Maximum likelihood conjoint measurement of lightness and chroma

MARIE ROGERS,¹,* KENNETH KNOBLAUCH,² AND ANNA FRANKLIN¹
¹The Sussex Colour Group, School of Psychology, University of Sussex, Brighton, UK
²Inserm U1208, Stem Cell and Brain Research Institute, Department of Integrative Neurosciences, Bron; Université Lyon 1, Lyon, France
*Corresponding author: M.Rogers@sussex.ac.uk

Received 30 September 2015; revised 19 December 2015; accepted 22 December 2015; posted 23 December 2015 (Doc. ID 251163); published 8 February 2016

Color varies along dimensions of lightness, hue, and chroma. We used maximum likelihood conjoint measurement to investigate how lightness and chroma influence color judgments. Observers judged lightness and chroma of stimuli that varied in both dimensions in a paired-comparison task. We modeled how changes in one dimension influenced judgment of the other. An additive model best fit the data in all conditions except for judgment of red chroma where there was a small but significant interaction. Lightness negatively contributed to perception of chroma for red, blue, and green hues but not for yellow. The method permits quantification of lightness and chroma contributions to color appearance. © 2016 Optical Society of America

OCIS codes: (330.0330) Vision, color, and visual optics; (330.5020) Perception psychology; (330.5510) Psychophysics.

http://dx.doi.org/10.1364/JOSAA.33.00A184

1. INTRODUCTION

A. Dimensions of Color

Color appearance can be specified in a three-dimensional perceptual space, composed of the dimensions lightness, hue, and saturation (or chroma). Perceptual color spaces, such as those defined by the CIE or the Munsell color system, attempt to characterize lights and surfaces according to dimensions such as these so that the space approximates perceptual uniformity [1,2]. In this paper, we seek to evaluate whether stimulus variation along one dimension influences perception along another. Specifically, whether physical changes in a dimension influence perceptual changes. The results of such an investigation may be complex, for example, if the underlying psychological dimensions prove to interact nonadditively and lack independence. When this occurs, the dimensions are called “integral” as opposed to “separable” [3]. In color, such interaction has been demonstrated for similarity judgments in which observers’ classifications of stimuli varying in lightness and chroma were found to depend upon overall similarity rather than that predicted by their dimensional components [4]. It would be expected, then, that stimulus dimensions that are “integral” cannot be processed separately, or disentangled in perception [5,6], even when integration of information from multiple dimensions hinders task performance [7].

Several methods have been exploited to investigate the dimensional structure of color perception. For example, the method of direct estimation [8,9] requires that observers explicitly choose stimuli of equal spacing along a continuum. This is a highly subjective task, and observers show considerable variability when the experiment is repeated. It is also prone to task-unrelated bias, such as that the difference between a given pair of stimuli influences judgments of a subsequent pair [10]. Multidimensional scaling has also been used to investigate the interaction of color dimensions (e.g., [4,11–15]). This technique uses similarity or difference judgments or ratings between stimuli to determine the number of dimensions necessary to describe the perceptual space of the stimuli and the distribution of stimuli in that space. However, it does not specify what the actual dimensions are [10].

Given the limitations of previous methods, there is interest in techniques that can model how multiple perceptual dimensions interact to constrain observers’ judgments. It would be particularly useful to be able to quantify individual differences as there is evidence that some groups of people may weight color dimensions differently to others. For example, an analysis of 6-month-old infants’ looking time data from four previous studies led the authors to hypothesize that prelinguistic infants’ color preferences ignore lightness, and only pay attention to hue [16]. There is also evidence that experience with color dimensions may influence observers’ ability to separate dimensions. For example, color scientists are better than nonexperts at extracting information about one dimension of a color, without being influenced by the other dimensions [4]. Furthermore,
participants trained to categorize a new hue-based color boundary (not pre-existing in the basic color term lexicon) also extended their learning to lightness-based distinctions. This may reflect the integrality of these two dimensions, i.e., attending to hue variation in the presence of (task unrelated) lightness variation resulted in sensitization to lightness distinctions [4].

B. Conjoint Measurement

Conjoint measurement is a psychophysical method designed to investigate how specific dimensions contribute to perception [17–20]. It is based on paired comparisons of stimuli covarying independently along two or more dimensions. The conjoint measurement approach is useful because it allows the investigation of more than one variable, the construction of measurement scales and how the variables combine in perception [17].

