights present, this "yellow" will look like a dark drab brown, not appearing yellow at all (Buck, 2014). It will *never* balance against the bright blue. The closest "balance" between the primary colors of the display unit for graphical applications appears to be more like R:G:B = 1:1:1, the "equipollent" condition. Indeed, the "principal" colors

$$\mathrm{R} = \{1,0,0\}, \mathrm{Y} = \{1,1,0\}, \mathrm{G} = \{0,1,0\})\mathrm{C} = \{0,1,1\}) \\ \mathrm{B} = \{0,0,1\}, \mathrm{M} = \{1,0,1\} \quad (1)$$

appear to mutually balance each other, implying that the "weight" functional is all but linear, but something more like **max[***r*,*g*,*b***]**. On the display unit white (W = {1,1,1}) is the union of R, G and B, thus these primary colors appear like "parts" of white.

In this study we compare a set of eight mutually quite distinct paradigms that require observers to "balance" all pairs of the principal colors R (red), G (green), B (blue), C (cyan), M (magenta) and Y (yellow). We attempt to fit the results with a linear functional of the type $\mathrm{w_R}$ R+$\mathrm{w_G}$ G+$\mathrm{w_B}$ B (the numerical coefficients to be adjusted appropriately). We focus primarily on color alone. However, it is in no case necessary to explain colorimetric concepts to the naive observers. Indeed, we avoid any discussion of color as much as possible. Our study is aimed at addressing "weight", not "luminance", "saturation", or something like that. The "weight" will typically be in the compositorial sense (Monroe, 1926; Morriss, Dunlap, & Hammond, 1982; Pinkerton & Humphrey, 1974; Wright, 1962; MacManus, Edmonson, & Rodger, 1985; Locher, Overbeeke, & Stappers, 2005; Mokaran, 2007; Parada-Castellano, 2016; Wise & Wise, 1988).

2. Methods

2.1. Display

The display is the screen of an Apple MacBook Pro 15" (mid 2007 model). It was spectrophotometrically calibrated with a Konica Minolta Spectroradiometer CS-2000A (Konica Minolta Holdings Inc., Marunouchi, Tokio, Japan). The luminance of the display white was 317 cd/m^2. The display was linearized using the Bergdesign SuperCal method (To, Woods, Goldstein, & Peli, 2013). The white point was x = 0.312, y = 0.339. The R,G,B primaries of the monitor had CIE 1931 xyY color coordinates of R = {0.5995, 0.3406, 68.9}, G = {0.3259, 0.5723, 197.4}, and B = {0.153, 0.1346, 53.2}. In the following, we will indicate colors in terms of the linear display Red, Green and Blue.

Notice that the relative R, G, B colorimetric coordinates of most monitors are approximately the same. This is no coincidence, as the optimal choice maximizes the volume of the monitor gamut relative to that of the color solid. The optimum can be computed from the CIE color matching functions and the standard daylight spectrum (see Koenderink, 2010). Not surprisingly, all monitors come close. The choice of the primaries R, G, B automatically fixes the C (=G+B), M (=R+B), Y (=R+G) secondaries. Here we use "principal colors" for the set RYGCBM (these are special in the above mentioned sense, we do not imply any relation to opponent colors). The principal colors are also constrained by our visual system. The sum of R and G has to be balanced with respect to the red-green color-opponent mechanism, resulting in a unique yellow. Similarly, the sum of R, G, and B has to be balanced with respect to both opponent mechanisms, resulting in a neutral white.

2.2. Presentation

Observers viewed the display binocularly from a distance of 57cm, using their preferred optical correction when necessary. The full screen measures 20 by 33 degrees of visual angle, although many of the paradigms use only a minor part of that. See Fig. 1 for an illustration of the paradigms.

The presentation software was written in **Processing2+** (http://procesing.org), a variety of Java aimed at artists and designers. It allows full spatial, temporal and colorimetric control for our needs and allows fast development. User interaction was limited to the use of the left–right and up-down arrow keys, whereas the space bar signified user initiated next trial.

Unless mentioned otherwise (see paradigm VIII below), a fixation cross was omitted, indeed, free viewing was considered natural.

2.3. Observers

Seventeen observers participated in the experiment. Ten observers were students of the Justus Liebig Universität at Giessen, Germany. These were predominantly female and in their early twenties. Their color vision was checked by the Ishihara test, all passed. Seven observers were staff members with some knowledge of colorimetry. Where this becomes critical in the analysis it will be mentioned. Our experiment was in agreement with the Helsinki declaration, was approved by the local ethics committee (LEK 2013-0018) and all observers provided informed consent.

3. Experiments

In all cases we seek to determine a ratio of "equal weight" for a pair of principal colors, say F_1 and F_2. It is repeated for all pairs of (distinct) principal colors R, G, B, C, M and Y. Because we want to stay as close to the full-strength principal colors as possible, one color is attenuated, the other kept at full strength.

Thus an equality $aF_1 = F_2$ will be reported as a ratio $R_{12} = 1/a$, whereas the equality $F_1 = \mathrm{a}\ F_2$ is reported as $R_{12} = a$. In the analysis we work with logarithmic representations. This implies $\log R_{12} = -\log a$ in the former and $\log R_{12} = \log a$ in the latter case. Thus the equipollent case $F_1 = F_2$ leads to $\log R_{12} = 0$. The parameter a automatically switches over from one side of the equation to the other as the observer passes equipollence. None of the observers noticed this happening.

In all cases a warning signal (far outside the stimulus area) is provided to notice the observer that one or the other limit of the parameter range was reached.

Observers were instructed to start a trial by looking at either extreme, then use large increments to find the approximate environment of the point of balance, finally switch to small increments to do their setting. In some cases the region of subjective equality is ill defined and all the observer can do is indicate the approximate ratio using the large increments. In other cases a single small increment might be close to a just noticeable difference. Trials are started at random settings of the ratio.

The various paradigms are visited in random order, different for each observer.

3.1. The paradigms

Heterochromatic photometry has a long history, perhaps starting with Abney's work (Abney & Festing, 1886; Burns, Smith,