Ho, Landy, and Maloney [21] recently recast the decision process of the conjoint measurement task within a signal detection framework, which, because of explicit assumptions about noise in the judgment process, permitted the perceptual scales underlying the judgments to be estimated by a maximum likelihood procedure, yielding maximum likelihood conjoint measurement (MLCM). Ho et al. [21] showed that the data could be fit with a series of three nested models, corresponding to three different decision rules of how observers combine the responses of the dimensions: (i) the independence model, in which the judgments depended on only one of the dimensions; (ii) the additive model, in which the judgments depended on an additive combination of component response functions from each dimension; and (iii) the saturated model, in which an interaction term was needed to model the judgments beyond the additive contributions of each dimension. An added benefit of this approach is that the estimated scales can be specified in terms of the signal detection parameter $d^\prime$ [22]. The method allows the researcher to determine how two or more physical dimensions interact to influence the observer’s judgments along a perceptual dimension. For example, it has been used to investigate mutual influences of physical surface roughness and glossiness on perception of texture [21,23] and to measure how the strength of the watercolor effect is influenced by several stimulus dimensions [24]. Similarly, the current study manipulates physical lightness and chroma to investigate their combined influences on the perception of lightness and chroma.

C. Current Study

In the current study we used MLCM to investigate how chroma and lightness contribute to chroma and lightness judgments for several hues. In order to maximize the possibility of interactions between these two dimensions and to simplify eventual interpretations, we sought to use stimuli that extended over a similar perceptual range and that were equally spaced perceptually along each dimension. To accomplish this, we conducted a preliminary experiment using maximum likelihood difference scaling (MLDS, [22,27,28]). MLDS is a psychophysical procedure used to estimate perceptual scales for stimuli distributed along a single physical continuum. The technique requires observers to make judgments comparing the perceptual intervals between pairs of stimuli (e.g., between which pair of stimuli is the difference greatest?). The signal detection model and maximum likelihood estimation procedure yield a scale with interval properties, i.e., equal scale intervals are perceptually equal. Since the task requires observers to report which interval between pairs of stimuli is greater and not simply which stimulus is stronger, as in discrimination experiments using paired comparisons, it allows larger stimulus difference to be evaluated. MLDS has previously been used in color research to equate color differences perceptually, specifically to ensure equivalence of perceptual differences between targets and distractors for a visual search task [29]. It has also been used as an independent measure of the perceived difference between stimuli [30], as so-called uniform color spaces are notoriously nonuniform [31,32]. Interestingly, it was found that perceived differences estimated by MLDS better predicted visual search performance than color category boundaries, contrary to previous work without this control.

In the second experiment, we used MLCM [21,22] to quantify the “contamination” of lightness when making judgments about chroma, and vice versa. In separate experiments, participants were asked which of two stimuli (varying independently in lightness and chroma) were (i) lighter or (ii) more chromatic (phrased as “redder,” “yellower,” “greener,” or “bluer” depending on the condition). The contribution of the dimensions in this task may vary between individuals depending on how the instructions are understood. The model assumes that the human visual system calculates a perceived lightness (or chroma) for the stimuli, which could depend on responses to both their physical lightness and chroma, and that the observer uses this information to make a judgment. Here, lightness and chroma are described as “physical dimensions,” despite the fact that units in color space are often described as “perceptual.” This is to distinguish between the manipulations of the stimuli by the experimenter in the physical world, and the internal psychological/perceptual response of the observer. Experiments were conducted for four hues (red, yellow, blue, and green), to determine whether there are different patterns of lightness and chroma interaction for different hues. We found that, generally, lightness and chroma interact in an additive manner when making judgments about the lightness or chromaticity of a pair of stimuli.

2. EXPERIMENT 1: METHODS

We designed the first experiment to choose stimuli to be used in the subsequent experiment. We used MLDS [27,28] to
estimate four perceptually equidistant levels along the dimensions of chroma (when hue was set to red, yellow, blue, or green) and achromatic lightness, spanning a perceptually equal range across dimensions.

### A. Observers

Ten observers participated in the experiment (female = 7; mean age = 25.3, SD = 2.83, range = 23–32). Five observers completed lightness, red chroma, and yellow chroma judgments. An additional five observers completed green chroma and blue chroma judgments. All were assessed as having normal color vision using Ishihara plates [33] and the Lanthony test [34].

### B. Apparatus

Stimuli were presented on a 22 in. 2070SB Mitsubishi Diamond Plus Diamondtron CRT monitor, with a resolution of 1600 x 1200 pixels, 24 bit color resolution and a refresh rate of 100 Hz. Experiments were performed in a dark room. The experiment was run using custom software written in MATLAB R2012b and the Psychophysics Toolbox extensions [35–37]. The monitor was calibrated using a ColorCAL colorimeter (Cambridge Research Systems).

### C. Stimuli

Stimuli were specified in CIE LChuv, as chroma and lightness can be varied independently in this space. CIE LChuv is a cylindrical version of CIELUV, where L is the lightness, C* is the chroma, and h° is the hue [38]. Stimuli varied along the dimensions of lightness or chroma. For chroma, there were 10 levels specified at four fixed levels of hue (red, yellow, blue, or green), with lightness fixed at L = 50. For lightness, there were also 10 levels specified with a fixed chroma (achromatic; C = 0). Hues angles were selected on the basis of the category exemplars for English speakers [39]. For the lightness condition, chroma = 0 and hue = 0. In the chroma condition, red = 14.3°, yellow = 80.2°, blue = 234.3°, and green = 143.2° hue angle in CIE LChuv space. Red chroma had a higher range because the monitor gamut was much wider at the specified lightness level (L = 50), whereas for yellow, blue and green, higher chroma values were not possible within the gamut. A gray background [xyY (1931): 0.31271, 0.32902, 0.15370] was used throughout the experiment.

### D. Design and Procedure

On each trial of the experiment, three stimuli (a triad) were presented, horizontally in a row on the monitor. Observers were asked to judge whether the left or right stimulus was more similar to the middle stimulus. The stimuli varied along either the dimension of lightness or chroma (with four hue conditions: red, yellow, blue, or green). The values along each dimension were chosen from a set of 10 predetermined values indicated in Table 1.

The triads were always ordered in intensity from left to right on the screen, randomly assigned to ascending or descending order along the dimension (e.g., either levels 1 2 3 or 3 2 1, but never 1 3 2). Every element in a triad differed in stimulus level; that is, there were no repeats of the same level within a triad (e.g., no 1 2 2). For 10 levels along each dimension, there are 120 unique trials.

A trial began with presentation of a central fixation cross for 200 ms, followed by a triad presented until the observer responded, thereby initiating the next trial. Observers were instructed to judge whether the left or right stimulus was more similar to the middle stimulus. They responded by pressing the left or right button on a button box. Observers’ responses were coded as their choice of left or right (0/1).

One group of observers (N = 5) judged red, yellow, and achromatic triads, resulting in 360 trials. A second different group of observers (N = 5) judged blue and green triads, resulting in 240 trials. Trials were randomized, interleaved, and presented in one session with blocks to allow for breaks.

### EXPERIMENT 1: MODEL

The analysis is described in detail elsewhere [22]. A difference scale was estimated from each session using functions from the MLDS package [27] in the open-source software R [40], and the scales from individual observers were averaged to obtain means and standard errors for the estimated scale values.

The analysis is derived from a signal detection model in which the observer’s judgments depend on a decision variable based on comparing perceptual intervals between pairs of stimuli, e.g., with stimulus triad (a, b, c), intervals (a, b) and (b, c). It is assumed that on each trial the decision variable is perturbed by Gaussian noise with variance . Intuitively, if for a given stimulus b, the observer is equally likely to choose either of the two stimulus intervals, the perceptual intervals between the two pairs are equal. Scale values and the variance are estimated by maximum likelihood so as to predict best the set of the observer’s responses over the course of an experiment. The estimated scale has the property that equal scale differences correspond to equal perceived differences. The scale is unique, however, only up to a linear transformation, i.e., adding and/or multiplying all scale values by a constant does not affect the predictions. Thus, we fix the scales to be zero at the lowest stimulus values tested and to have at each stimulus level. Parameterized in this fashion, the estimated scale values of an MLDS experiment are on the same scale as the measure .

### Table 1. CIE LChuv Values for the 10 Levels of Stimuli in Experiment 1

<table>
<thead>
<tr>
<th>Level</th>
<th>Lightness (L in LCh space)</th>
<th>Red</th>
<th>Yellow</th>
<th>Blue</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>30.5</td>
<td>21</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>31</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>41.5</td>
<td>41</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>47</td>
<td>51</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>52.5</td>
<td>61</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>58</td>
<td>71</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>63.5</td>
<td>81</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>69</td>
<td>91</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td>74.5</td>
<td>101</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

In the lightness condition, chroma and hue were kept constant. In the chroma conditions, lightness was kept constant.
from signal detection theory \[41,42\], i.e., in units of the standard deviation of the internal noise (see \[43\]). We used this parameterization in the graphs throughout the article.