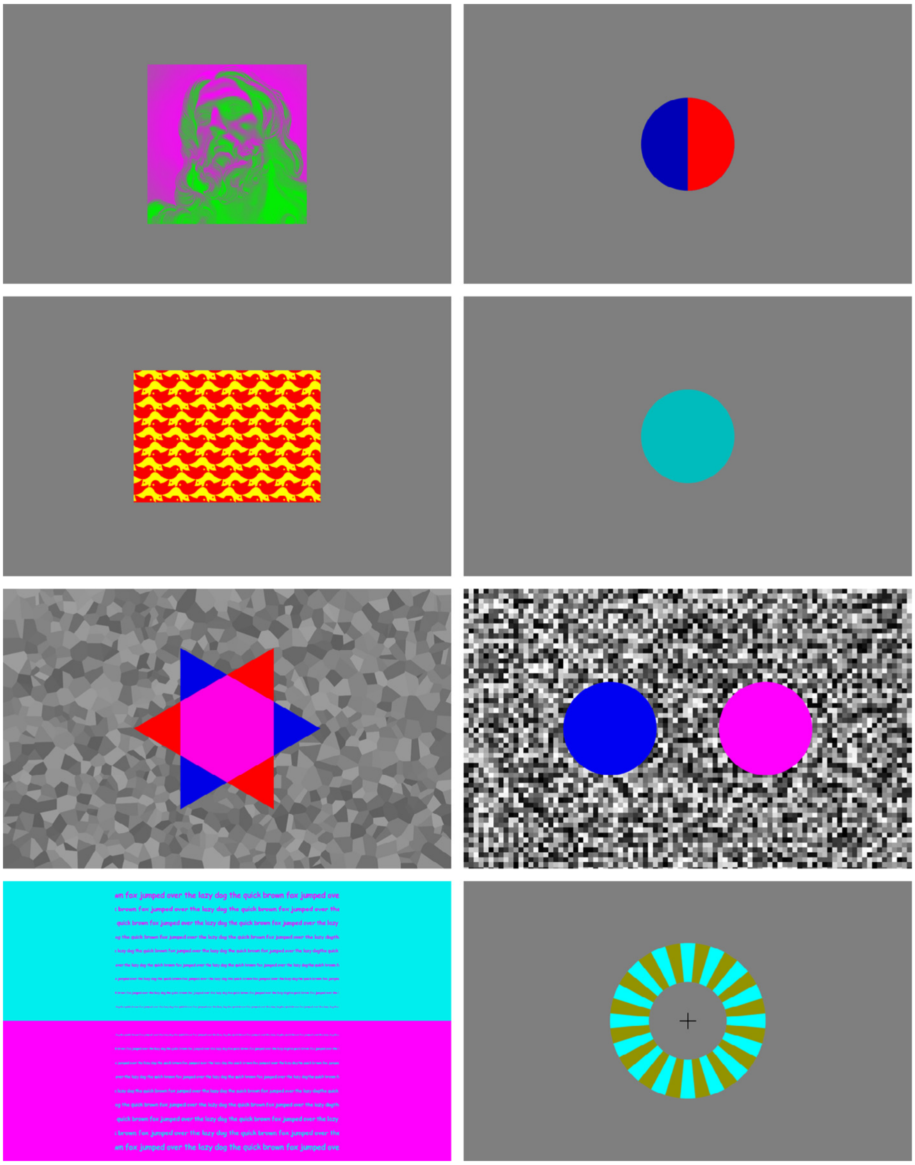
wn fox jumped over the lazy dog the quick brown fox jumped ove
: brown fox jumped over the lazy dog the quick brown fox jumped over the
quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy
og the quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy dogth
: lazy dog the quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy dogthe quick
over the lazy dog the quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy dogthe quick brown f
over the lazy dog the quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy dogthe quick brown f
: lazy dog the quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy dogthe quick
og the quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy dogth
quick brown fox jumped over the lazy dog the quick brown fox jumped over the lazy
: brown fox jumped over the lazy dog the quick brown fox jumped over the
wn fox jumped over the lazy dog the quick brown fox jumped ove

Fig. 1. This figure shows momentary screen grabs for the eight experiments. Although they yield a true view of the geometry (screen size was 20 by 33 degrees of visual angle), they cannot suggest the nature of the dynamic structure. Moreover, they show just one colorimetric setting, whereas the program changes settings from case to case and the observer has continuous control over certain parameter combinations. Thus the actual view may be very different from these examples. From left to right, top to bottom: I. Pictorial content. The two hues alternate at a fairly low rate. At some setting the picture becomes unrecognizable. II. Split-field comparison. The presentation is static. At some settings the figure "looks balanced". III. Figure-ground segregation. The presentation is static. At some setting figure/ground switches or becomes undefined. IV. Minimum flicker. The color of the field alternates at a high rate. One is aware of an "unrest". At some setting the unrest is minimally noticeable. V. Perceptual Grouping. The figures may be seen as a star or as two triangles, usually a left and a right pointing one. At some settings neither the left nor the right pointing "wins". VI. Indirect comparison. Here the two disks are static, but the background very dynamic: the noise pattern is refreshed at frame rate. This renders the gray level of the background ambiguous ("both light and dark"). At some setting the "weight" (in the sense of pictorial composition) of the two disks balances. VII. Legibility The presentation is static. Large type tends to be legible, but small type may be illegible. The observer attempts to find the setting that allows the finest print to be read. VIII. Apparent rotation. This uses a dynamic presentation, so the figure only suggests the geometry. Typically the observer is aware of either clockwise or counterclockwise rotational movement. At some setting the rotation gives rise to a "vibration" and moves neither clockwise, nor counterclockwise.

Pokorny, & Elsner, 1982; Lennie, Pokorny, & Smith, 1993; Wagner & Boynton, 1972; Walsh, 1965). "Color weight" is something else again (Alexander & Shansky, 1976; Bullough, 1907), although at least "related". We use "color weight" in the sense of pictorial composition, related to what has been termed "Prägnanz" by Wertheimer (1923). It is somewhat related to measures of salience used in visions science and computer vision (see Itti, Koch, & Niebur, 1998; Tatler, Hayhoe, Land, & Ballard, 2011 for a recent review and criticism), although both terms are used in various senses. We prefer to use "color weight" in an operational sense, leaving open the possibility that it might be method dependent.

We used the following paradigms, listed in conceptually arbitrary order. We will use the roman numerals I, II…VIII, to refer to these instances in the main text. There are various alternatives that we considered, but eventually did not include (Cooper & Lee, 2014). Several of these methods are in common use, or have at least been described and investigated before. Some methods are apparently novel, such as the indirect comparison on a nondescript background. We make no particular claims as to originality here. We also did not try to emulate existing methods. We did decide on the methods on purely phenomenological grounds. All the stimuli are illustrated in Fig. 1.

3.1.1. Paradigm I: Pictorial content

This is a case on which some literature exists (Cavanagh, 1991; Kindlmann, Reinhard, & Creem, 2002; Livingstone & Hubel, 1987, 1988). A monochrome photograph is default rendered by mapping pixel intensity i (say) on the achromatic color $i\,W + (1 - i)\,K$, where "K" ("key" in the printer's jargon) stands for black (that is {0,0,0}), but it may also be chromatically rendered as $i\,C_1 + (1 - i)\,C_2$, where C_1 and C_2 are *any two distinct colors.* When either $C_1 = K$ or $C_2 = K$ one obtains a "normal" (though "colored") image that is easily recognized for what it is. But for certain choices of C_1 and C_2 it may be very hard to become aware of the pictorial content. This is the case for so-called "equiluminant" images. In the presentation images toggle between a "positive" and a "negative" (in the sense of the silver-based photography of the past). The toggle frequency is 2 Hz.

3.1.2. Paradigm II: Splitfield comparison

This case is familiar as "photometry by direct comparison". The test field is composed of two abutting geometrically identical patches, classically two "half fields" of a circular disk split by its vertical diameter (Boynton & Kaiser, 1968). The conventional task is to "minimize the distinctness of the border", whereas our task is explicitly to "balance the left–right weights". In the case of this paradigm the background was a constant medium gray.

3.1.3. Paradigm III: Figure-ground segregation

A chess board is perceived as a regular pattern of white and black squares. When the squares are filled with distinct colors it may be hard to become aware of the chess board pattern at all. This is the case for so-called "equiluminant" colors (Livingstone & Hubel, 1984), that is to say, this paradigm is one possible operational definition of "equal weights" out of many. In our implementation we use a texture composed of a mosaic of two different shapes, either of which can appear as either figure or ground, much like the familiar Rubin faces–vase figure. At one side of the equilibrium point one sees one pattern, at the other side the other pattern. Near equilibrium there is no evident pattern. Depending on the hues one hardly notices a texture at all, the colors appear to "melt into each other".

3.1.4. Paradigm IV: Minimum flicker

Look at a circular disk filled with one of two colors that toggle at a fairly fast rate (18 Hz frame rate). For certain pairs of colors the alternation becomes hard to notice. This is the classical case of "flicker photometry". In some cases the flicker will completely vanish at some ratio, in others there will remain a hard to describe spatiotemporal "unrest", which has to be minimized. The method is generally preferred because known to lead to linear behavior (de Vries, 1948; Ikeda, 1983; Ives, 1912; Le Sueur, Mollon, Granzier, & Jordan, 2014; Pokorny, Smith, & Lutze, 1989).

The nature of the "unrest" left at the setting of "minimum flicker" is very hard to describe and probably idiosyncratic. One simply has to trust that observers "know" how to minimize it (see Pokorny et al., 1989). The circular flickering patch was presented on a medium gray background.

Residual flicker can be minimised using a phase shift, although it is rarely possible to achieve the impression of a static field. We did not use such an additional control parameter, as it would ill fit our general paradigm. Moreover, it is unclear how to interpret a phase shift in the context of pictorial weight. In practice, no participant had specific problems in minimising the degree of residual unrest in the field.

3.1.5. Paradigm V: Perceptual grouping

Perceptual grouping depends on the balance of colors too (Takahashi, Ohya, Arakawa, & Ishisaka, 2010). We use a configuration of two equilateral triangles each oriented with one side vertical, one triangle pointing to the left, the other to the right. The triangles were superimposed, their barycentra coinciding, thus appearing as a hexagonal star. The intersections define seven areas, a regular hexagon and six small equilateral triangles. The latter we group in sets of three, "belonging" to one of the large equilateral triangles. Depending upon the coloration one becomes immediately aware of an arrow pointing left, or, in other cases of an arrow pointing right. It crucially depends upon the colors of the parts. For some choices of colors it becomes very hard to decide whether the configuration points to the left or to the right. This may be taken as an operational definition of "equal weight" of the colors. Experienced observers might note that the triangles might not point left or right, but in some oblique directions. In practice, none of the participants remarked upon this. We used a random mosaic of gray tones as background, uniformly distributed on (0.25–0.75). Thus the luminance of the background is not precisely defined.

3.1.6. Paradigm VI: Indirect comparison

We present two equally sized circular disks at some separation on a neutral background. The geometry is symmetrical with respect to the frame. The disks are filled with distinct colors. In some cases one notices an unbalance to one side, in other cases an unbalance to the other side. For certain color choices the "composition" may become ambiguous, or "perfectly balanced". In this paradigm a random noise background is used, refreshed at frame-rate. It looks a bit like a snowstorm, having no particular gray-tone, which is the reason for this choice. This is an important condition, since we want to compare the two disks with each other, not each disk with its background. With this type of dynamic background the disks are neither lighter nor darker than the ground. This is the reason for this choice.

3.1.7. Paradigm VII: Legibility

Black type (of some reasonable size) printed on a white page is perfectly legible, so is white type printed on black paper. But type of certain colors printed on certain (different!) backgrounds may become illegible. This may be taken as another operational definition of "equal weight" of the colors. The paradigm seems perhaps somewhat related to the "minimally distinct border" method (Gunther & Dobkins, 2005; Kaiser & Greenspon, 1971; Lindsey & Teller, 1989; Pokorny, Graham, & Lanson, 1968; Pokorny et al., 1989; Schwarz, 1956) and acuity criteria (Ingling, Grigsby, & Long, 1992).

In our implementation we used an array of text set in various font sizes, similar to an acuity test chart. The largest type was legible in all color combinations, the smallest type was never legible. The participants were free to base their setting on the top or bottom half of the display, but were asked to divide their attention about equally.

3.1.8. Paradigm VIII: Apparent rotation

Richard Gregory proposed an ingenious method in which to compare two colors with a pair of light and dark achromatic colors, using the phenomenon of "apparent motion" (Cavanagh & Anstis, 1991; Cavanagh, MacLeod, & Anstis, 1987; Chaudhuri & Albright, 1990; Gregory, 1985; Kaiser, Vimal, Cowan, & Hibano, 1989). This may be taken as yet another operational definition of "equal weight" of the colors. It is usually implemented as a translation, the observer typically being aware of vertical bars, either moving to the left or to the right. At some well defined ratio of the weights the motion apparently stops. In our implementation we used a rotating wheel layout. The observer typically notices clockwise or anti clockwise rotating sectors (instead of bars), except in a "balanced" condition where the wheel ideally would come to a standstill. For this paradigm a fixation mark has to be provided, because in this case small eye movements tend to interfere with the task.

The rotational method implemented here is advantageous because fixation is much easier than in the case of translational motion. Far from equilibrium the rotation movement dominates the awareness, near equilibrium one has a confused experience of "flicker", or "irregular motion reversals". Various participants complained it gave them a headache. However, the equilibrium point could be precisely set by all participants.

In all case the participants used the arrow keys of the keyboard to control the relevant parameter. Pairs UP/DOWN and LEFT/RIGHT allow slow and fast rates of change. The space bar was used to signal completion of a setting. When the parameter reached the upper or lower limit its value was clipped and the participant received a notice in one corner of the screen. After each setting the next presentation appeared with randomly chosen initial parameter value. The sequence of presentations was randomised for each participant.

4. Analysis

Overall, the experiment yields a large dataset, since there are 6 (6 − 1)/2 = 15 pairs of principal colors, thus 15 * 8(number of paradigms) = 120 measurements per observer. With 17 observers this implies 2040 measurements, each measurement represented as a real number (the natural logarithm of the ratio) between minus and plus infinity. In practice the range turned out to be limited to (−0.15,+1.1).

There are numerous ways to analyze such data. We start by discussing the initial analysis for two paradigmatic cases, namely minimum flicker (IV) and indirect comparison (VI). In the first case one expects to find that CIE luminance explains the data, in the second case probably not. We show results of observer #2, but very similar results are encountered for the others.

The raw settings of observer #2, paradigm IV are illustrated in Fig. 2 left. These raw data form the input for all further analysis. Can these ratios be explained through a linear functional $w_R R + w_G G + w_B B$ for some triple of weights (subject to the condition $w_R + w_G + w_B = 1$)? Then one could summarize 15 independent observations through only 2 degrees of freedom, a major gain. Of course, the very existence of such a functional implies that there should be no "intransitive triangles" (Thurstone, 1927), that are cases of ratios a/b, b/c, c/a such that

$$\log[(a/b)(b/c)(c/a)] \neq 0 \tag{2}$$

A numerical check yields 20 intransitive triangles in this case, that is 100% of the possible triples. Of course, this is trivial because exact zeroes never occur in real life, a better measure of the intransitivity is the range or variance of these values. Granted the existence of intransitivities, one needs to fit the function in some "best" sense. It is of some interest to see which triples are mostly involved in such intransitive behavior though. Here R-C-Y is worst, R-G-B next.

We attempt to fit such a functional to the observations by minimizing the sum of squares of the differences of the predictions of a given model with the actual responses. The result is w_R:w_G: w_B = 0.231:0.590:0.179, which happens to be very close to the CIE luminance ratios of the display primaries (0.223:0.602:0.175), as measured with a spectroradiometer. Using these weights one can "explain" all observations. The fit is very good, with R-squared at 0.96. Thus, for observer #2 and paradigm IV, the CIE convention works fine and one might say that our experiment was really superfluous and only affirmed general wisdom.

Of course, (non-)linearity comes in degrees. For instance, one might attempt to fit a more general Minkowski functional such as

$$W(r, g, b) = (w_R r^n + w_G g^n + w_B b^n)^{(1/n)} \tag{3}$$

The best fitting Minkowski exponents have values close to one, scattering a bit, depending upon the observer. (This may be taken as an empirical proof of the CIE luminance notion.) Fig. 3 illustrates the case for observer #2 and the minimum flicker paradigm (IV). Here the best fitting exponent is slightly larger than one, about 1.2.

Next consider the case of indirect comparison (VI) for the same observer #2. Here the raw settings look rather different (Fig. 1 right). In this case (paradigm VI) the transitivity violations are about two times as large as in the previous case (paradigm IV). (For observer #2 the median transitivity violation for paradigm VI is 1.99 as large as that for paradigm IV.) The fit of the linear functional does not so well, the R-squared being only 0.230. In this case a very *nonlinear* functional, **max**[$w_R\, r$, $w_G\, g$, $w_B\, b$], works much better, for R-squared increases to 0.570. The best fitting weights are w_R:w_G:w_B = 0.332:0.361:0.307, very close to equipollent. The difference between these two cases is huge, as Fig. 4 illustrates.

Whereas one finds best fitting Minkowski exponents near to 1 in the flicker case, one finds about 4 or higher in the indirect comparison case. The maximum rule does about as well as a high exponent, whereas an exponent of two ("Guth's vector model" [Guth & Lodge, 1973]) is significantly worse, although better than 1 (the CIE luminance).