4. EXPERIMENT 1: RESULTS

Figure 1 shows the average perceptual scale values across observers within conditions. The red chroma condition displays a steep slope at low chroma values indicating a “crispening effect” \[44–47\], i.e., enhanced sensitivity to chroma differences at low chroma values. Subsequent to this high slope region, the scale values increase approximately linearly. The scales for the other dimensions are similar in their trends and more nearly linear though the yellow and blue scales display a slight crispening at low chroma and the slope of the achromatic scale decreases slightly at high values. The standard error indicated for each point reflects differences between individuals and grows with increase along the physical dimension. This results because observer differences are manifested primarily as sensitivity differences, i.e., the heights of the curves vary across individuals rather than their shapes. As the curves are anchored at \(d' = 0\) by construction, larger responses will display a greater interindividual variability. A different pattern of interindividual variation is evident when the scales are parameterized so that the minimum and maximum values are at 0 and 1, respectively \[28\].

A. Interpolation

The purpose of conducting MLDS was to choose a stimulus set spanning perceptually equal ranges across dimensions with perceptually equidistant levels along each dimension for use in the subsequent MLCM experiment. Four levels along each dimension were selected with a range of \(d' = 12\), the maximum value within the gamut along all dimensions and then interpolating values in LCH space (on the x-axis) from the specified \(d'\) (y-axis) values. These points are indicated on the graphs in Fig. 1 as red triangles and their LCh\(_{uv}\) values are shown in Table 2. The units for the chroma conditions for each hue are specified in \(C\) and the achromatic lightness condition in \(L\).

5. EXPERIMENT 2: METHODS

As outlined in the general introduction, MLCM is a psychological procedure that models how two physical dimensions contribute to a perceptual judgment \[21,22\]. In this experiment, the two investigated dimensions were lightness and chroma. The levels of these dimensions were selected to be perceptually equidistant using the MLDS procedure outlined in experiment 1.

A. Observers

Thirty observers participated in the experiment (female = 21; mean age = 20.8, SD = 3.05). All were tested for color vision deficiencies using the Ishihara plates \[33\] and the Lanthony test \[34\]. Observers were paid £8 for their participation.

B. Apparatus

The same apparatus was used as in experiment 1. A new program was written in MATLAB R2012 using Psychophysics Toolbox extensions to run the experiment \[35–37\].

C. Stimuli

Four levels of lightness and four levels of chroma that were identified in experiment 1 as giving four equally spaced points perceptually were used for each hue. These formed a \(4 \times 4\) matrix of stimuli varying along the dimensions of lightness and chroma for each of the hues: red, yellow, blue, and green (see Fig. 1 for an illustration of the green stimulus matrix).

D. Design and Procedure

On each trial, a pair of stimuli chosen at random from the \(4 \times 4\) grid (Fig. 2) were presented and observers were asked to judge which one was lighter (condition 1) or which was more chromatic (condition 2). Given the 16 pairs of stimuli per hue condition, there were \(16 \times 16 = 256\) pairs to test (self-comparisons were included but not analyzed as they do not contribute to the estimated scale values). Each pair was presented twice with left-right positions reversed, resulting in

<table>
<thead>
<tr>
<th>Dimension</th>
<th>(d') Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Chroma Red</td>
<td>12.96</td>
</tr>
<tr>
<td>Chroma Yellow</td>
<td>5.00</td>
</tr>
<tr>
<td>Chroma Green</td>
<td>11.26</td>
</tr>
<tr>
<td>Chroma Blue</td>
<td>7.65</td>
</tr>
<tr>
<td>Lightness</td>
<td>42.75</td>
</tr>
</tbody>
</table>
Interval responses to lightness that is a function of the stimulus when the observer judges chroma. The noise term, $\epsilon$, is a box, which initiated the next trial. Observers were asked the observer made a response (left or right) with the button 200 ms, followed by a pair of stimuli presented on screen until it is assumed that when judging the lightness of stimuli.

**6. EXPERIMENT 2: MODEL**

It is assumed that when judging the lightness of stimuli $i$, $j$ and $k$, $l$, the observer forms the noise-contaminated decision variable:

$$\Delta = \Psi^L_i(q^L_{ki}, q^L_{lj}) - \Psi^L_j(q^L_{ki}, q^L_{lj}) + \epsilon. \quad (1)$$

According to the model, if the decision variable ($\Delta$) is negative, then the observer chooses the left stimulus, whereas if it is positive, she chooses the right stimulus. The $\Psi^L$ terms are interval responses to lightness that is a function of the stimulus lightness and chroma. Similar terms ($\Psi^C$) appear for chroma when the observer judges chroma. The noise term, $\epsilon$, is included in order to account for the fact that observers will not necessarily make the same choice on repeated trials when stimulus differences are small. Given a specific form for the combination of responses from the stimulus attributes, the MLCM model estimates the scale values ($\Psi^L_i, \Psi^C_j$) by maximum likelihood so that the estimated scale values best predict the observers’ choices over the experiment.