Almost all observers yield results very similar to what was shown for observer #2. When analyzing the results of all observers the main choice is to use a kind of "exploratory data analysis", or to

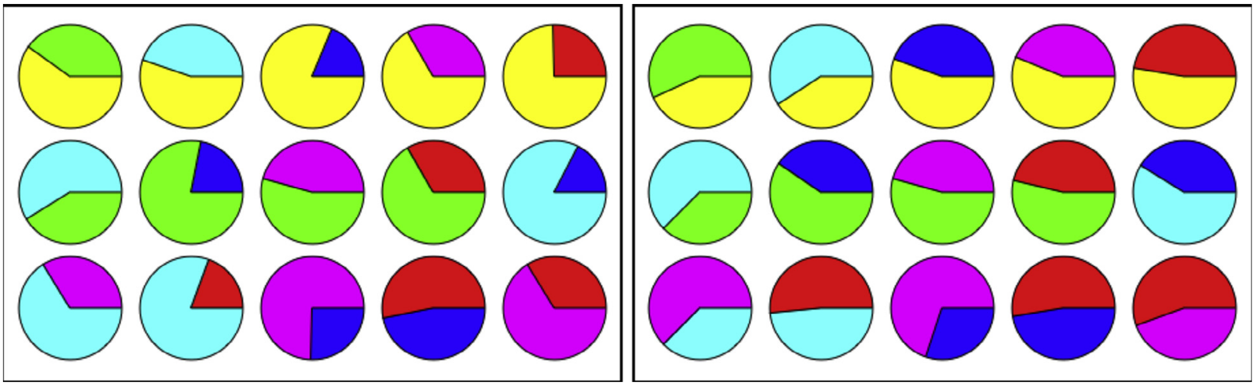

Fig. 2. Left. The raw settings of observer #2 for the flicker paradigm (IV). There are 15 distinct principal color pairs that have been compared. The angular sections of the pie charts indicate the ratios set by the observer. Right. The raw settings of observer #2 for the indirect paradigm VI. One should compare these settings to those in the left figure for the case of flicker (IV). The settings are very different.

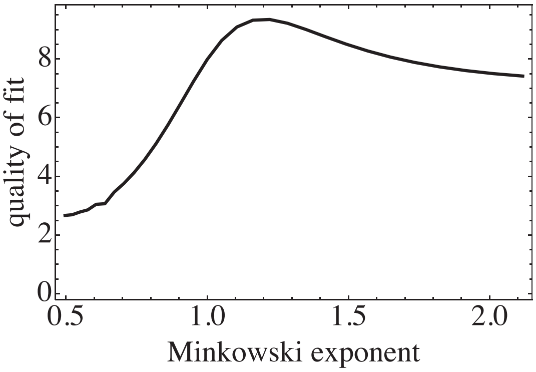

Fig. 3. A search for the best fitting Minkowski exponent (Eq. (5)) yields values close to one for the flicker paradigm (IV; this is observer #2). Of course, it is only to be expected that one often finds values slightly different from one that do a little better than exactly one. For strong conclusions one needs to consider the influence of scatter in the data.

apply a more focussed method akin to "factor analysis". It is found that many different methods eventually converge on very similar, or even identical results. We discuss a method that is focussed and direct, only summarily mentioning similar results that can be obtained by (very) different means.

A major shortcut would be to represent the raw settings of an observer (15 numbers) through a single scalar. This is only possible if one imposes some prior structure though. Here we introduce the notion that the anchor points of the results should be the linear CIE luminance functional and the equipollent, nonlinear (max-rule) condition. This turns out to be reasonable because essentially the same results are obtained using clustering on the raw data or principal component analysis (PCA). Given these empirical findings, the focussed method is to be preferred, since – for instance – the first principal component (PC) is very close to CIE luminance, though inevitably different because of the random variations in the data. That is why focussed methods yield "cleaner" results. Their main drawback is that one might miss unexpected features. We have explored the general methods (like PCA, or clustering) far enough that one can rest assured that nothing of possible interest is ignored.

In the case of the CIE linear luminance functional one expects a certain set of ratio observations, for the equipollent max-rule case one expects all ratios to be essentially equal to unity. Thus one way to compress the data is to project the ratios on the vector predicted from CIE luminance. By suitable normalization one obtains a scalar that will be zero for the equipollent and one for the CIE luminance case. Doing this reduces the data set to a data matrix of 8 (number of chromatic paradigms) by 17 (the number of observers). That this is "reasonable" follows from the fact that the R-squared between the CIE luminance and the first principle component of the raw data is 0.981. The first principle component explains 32.4% of the variance. (Fig. 5 left shows an example for one observer: the observations correlate very well with the luminance prediction.) Apparently the difference between the "focussed" and the "exploratory data analysis" (no prior assumptions) approaches is not so large. However, it needs 7 PCs in order to explain 75% of the variance, so there is quite a bit of non-luminance-related variation. Since the equipollent max-rule prediction is the null-vector, there can be no PC that specifically correlates with it. The unavoidable spread in the observations (Fig. 5 right) causes contributions to all PCs, including the first one.

This is a really major reduction. One expects the matrix elements to be in the range zero to one, perhaps clustering on zero, one, or both zero and one. The actual result is something a little more diffuse, as is evident from the histogram shown in Fig. 6.

The next simplification is to reorder the paradigms and the observers so as to bring the data matrix into a more meaningful, especially simple order. It is achieved through sorting on the sum of matrix elements in rows or columns. The result is shown in Fig. 7. Apparently the choice of paradigms is such that roughly equal amounts of the two different types are present.

The sorting yields a meaningful ordering of paradigms as well as a meaningful ordering of observers, whereas the raw data matrix is in strong disarray. The paradigms are in the order shown in Table 1.

The top of this order is most like the linear CIE luminance functional, the bottom represent cases that are much more like the equipollent, max-rule case. A clustering on the raw data (exploratory data analysis without prior assumptions) yields four clusters. The first cluster contains mainly (92%) paradigms I, VII and IV, the second cluster mainly (72%) paradigms II and VI, the third cluster mainly (69%) paradigms III, VIII and (some) IV, whereas the fourth cluster is 91% paradigm V. Apparently, the first cluster contains the linear cases, the second and fourth clusters the most nonlinear ones. Another way to check this is to look at the fraction of cases in which a linear function fitted the observations better than a nonlinear one. This is 89% for cluster 1, 97% for cluster 2, 49% for cluster 3 and 0% for cluster 4. Thus only cluster 3 (paradigms III,

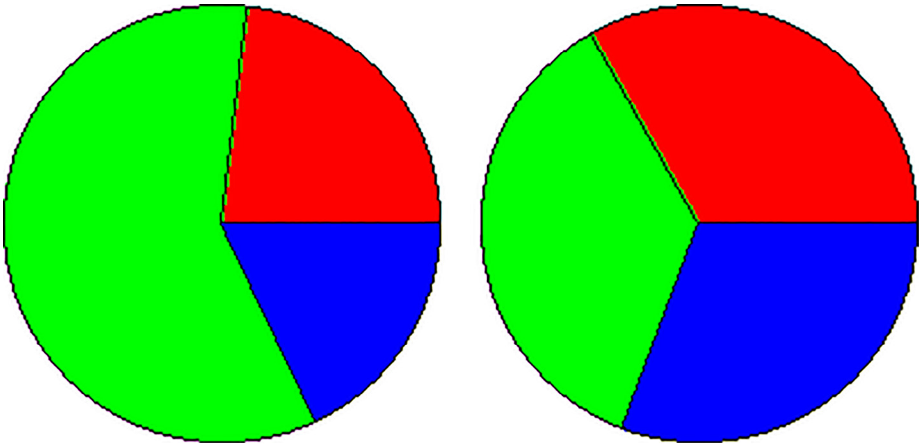

Fig. 4. The best fitting weights for observer #2 in flicker (IV, shown at left) and the indirect paradigm (VI, shown at right). The former is close to the CIE luminance, the latter to the equipollent condition. Moreover, the former applies to a linear, the latter to a very nonlinear functional, the max-rule.