We considered three nested models to fit the data, with the first the most constrained and the last the least constrained. First, the independence model assumes that the observers’ judgments depend on only one of the component dimensions. For example, there is no contamination of chroma (lightness) when making lightness (chroma) judgments; only the difference in lightness (chroma) between the two stimuli influences the observers’ choices. Second, the additive model assumes that the observer’s response to a stimulus is a simple sum of the component psychological responses to the physical dimensions.

Third, the saturated model has the highest number of free parameters and allows for an interaction between the dimensions. The model is termed saturated because it includes the maximum number of parameters to estimate given the number of stimulus pairs presented.

Initially, we fit the additive model to the data using the MLCM package [48]. This model assumes that the physical lightness level ($\phi^L_i$) and chroma level ($\phi^C_j$) of the stimulus separately and additively contribute to perceived lightness ($\Psi^L$) and perceived chroma ($\Psi^C$). Here, upper case $L$ and $C$ are used to refer to perceived dimensions, whereas lower case $l$ and $c$ are used to refer to physical dimensions. For the additive model, the decision variable is estimated by

$$\Delta = [\Psi^L_{ki} - \Psi^L_{lj}] + [\Psi^L_{li} - \Psi^L_{kj}] + \epsilon, \quad (2)$$

where $\Psi^L_{kl}$ is an additive contribution of physical lightness to perceived lightness that is constant along a row ($i$) in Fig. 2 and $\Psi^L_{kl}$ is the contribution of physical chroma and is constant along a column. The additive model for lightness judgments estimates the difference between the perceived lightness of the two stimuli plus an additive contribution of perceived chroma. The independence model is obtained by suppressing the responses to one of the dimensions in the equation above. The saturated model is obtained by including an interaction term that depends on the levels, $i$ and $j$ of each stimulus.

**7. EXPERIMENT 2: RESULTS**

The additive model was fit to judgments of lightness and chroma for the four conditions of red, yellow, green, and blue to estimate how much each dimension contributed to the decision using the MLCM package [48]. The average contributions of chroma and lightness to the judgments for each stimulus condition are shown in Fig. 3. The column labels indicate the hue of the chromatic component of the stimuli. Circles indicate the lightness contribution and triangles the corresponding chromatic contribution to the judgments. The abscissa values indicate the four stimulus levels for each component as indices varying from 1–4. This corresponds to the four levels of lightness and chroma used in the experiment. The additive model assumes no interaction between the
levels; therefore, the data can be plotted along just four levels (rather than $4 \times 4$ levels).

In Fig. 3, the top left plot shows the additive model fit to judgments of the redness of stimuli. Interestingly, lightness negatively contributes to judgments of the redness of stimuli. This means that a higher lightness component in a stimulus will tend to diminish its chromatic appearance leading the observer to be less likely to choose that stimulus as more chromatic than a stimulus of equal chroma but lower lightness. Chromaticity of the stimulus positively and linearly contributes to judgments of redness. A similar behavior was observed for both the blue and green stimuli, but notably not for the yellow. For the yellow stimuli, increased lightness slightly increases the chromatic appearance of the stimuli.

The second row right plot shows judgments of lightness for yellow stimuli. Here, there appears to be little contribution of the chromaticity of the stimuli to judgments, because the chroma component curve is close to zero at all stimulus intensity levels. The lightness component, however, positively contributes. The same relations are observed with the blue and green stimuli but not for the red. For the red, the data show a small contribution of chroma to lightness for the most chromatic stimuli. Recall that we limited the gamut of the stimulus sets so that the range of stimuli would be perceptually equal. This seems to be born out in that the range of responses for the judged component are approximately equal across all conditions. The contributions of the levels of the dimensions to judgments also appear to be linear in the additive model graphs. The linearity of the responses likely reflects the preselection of equally spaced stimuli with the MLDS method.

The additive model corresponds to just one possibility to describe the data. Another possibility is that the judgments depend on the contributions of only one of the dimensions, as appears to be approximately the case for the chromatic contributions to lightness for three of the hues. This model has been referred to as the “independence” model and is nested within the additive model [21,22]. There is also a more complex model in which the additive combination of the responses does not suffice to describe the data and an interaction term is included to describe the deviations from additivity of the component responses. This model is referred to as the “saturated” model because it includes the maximum number of parameters to describe all of the stimulus conditions. The additive and independence model are nested within the saturated model [22]. We could fit and evaluate the three models with a likelihood ratio test for each individual but this would entail performing multiple tests (30 observers with two tests per observer would give 60 tests), which would be efficient neither computationally nor statistically.