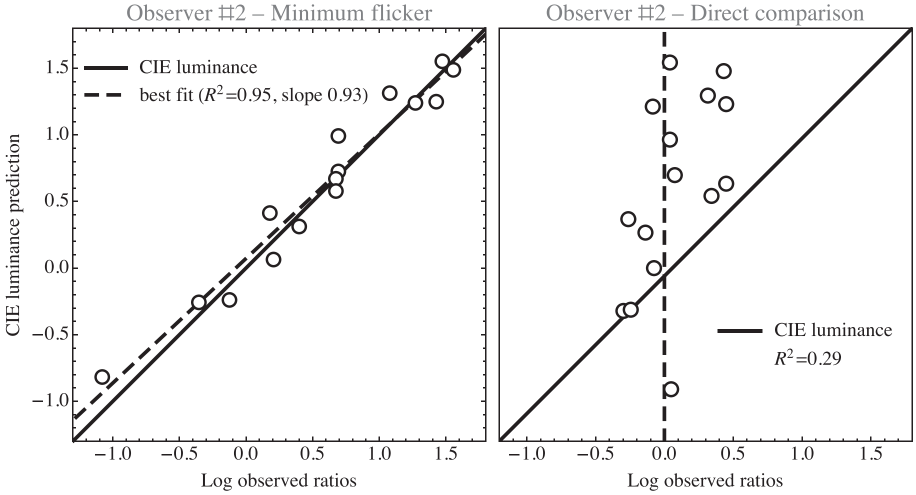

Fig. 5. Regression of the observations against the luminance functional predictor. The log-ratios as determined by the linear CIE luminance functional are plotted on the vertical, the observed log-ratios on the horizontal axis. For the minimum flicker paradigm (shown at left) the data closely follow the CIE luminance as derived from the spectrophotometric data. (Of course, using the best fit linear functional for this observer does even better.) In case of the direct comparison paradigm (shown at right) the R^2 for this observer drops to 0.29. The observed ratios all cluster on zero (the dashed vertical line), that is the equipollent condition.

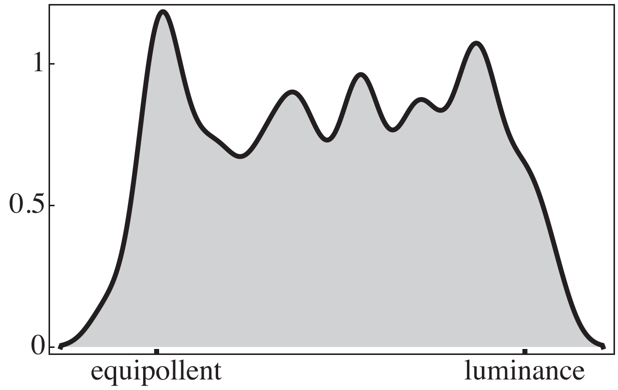

Fig. 6. The smoothed histogram of all (17 * 8 = 136) coefficients of the data matrix. Notice that it is far from normal, but more akin to a uniform distribution. Of course, things are smeared out because all observers and all methods have been pooled. There is already a slight hint that both the equipollent and the CIE luminance limits might prove to be of relevance. (Vertical scale is the probability density.)

Table 1
The paradigms in the order indicated by the doubly sorted data matrix.

1. Legibility (VII)
2. Pictorial content (I)
3. Minimum flicker (IV)
4. Figure-ground segregation (III)
5. Perceptual grouping (V)
6. Apparent rotation (VIII)
7. Direct comparison (II)
8. Indirect comparison (VI)

VIII and IV) can be considered "mixed". The dendrogram shown in Fig. 8 yields additional detail. Clustering on the raw data thus yields results that reflect the order quite well.

The order of observers is also highly relevant, but it makes little sense to list it here because the observers are treated anonymously. In examples it evidently makes a difference whether

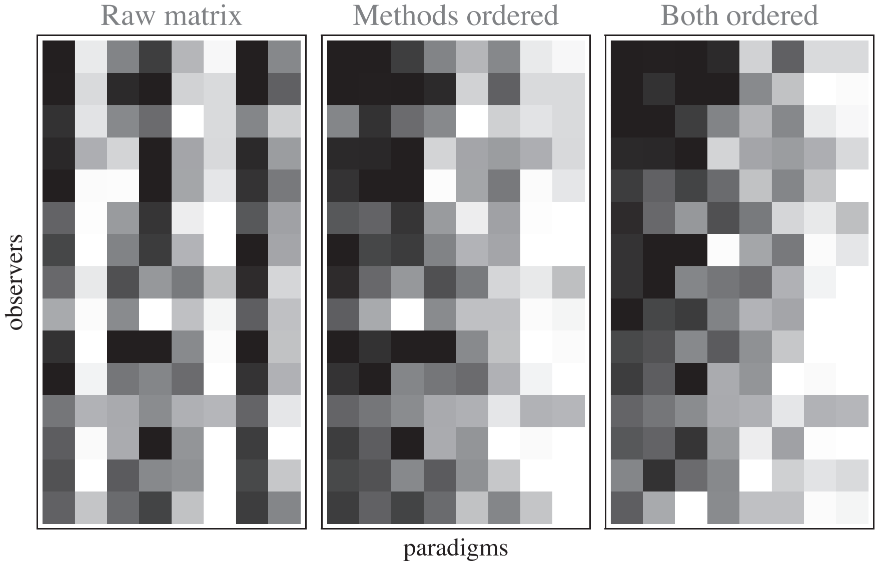

Fig. 7. At left the data matrix. At center the matrix has been sorted by column, followed at right sorted by row. This already shows much of the relevant structure. Each row represents an observer, each column a paradigm. The white-black color scale is centered at the median and runs between the 10% and 90% quantiles. Black means affinity to the CIE linear functional, white to the equipollent max-rule case.

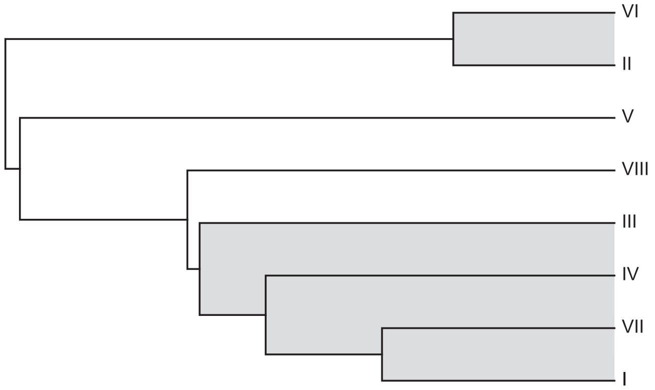

Fig. 8. Dendrogram on the raw data using Ward linking showing some major clusters. The comparison methods (II and VI) split off first, then Perceptual grouping (V). Figure-ground, flicker, legibility and pictorial structure cluster together, as shown by the highlighting.

observers are selected from the top or the bottom of the order. The colorimetric "professionals" prove to be not different from the "naive" observers.

The top and bottom rows and left and right columns of the data matrix summarize much of the results. These are shown as smoothed histograms in Fig. 9.

Yet another way to look at the result is to check the conformity to the CIE linear luminance functional and the equipollent max-rule for all paradigms. A first overview is obtained by studying the observations in the RGB-chromaticity diagram (Fig. 10).

The data suggest that there are significant idiosyncracies, although it is not really possible to distinguish these from random scatter. In order to check for this we had a number of observers repeat settings fifteen times. Because this is very time consuming, the number of principal colors was limited to just red, green and blue and the methods to indirect comparison (VI, strongly nonlinear) and legibility (VII, linear). The result is presented in Fig. 11. The conclusion is clearcut: the spread for individual observers is much less than the spread for the group. Especially the weights for the linear case (VII, legibility) are seen to be somewhat idiosyncratic. The weights for the nonlinear case (VI, indirect comparison) cluster closely on the equipollent point.