In the software used to fit the models [22], the maximum likelihood procedure is implemented as a generalized linear model (GLM) with a binomial family [49]. This suggests the more efficient alternative of estimating the values using a generalized linear mixed-effects model (GLMM) in which average (population) responses are estimated as fixed effects and observer sources of variability are estimated as random effects [22,50,51]. See Appendix A for an explanation of how the GLMM was fit. The lines drawn through the data

---

**Fig. 3.** These graphs show the normalized parameter estimates for the additive model, averaged across all observers and judgments. Rows show the different hue conditions, i.e., red, yellow, green, and blue. Columns show type of judgment, i.e., lightness judgment (“which is lighter?”) or chroma judgment (“which is redder/yellower/bluer/greener?”). Parameter estimates of the contribution of chroma to judgments are indicated by triangles and contributions of lightness to judgments are indicated by circles. The lines show GLMM fit to the data. Error bars show ±1 SEM.
points in Fig. 3 are those based on the predicted slopes from the additive model of the GLMM analysis. The lines predict the data well in that they are all within 1 SE of the averages but there is a marked tendency in several cases to underestimate the magnitude of the slopes. This results from the influence of the random effects in the model that force a bias-variance tradeoff to choose between estimates that are close to the data but also would be predictive of future data. That the line describes the average data well for the red chroma condition also supports that the significant interaction observed in this condition is not very large.

8. DISCUSSION
In this study, we used MLCM and MLDS to estimate perceptual scales of lightness and chroma for hues red, yellow, green, and blue.

A. Maximum Likelihood Difference Scaling
In experiment 1, we employed MLDS to estimate difference scales for lightness and chroma. We then used these scales to interpolate four perceptually equally spaced points spanning an equal range along all of the dimensions tested, in order to create the stimuli for the following experiment. The fact that the estimated scales vary nonlinearly with separation in CIE LCh_uv simply reflects that this space only approximates uniformity, and that for an average observer [31]. Prior work has questioned the perceptual spacing of uniform CIE color spaces (e.g., [31,32]), in particular, that CIE LCh_uv perceptual differences are only valid over small differences, and are not reliable over larger distances. The differences in CIE LCh_uv units between our stimulus intensity levels are all quite large with respect to discrimination experiments. Deviation from linearity in our data is particularly evident at low chroma values for the red hue: observers in our experiment were more sensitive to differences in chroma here than CIE LCh_uv space would predict. This may be an example of a crispening effect [46] in which observers show heightened sensitivity to small stimulus differences near the background level. The effect may be linked to changes in stimulus appearance at those intensity levels. At the lowest chroma level, the stimulus appears gray, whereas by the second and third level the color is better described as “red” or “pink.” This categorical jump between an achromatic stimulus and a chromatic stimulus may be related to the high difference sensitivity in this area, although differences in sensitivity could also be the source of the categorical jump rather than a result of it.

B. Maximum Likelihood Conjoint Measurement
In experiment 2, we used the technique of MLCM to analyze paired-comparison judgments of lightness and chroma. In separate experiments, we asked observers to judge either the lightness or the chroma of the stimuli, to determine the “contamination” of one dimension when making judgments about the other. The results indicate that the lightness and chroma contributions vary linearly as a function of the stimulus index. This reflects, and is an advantage of, our having prescaled the stimuli using the MLDS procedure in order that the stimuli be perceptually equally spaced. In addition, the linearity of response facilitated the application of a more sophisticated statistical analysis of the contributions of the different dimensions and individual sources of variability.

Observers tended to perceive physically lighter stimuli as less chromatic in the red, blue, and green conditions, demonstrating a veiling effect of lightness [52], but not in the yellow. One potential explanation for this effect is that yellowness and lightness are both defined as an additive combination of the long and medium wavelength sensitive cones, and this may cause the two signals to be more easily confounded or alternatively, one masks the other. An alternative explanation is that it is due to the wording of the task. Observers were asked to decide which of two stimuli was “redder/yellower/bluer/greener” (depending on the hue condition) when making a chromatic judgment. This may have added a linguistic effect to the decision, such that the observer’s judgment was influenced by the typical lightness of the focal color of each term. Focal red, blue, and green are medium lightness whereas focal yellow is a lighter color [53]. This may explain why observers judged lighter stimuli as “yellower” for the yellow hue condition, but darker colors as “redder,” “bluer,” and “greener.” One way to investigate this would be to use an alternative instruction that does not refer to color terms, such as “pick the more colorful one.” Asking observers outright to choose the more saturated stimulus is problematic as those inexperienced with color theory find the concept unintuitive [4]. Pilot testing revealed that observers were unclear on the definition of “saturation” and often confused this for lightness.

Another pattern in the data is that the physical chroma of the stimuli tended not to influence observers’ perception of the lightness for yellow, green, and blue colors. However, in the case of red, increasing the chroma increased the probability that the observer would choose a stimulus as “lighter.” This may be a manifestation of the Helmholtz–Kohlrausch effect [54] in which chroma contributes to perceived brightness. The effect is small here but it is possible that larger effects would be observed with stimuli of higher chroma.

The three nested models applicable to MLCM were evaluated within the framework of generalized linear mixed-effects models. It was found that an additive model best described the data in all conditions except for the condition in which observers judged the chroma of red stimuli. For that condition, the additive model of the physical lightness and chroma contributions to the chroma judgments did not suffice to describe the data and a model including an interaction term described the pattern of observers’ judgments better. The contribution of this term was small, however, and more data over a greater stimulus range should be collected to evaluate if it is playing an important role in perception.

C. Integral and Separable Dimensions
Do the MLCM results address the issue of whether lightness and chroma are separable or integral dimensions? This is a more subtle question. The MLCM method shows that we can isolate two response components associated with manipulation of two (psycho-)physically specified dimensions or scales, that are called lightness and chroma in the CIE LCh_uv space. We introduce a psychological dimension to the task by asking the
observer to order the stimuli on the strength of their perceived lightness or chroma. When the observer makes a lightness judgment, the response component that depends on the scale lightness increases monotonically with the scale value, suggesting that we are tapping a lightness response, and conversely for a chroma judgment. In some cases, the second component follows monotonically the strength of the other scaled component, as estimated in the additive model. Therefore, we are tempted to attribute these components to the psychological lightness and chroma response contributions to the judgments of the observer. In the case of the lightness judgments, three out of four of the cases studied show the lightness component to be independent of the chroma component. It seems reasonable to conclude that the lightness scale is separable from the chroma scale under these circumstances. In one of the cases for the chroma judgments, simple additivity of the response components does not suffice to adequately explain the judgments. It is reasonable to conclude that under this condition the two responses are not separable and are, therefore, integral. But what of the cases in which the response components are additive and how should we interpret the asymmetry in the decision rule for the combination of components depending on which judgment was made?

On the one hand, the fact that observers cannot judge chroma without the contamination of the secondary scale could be taken as evidence that the two scales are psychologically confounded; this has been argued to be a signature of integral dimensions [3,5,7]. On the other hand, the fact that the judgments can be shown to depend on two responses that add in a way such that the level of one does not influence the level of the other could be taken as a signature of separability. If the former is assumed then the asymmetry in performance is taken as evidence for integrality of the dimensions and if the latter, separability. In short, our conclusions about these results ultimately depend on our definitions of the concepts of separable and integral dimensions.

9. CONCLUSION

This study has demonstrated the use of MLCM as a technique for investigating how the psychophysical dimensions of color interact to influence observers’ judgments about color. The utility of prescaling the stimulus dimensions to obtain stimuli of equal perceptual spacing and range was demonstrated in that it yielded linear response functions that were simpler to analyze. The results reveal that lightness and chroma components combine additively or nearly so in chroma judgments, but that lightness judgments are largely independent of chroma level over the stimulus range that we studied. The exception to the above results involved the conditions in which the hue component was red, in which the additive model was rejected for chroma judgments. Interestingly, increased lightness was found to decrease the appearance of chromatic strength, evidencing a veiling effect, for all of the hues tested except yellow.

This study paves the way for further investigation into the dimensions of color using MLCM, for example, to further investigate evidence of individual differences in color perception (e.g., [55,56]), to investigate how infants’ preferences for color dimensions develop [16], or to study the interaction of color dimensions with other dimensions, such as texture [7].