The segregation into two different regimes is apparent from the pooled data of all individual settings in Fig. 11. There exists a linear discriminant (e.g., the line connecting yellow and cyan) such that the data for the two methods are fully separated. This implies that the odds against the data to derive from a single distribution are

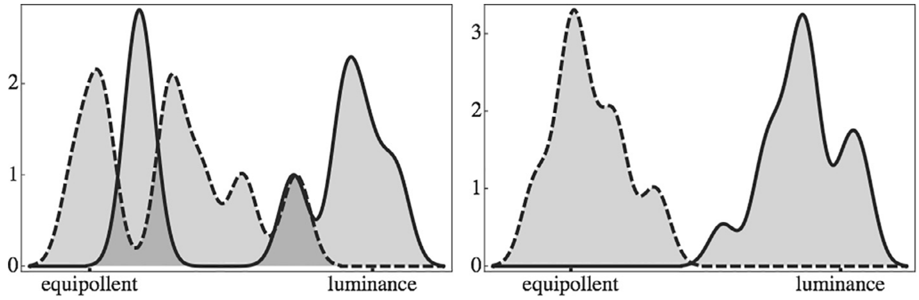

Fig. 9. Left: The top (drawn) and bottom (dashed) rows of the sorted data matrix (the extremes of distribution over methods). Right: The left (drawn) and right (dashed) columns of the matrix (the extremes of distribution over observers). Here the fundamental dichotomy between the linear CIE luminance functional and the equipollent max-rule becomes very explicit. (Vertical scale indicates probability density.)

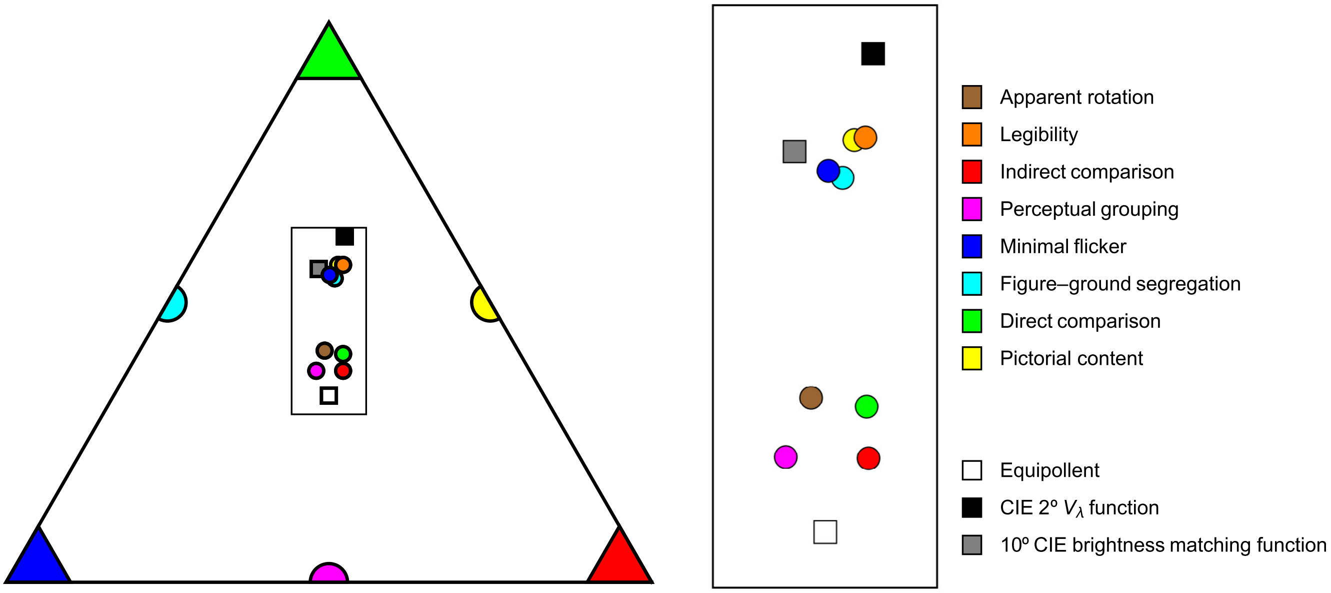

Fig. 10. At left the RGB-weights for each paradigm over all observers. At right the cut-out rectangle (indicated in the RGB-triangle at left) has been magnified.

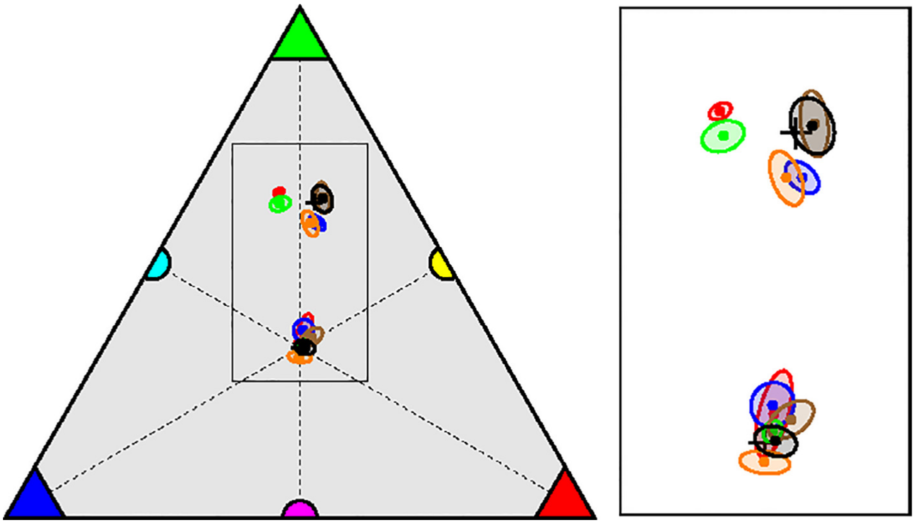

Fig. 11. The spread for settings by individual observers (distinguished by color) in the RGB-chromaticity diagram (left). At right the enlarged region indicated at left as rectangular outline. The covariance ellipses are for one standard deviation. The black crosses indicate the CIE luminance prediction and the equipollent point. The results for paradigm VII (legibility) cluster around the former, those for paradigm VI (indirect comparison) about the latter. For analysis purposes we arbitrarily set the height of the triangle to have unit length.

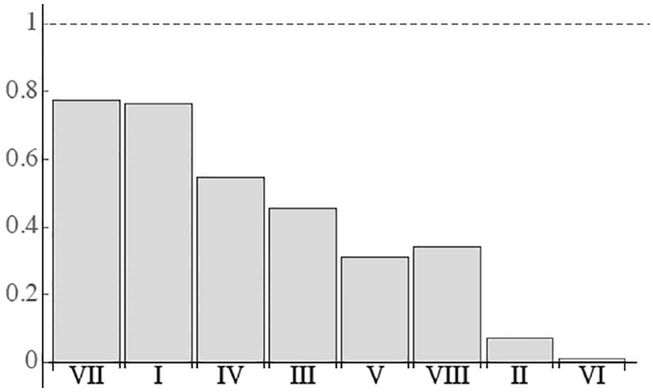

Fig. 12. The fraction of variance of the raw data (log-ratios) explained by the CIE linear luminance functional derived from the spectrophotometric display calibration. The paradigms are in the order established above. Apparently, the fraction explained decreases almost monotonically in this order. Left to right: legibility (VII), pictorial content (I), minimum flicker (IV), figure-ground segregation (III), Perceptual grouping (V), apparent rotation (VIII), direct comparison (II), indirect comparison (VI).