APPENDIX A

To fit the GLMM, we take advantage of the fact that because of our choice that stimuli be equally spaced perceptually, the estimated scales are approximately linear as a function of the stimulus indices, and additionally, they pass through the origin by construction. Thus, to simplify the GLMM specification, we will treat each response scale as a linear function of the indices. Then, we only need to estimate the slopes for each component scale. In this case, the fixed-effects component of the decision variable for the additive model can be notated as

$$\Delta_{ijkl} = (\beta_L L_i + \beta_C C_j) - (\beta_L L_k + \beta_C C_l) + \epsilon, \quad \text{(A1)}$$

where the $\beta$s are the slopes (or gains) and $L$ and $C$ the indices of the luminance and chroma components, respectively, and $\epsilon$ is the Gaussian distributed judgment noise of the observer. This can be rearranged and simplified as

$$\Delta_{ijkl} = \beta_L (L_i - L_k) + \beta_C (C_j - C_l) + \epsilon = \beta_L \Delta L_{ik} + \beta_C \Delta C_{jl} + \epsilon. \quad \text{(A2)}$$

Thus, the decision variable is a function of only the differences between stimulus indices. Any offsets or intercepts in this formulation would be cancelled by the differencing of levels. In the independence model, we constrain one of the slopes to be zero and for the saturated model, we introduce a coefficient $\beta_{CL}$ applied to a product of the indices to obtain an additional interaction term. The advantage of the GLMM formulation, however, is that we can also estimate and test variance components associated with individual differences in the slopes of the components for each condition. For example, a full GLMM including both fixed and random effects can be notated as

$$g(E[Y_i]) = (\beta_{L,TX} + b_{L,O} + b_{L,OX} + b_{L,OTX}) \Delta L_{ik} + (\beta_{C,TX} + b_{C,O} + b_{C,OTX} + b_{C,OX}) \Delta C_{jk}. \quad \text{(A3)}$$

In this model, the dependent variable, $Y$, codes the choices of the observer (left/right as 0/1) and their expected value, here, is the probability of choosing the right (the probability of left being 1 minus this value). This expected response is related to the linear predictor through a link function, $g$, that we take to be an inverse cumulative Gaussian, making this a type of probit analysis. The linear predictor on the right is the sum of the luminance and chromatic difference terms, each with a complex sum of components determining its gain. Greek letters are used for the fixed-effect terms. The fixed-effect term indicates an average or population estimate across the dataset. The subscripts $L$ or $C$ indicate whether the term corresponds to a luminance or chroma slope, $T$ (indexing lightness or chroma tasks), and $X$, the hue of the stimulus set. Roman letters are used for the random effects whose variance contributions are estimated. The term with subscripts $L$, $O$ accounts for observer-specific variations in the slopes for the lightness component with a similar term for the chroma component. Successive random effects, $b_{L,OT}$, $b_{L,OX}$, and $b_{L,OTX}$ refer to random observer...
effects of the slope of the luminance component that are specific to the task ($T$), the hue component ($X$), or both the task and the hue component with matching terms for the chroma slope. Equation (5) is often simplified in matrix notation by separating out the fixed and random effects as

$$g(E[Y]) = X\beta + Zb,$$

where $X$ is the model matrix for the fixed-effect terms, $Z$ the model matrix for random effects, and $\beta$ and $b$ are vectors of the fixed- and random-effect coefficients, respectively. The fixed-effect coefficients are estimated directly in the model as are the variances and covariances of the random effects. In this form, it is evident that the random effects enter the model as a regularizer formally equivalent to a prior in a Bayesian model that results in estimators that balance the trade-off between the task variance components 1.1), it was highly significant ($\chi^2(4) = 55.7, p \ll 0.001$) but the test between the additive and the saturated model was not significant ($\chi^2(4) = 8.5, p = 0.075$). Thus, for the blue and green conditions, the additive model describes the data most parsimoniously.

Examining the red and yellow experiment in more detail indicated that the significance of the interaction term could be traced to the chroma judgments for the red hue condition. Though the coefficient for the interaction term was small for this condition (0.034 when the slope of the chromatic component was 1.1), it was highly significant ($z = 3.36, p = 0.0008$). This compares with the small and not significant interaction coefficients in the other three conditions for this experiment (chroma task, yellow hue: 0.004, $p = 0.66$; lightness task, chroma red: $-0.010, p = 0.40$; lightness task, chroma yellow: 0.004, $p = 0.76$). Thus, the additive model suffices to describe the data in these three conditions.

**Funding.** Agence Nationale de la Recherche (ANR) LabEx CORTEX (ANR-11-LABX-0042); European Research Council (ERC) (283605).

**Acknowledgment.** A European Research Council grant was provided to A.F. under the project “CATEGORIES.”

**REFERENCES**

20. J. Falmagne, Elements of Psychophysical Theory (Oxford University, 2002).
33. S. Ishihara, Ishihara’s Test Chart for Colour Deficiency (Kanehara Trading Inc., 2010).
34. P. Lanthony, Album Titian, 2nd ed. (Springer, 1987).
55. H. von Helmholtz, Treatise on Physiological Optics, in The Sensations of Vision (Optical Society of America, 1924), Vol. II.