6.18×10^{26}. This is overwhelming evidence that the distributions for the two methods are perhaps not the same. Next consider the case per observer in more detail. At a first, coarse, estimate, the distances in the chromaticity diagram between the means of their settings for the two methods ranges from 0.317 to 0.439 (median 0.410), whereas the spreads range from 0.014 to 0 0.036 (median 0.025). The ratios of distance over standard deviation (z-scores) range from 12.1 to 21.8 (median 14.7). This strongly suggests that the segregation between the two regimes extends to each individual observer. These numbers are perhaps somewhat suspect, because the chromaticity diagram has the structure of the projective plane (no Euclidian metric) and the individual settings are not necessarily normally distributed. A method based on rank orders is perhaps to be preferred. In comparing two point sets one such statistics uses the frequency distribution of in-cluster distances to that of the between-cluster distances (Liu & Modarres, 2011). The standard Mann-Whitney test can then be used to compare these distributions. For the comparison of methods per observer we found that the ratio of the medians of the distances varies between 7.64 and 12.8 (median 9.22). The Mann-Whitney tests yielded p-Values that are all less than 0.0001. Thus we can indeed extend the above finding to all individual observers. This raises the question whether the observers are mutually different? The same non-parametric method can be used for this. We found that in the vast majority of cases the p-values were smaller than 0.001, indicating that the observers are indeed mutually different. There was only one exception when the p-value was larger than 0.05. The overall conclusion is that there exist idiosyncratic differences between observers but that all observers are very different (in the same way) for the two tasks VI and VII. Therefore it is highly likely that increasing the number of repeats per observer would not change this conclusion.

The conformity to the CIE linear luminance functional can be checked by looking at the difference between the actual observations and the prediction by the CIE luminance functional, pooled over all observers for any given paradigm. This is a very conservative test, since any observer apparently has “best fitting” weights different from the colorimetric weights. The latter have been derived from the spectrophotometric calibration of the display unit and the CIE 1961 xyz tables. For some paradigms this explains more than three-quarters of the variance, for others as little as one percent (Fig. 12). In the order established above one obtains a near monotonic relation (Kendall’s tau 0.93). This neatly corroborates the interpretation.

For the equipollent max-rule such an analysis is not possible since the prediction would be that all observations – the log-ratios – should be identically zero. Here the luminance prediction should suggest a certain pattern that simply would fail to apply. A way to check this is to look at the ranges of the raw data and the difference between the raw data and the luminance prediction (Fig. 13). Indeed, one encounters exactly the expected pattern. For paradigms that conform to the CIE luminance the data range is large and subtracting the prediction brings them down to a low level. In contradistinction, for paradigms that do not conform to the CIE luminance the data range is small and subtracting the prediction puts them up to a high level. Again, the distribution over the paradigms is close to monotonic (Kendall tau 0.86).

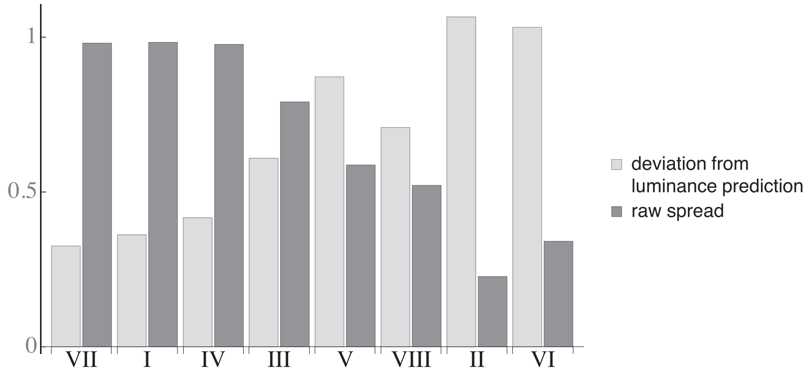

Fig. 13. The interquartile ranges of the raw data (dark gray bars) and the differences between the raw data and the CIE luminance prediction (light gray bars). The paradigms are in the order established above. At the left side almost all of the range of the raw data is due to luminance. At the right end of the scale subtracting the luminance prediction increases the scale. The range of the raw data is probably simply random variation. Notice the almost monotonic dependence on the order.

5. Conclusions

As evident from the formal analysis, there exists a clearcut dichotomy between the linear CIE functional and the equipollent max-rule. This is apparently the difference between colorimetry proper (in the conventional CIE sense) and what graphical artists consider "simple eye measure". This suggests that there is a very real need for a variety of "color science" that deviates from conventional scientific (that is colorimetric) wisdom. As mentioned in the introduction, this is hardly new, nor surprising, but – perhaps unfortunately – it fails to be generally recognized in vision science. Note that a similar "max-rule" had been proposed by Land for his L, M, S channel normalization (see Land, 1964; for review, see McCann, 2017).

There remain a few observations that appear somewhat surprising and perhaps not without interest. One is that the simple max-rule beats the familiar proposals for nonlinear brightness functions, such as Guth's "vector model", for our data set. More importantly, the dichotomy is not perfect. There appears to be a spectrum of results that depend upon the particular operationalization of "color weight". Of course, it is common knowledge that the various methods usually designated "heterochromatic photometry" mutually diverge widely. Some are linear, some are not, and that to various degrees. The major advantage of additive measures such as luminance is, of course, that they can be directly computed from the wavelength distribution of lights. This is not possible with the non-linearities that are present in some of our color-weight measures.

We propose that this indicates an interesting field of endeavor, although it is usually considered a nuisance, the idea being that "correct" methods should all converge to luminance. However, "luminance" is perhaps more of a convenience than a fact, or – perhaps a happier formulation – a singular perspective. Naive observers couldn't care less, they simply go on their guts, with a wide spectrum of results.

Conventional wisdom in colorimetry would probably single out flicker photometry or minimally distinct border as the "most correct" methods. Indeed, these turn out to yield results that are close to the CIE linear functional. In recent times the method proposed by Gregory (our VIII "apparent rotation") has been added and is widely used because indeed very convenient and reliable. Direct comparisons have always been regarded with suspicion. Indeed, they end up closely to the equipollent max-rule. A priori, we expected pictorial structure, figure-ground and perceptual grouping to be close to the equipollent max-rule regime. This is not fully born out by the present results though.

As a matter of course, all these paradigms are sensitive – sometimes even *very* sensitive – to the particular parameter choices. Apparently there lies a wide area of investigation open here.

"Color weight" and CIE "luminance" are categorically distinct concepts. This may be considered an empirical fact as demonstrated by the present data. The bottom line is that given the application one should select an appropriate operational definition of "color weight". CIE luminance is not the universal panacea. Our results do show that heterochromatic photometry is in principle possible and clearly distinct from luminance.

Funding

This work was supported by Deutsche Forschungsgemeinschaft (DFG) grant Ge-879/9, DFG Sonderforschungsbereich SFB TRR 135 and BMBF-NSF grant 01GQ1304. Jan Koenderink was supported by a Humboldt-Award from the Alexander-von-Humboldt-Founda tion. The presentation program, data and Mathematica analysis methods presented in this article are shared and freely available at zenodo.org: doi:10.5281/zenodo.XXXX.

References

Abney, W., & Festing, E. R. (1886). Colour photometry. *Philosophical Transactions of the Royal Society of London, 177*, 423–456.
Alexander, K. R., & Shansky, M. S. (1976). Influence of hue, value and chroma on the perceived heaviness of colors. *Perception Psychophysics, 19*, 72–74.
Boynton, R. M., & Kaiser, P. K. (1968). The additivity law made to work for heterochromatic photometry with bipartite fields. *Science, 116*, 366–368.
Buck, S. L. (2014). Dark versus bright equilibrium hues: Rod and cone biases. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 31*, 75–81.
Bullough, E. (1907). On the apparent heaviness of colours. *British Journal of Psychology, 2*, 111–152.
Burns, S. A., Smith, V. C., Pokorny, J., & Elsner, A. E. (1982). Brightness of equal luminance lights. *Journal of the Optical Society of America, 72*, 1225–1231.
Cavanagh, P., & Anstis, S. M. (1991). The contribution of color to motion in normal and color-deficient observers. *Vision Research, 31*, 2109–2148.
Cavanagh, P., MacLeod, D. I. A., & Anstis, S. M. (1987). Equiluminance: Spatial and temporal factors and the contribution of blue-sensitive cones. *Journal of the Optical Society of America A, 4*, 1428–1438.
Cavanagh, P. (1991). What's up in top-down processing. In A. Gorea (Ed.), *Representations of vision: Trends and tacit assumptions in vision research* (pp. 295–304). Cambridge: Cambridge Univ Press.
Chaudhuri, A., & Albright, T. D. (1990). A new technique for estimating chromatic isoluminance in humans and monkeys. *Visual Neuroscience, 5*, 605–608.
CIE Proceedings (1988). 2° spectral luminous efficiency function for photopic vision. In CIE Proceedings, Paris (pp. 1–11).
Commission Internationale de l'Eclairage (1988). Spectral Luminous Efficiency Functions Based upon Brightness Matching for Monochromatic Point Sources, 2° and 10° Fields, Wien, CIE Publ. 075–1988.
Cooper, B., & Lee, B. (2014). Independence and interaction of luminance and chromatic contributions to spatial hyperacuity performance. *Journal of the Optical Society of America A, 31*, 394–400.

de Vries, H. L. (1948). The luminosity curve of the eye as determined by measurements with the flicker photometer. *Physica, 14*, 319–348.

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology, 357*(1), 241–265.

Eisner, A., & MacLeod, D. I. A. (1980). Blue-sensitive cones do not contribute to luminance. *Journal of the Optical Society of America, A70*, 121–123.

Gegenfurtner, K. R., & Hawken, M. J. (1995). Temporal and chromatic properties of motion mechanisms. *Vision Research, 35*(11), 1547–1563.

Gregory, R. (1985). Movement nulling: For heterochromatic photometry and isolating channels for 'real' and 'apparent' motion. *Perception, 14*, 193–196.

Gunther, K. L., & Dobkins, K. R. (2005). Induction effects for heterochromatic brightness matching, heterochromatic flicker photometry, and minimally distinct border: Implications for the neural mechanisms underlying induction. *Journal of the Optical Society of America, A22*, 2182–2196.

Guth, S. L., & Lodge, H. R. (1973). Heterochromatic additivity, foveal spectral sensitivity, and a new color model. *Journal of the Optical Society of America, 63*, 450–462.

Ikeda, M. (1983). Linearity reexamed for flicker photometry by the summation-index method. *Journal of the Optical Society of America, 73*, 1055–1061.

Ingling, C. R., Grigsby, S. S., & Long, R. C. (1992). Comparison of spectral sensitivity using heterochomatic flicker photometry and acuity criterion. *Color Research & Application, 17*, 187–196.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 11*, 1254–1259.

Ives, H. E. (1912). Studies in the photometry of lights of different colours. I. Spectral luminosity curves obtained by the equality of brightness photometer and the flicker photometer under similar conditions. *Philosophical Magazine, 24*, 149–188.

Kaiser, P. K., & Greenspon, T. S. (1971). Brightness difference and its relation to the distinctness of border. *Journal of the Optical Society of America, 61*, 962–965.

Kaiser, P. K., Vimal, R. L. P., Cowan, W. B., & Hibano, H. (1989). Nulling apparent motion as a method for assessing sensation luminance: An additivity test. *Color Research & Application, 14*, 187–191.

Kindlmann, G. L., Reinhard, E., & Creem, S. (2002). Face-based luminance matching for perceptual color-map generation. In Proceedings of the conference on Visualization 02, IEEE Comput. Soc., 299–306.

Koenderink, J. J. (2010). *Color for the sciences*. Cambridge, USA: MIT Press.

Land, E. (1964). The Retinex. *American Scientist, 52*, 247–264.

Le Sueur, H., Mollon, J. D., Granzier, J., & Jordan, G. (2014). Counterphase modulation photometry: Comparison of two instruments. *Journal of the Optical Society of America, A31*, 34–37.

Lee, B. B., Martin, P. R., & Valberg, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *The Journal of Physiology, 404*(1), 323–347.

Lennie, P., Pokorny, J., & Smith, V. C. (1993). Luminance. *Journal of the Optical Society of America, A10*, 1283–1293.

Lindsey, D. T., & Teller, D. Y. (1989). Influence of variations in edge blur on minimally distinct border judgments: A theoretical and empirical investigation. *Journal of the Optical Society of America, A6*, 446–458.

Liu, Z., & Modarres, R. (2011). A triangle test for equality of distribution functions in high dimensions. *Journal of Nonparametric Statistics, 23*(3), 605–615.

Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience, 4*, 309–356.

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for perception of form, color, movement and depth. *Journal of Neuroscience, 7*, 3416–3468.

Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement and depth: Anatomy, physiology and perception. *Science, 240*, 740–749.

Locher, P., Overbeeke, K., & Stappers, P. J. (2005). Spatial balance of color triads in the abstract art of Piet Mondrian. *Perception, 34*, 169–189.

MacManus, I. C., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology, 76*, 311–324.

McCann, J. J. (2017) Retinex at 50: Color theory and spatial algorithms, a review. Journal of Electronic Imaging. 0001;26(3):031204. doi:10.1117/1.JEI.26.3.031204.

Mokaran, M. A. (2007). *Visual balance in engineering design for aesthetic value*. Thesis: University of Saskatcewan.

Monroe, M. (1926). The apparent weight of color and correlated phenomena. *American Journal of Psychology, 36*, 192–206.

Morriss, R. H., Dunlap, W. P., & Hammond, S. E. (1982). Influence of chroma on spatial balance of complementary hues. *American Journal of Psychology, 95*, 323–332.

Parada-Castellano, R. (2016). Study of Balance of Images Using Visual Weight. *Color Research & Application, 41*, 175–187.

Pinkerton, E., & Humphrey, N. (1974). The apparent heaviness of colours. *Nature, 250*(5642), 164–165.

Pokorny, J., Smith, V. C., & Lutze, M. (1989). Heterochromatic modulation photometry. *Journal of the Optical Society of America, A6*, 1618–1623.

Pokorny, J., Graham, C. H., & Lanson, R. N. (1968). Effect of wavelength on foveal grating acuity. *Journal of the Optical Society of America, 58*, 1410–1414.

Schwarz, F. (1956). Weitere Untersuchungen über den Einfluss der Farbe auf Sehscharfe und Sehleistung. *Albrecht Von Graefe's Archiv Fur Ophthalmologie, 157*, 534–539.

Smith, V. C., & Pokorny, J. (1987). Is there a luminance channel? *Farbe, 34*, 123–128.

Stromeyer, C. F., 3rd, Chaparro, A., Tolias, A. S., & Kronauer, R. E. (1997). Colour adaptation modifies the long-wave versus middle-wave cone weights and temporal phases in human luminance (but not red-green) mechanism. *The Journal of Physiology, 499*(Pt 1), 227.

Takahashi, S., Ohya, K., Arakawa, K., & Ishisaka, Y. (2010). Perceived Strength of Edge, Depth and Brightness of the Kanizsa Illusion as a Function of the Color Contrast between Figures and Background. *Gestalt Theory, 32*(2), 155–166.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*, 5.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34*, 278–286.

To, L., Woods, R. L., Goldstein, R. B., & Peli, E. (2013). Psychophysical contrast calibration. *Vision Research, 90*, 15–24.

Wagner, G., & Boynton, R. M. (1972). Comparison of four methods of heterochromatic photometry. *Journal of the Optical Society of America, 62*, 1508–1515.

Walsh, J. W. T. (1965). *Photometry* (3rd ed.). New York: Dover.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung: Zeitschrift für Psychologie und ihre Grenzwissenschaften, 4*, 47–58.

Wise, B. K., & Wise, J. A. (1988). The human factors of color in environmental design: A critical review. NASA contractor report 177498.

Wright, B. (1962). The influence of hue, lightness and saturation on apparent warmth and weight. *American Journal of Psychology, 75*, 232–241.