

Color appearance of spatial patterns compared by direct estimation and conjoint measurement

Frédéric Devinck, Kenneth Knoblauch

▶ To cite this version:

Frédéric Devinck, Kenneth Knoblauch. Color appearance of spatial patterns compared by direct estimation and conjoint measurement. Journal of the Optical Society of America. A Optics, Image Science, and Vision, 2023, 40 (3), pp.A99. 10.1364/JOSAA.475040 . hal-04048552

HAL Id: hal-04048552 https://hal.science/hal-04048552

Submitted on 10 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Color appearance of spatial patterns compared by direct estimation and conjoint measurement

3 FRÉDÉRIC DEVINCK,^{1,*} AND KENNETH KNOBLAUCH^{2,3,4}

4 ¹Université Rennes 2, LP3C, EA 1285, 35000 Rennes, France

5 ²INSERM U1208 Stem-Cell and Brain Research Institute, Department of Integrative Neurosciences,

6 69675 Bron Cedex, France

7 ³Université Lyon 1, 69003 Lyon, France

8 ⁴National Centre for Optics, Vision and Eye Care, Faculty of Health and Social Sciences, University of

9 South-Eastern Norway, Kongsberg, Norway

10 *Frederic.devinck@univ-rennes2.fr

11 Abstract: Perceptual scales of color saturation as obtained by direct estimation (DE) and 12 maximum likelihood conjoint measurement (MLCM) were compared for red checkerboard 13 patterns and uniform red squares. For the direct estimation task, observers were asked to rate 14 the saturation level as a percentage, indicating the chromatic sensation for each pattern and 15 contrast. For the MLCM procedure, observers judged on each trial which of two stimuli that 16 varied in chromatic contrast and/or spatial pattern evoked the most salient color. In separate 17 experiments, patterns varying only in luminance contrast were also tested. The MLCM data 18 confirmed previous results reported with DE indicating that the slope of the checkerboard scale 19 with cone contrast levels is steeper than that for the uniform square. Similar results were 20 obtained with patterns modulated only in luminance. DE methods were relatively more variable 21 within observer, reflecting observer uncertainty, while MLCM scales showed greater relative 22 variability across observers, perhaps reflecting individual differences in the appearance of the stimuli. MLCM provides a reliable scaling method, based only on ordinal judgments between 23 24 pairs of stimuli and that provides less opportunity for subject-specific biases and strategies to 25 intervene in perceptual judgements.

26 © 2023 Optica Publishing Group

27

28 1. Introduction

29

Direct estimation (DE) is commonly used in color vision studies to measure hue and saturation
perception [1,2]. While the method has been carefully validated [3], it has been criticized as
being nonintuitive for naïve observers, subject to cultural or linguistic biases and dependent on
an observer's capacity to translate percepts into numbers [4]. A recent study sought to address
these criticisms by using a similarity rather than a rating procedure [4].

35 Over the last 20 years, scaling methods have been introduced that are based on simple ordinal 36 comparisons of stimuli, and the perceptual scales are estimated to maximize the likelihood of 37 the observer's judgments over the experimental trials within the framework of a signal detection 38 model of the decision process [5-7]. Ordinal comparisons would be expected to be easier to 39 make than subjective ratings, and these methods offer a more rigorous procedure for obtaining 40 scale values. Nevertheless, there are no guarantees that both methods would generate the same results. Here, we compare DE and Maximum Likelihood Conjoint Measurement (MLCM) in 41 a replication of experiments that investigated the influence of spatial complexity and cone 42 43 contrast on color appearance. MLCM is based on paired-comparisons and is used to estimate 44 perceptual scales associated with the integration of information along multiple dimensions [8-45 10].

46 Recently, Shapley, Nunez & Gordon (2019) argued that two separate systems contribute to 47 color appearance [11]. One system is sensitive to low chromatic contrasts and responds best to 48 low spatial frequencies while the other is most active at high chromatic contrasts and has a 49 band-pass frequency characteristic. This hypothesis is based on the neural responses of two 50 color-sensitive cell classes described in primate cortical area V1: single-opponent and double-51 opponent cells [12-18]. Single-opponent cells respond best to the low spatial frequency content of chromatic patterns (< 0.5 c/deg) and display a fall-off in sensitivity to high frequency 52 53 chromatic patterns. In contrast, double-opponent cells display a band-pass sensitivity, with 54 greater selectivity for patterns at higher spatial frequencies (2 c/deg) with response fall-off to 55 patterns at both low and high spatial frequencies [12,17,19].

56 Nunez, Shapley & Gordon (2018) used a direct estimation (DE) [20], saturation scaling method 57 to estimate the appearances of equiluminant uniform and high frequency checkerboard stimuli 58 under the assumption that these stimuli would preferentially activate single- or double-59 opponent cell classes, respectively. The perceived saturation of both stimuli increased as a 60 function of cone contrast, but the checkerboard pattern increased more steeply than the uniform 61 square field for the equivalent size and cone contrasts. Thus, the strength of the estimated color 62 appearance was stronger for stimuli designed to preferentially activate double-opponent rather 63 than single-opponent cell classes. These results were supported by VEP measurements which 64 provided evidence for two different mechanisms underlying the responses to the two patterns 65 [20,21].

Here, we replicate the DE procedure of Nunez et al. (2018) for reddish equiluminant stimuliand compare the results with those obtained using MLCM. We also compared the methodsusing patterns varying only in luminance.

69

70 2. General Methods

71

72 2.1 Observers

73 Seven observers participated in the experiments, six of whom were naïve to the purpose of the 74 study. Ages ranged from 26 to 46 years. All observers had normal color vision as tested by a 75 Farnsworth Panel D15 and had normal or corrected-to-normal visual acuity. We tested five 76 participants for each task. The same five observers participated in the DE and the MLCM tasks for patterns displayed along a given axis. For the MCLM and the direct estimation tasks with 77 patterns displayed along the L+M axis, three observers from the previous tasks remained, and 78 two additional participants were recruited. All studies were conducted in agreement with the 79 80 Declaration of Helsinki for the protection of human subjects.

81

82 2.2 Apparatus

83 The experiments were run in a dark room and were programmed using MATLAB 7.9 (MathWorks, http://mathworks.com) and the CRS Toolbox. The stimuli were presented on a 84 85 NEC MultiSync FP2141sb color CRT monitor driven by a Cambridge Research ViSaGe graphic board with a color resolution of 14 bits per gun (Cambridge Research Systems, 86 Rochester, United Kingdom). The monitor had a diagonal screen size of 22 inches, resolution 87 88 of 1024×768 pixels, and a refresh rate of 120 Hz. The screen was calibrated using a SpectroCal 89 spectroradiometer with the calibration routines of Cambridge Research Systems. A Cedrus 90 RB540 response pad was used to collect observer responses (Cedrus Corporation; San Pedro, 91 CA, USA). Observer position was stabilized by a chinrest so that the screen was viewed 92 binocularly at a distance of 80 cm.

94 2.3 Stimuli

95 Observers viewed a red-gray checkerboard pattern, or a red square displayed on a gray
96 background. In a separate session, a white-gray checkerboard pattern and a white square were
97 also used. The same spatial patterns were used for the DE method and the MLCM procedure
98 though more contrast levels were tested using MLCM. Stimuli were designed to be similar to
99 those used by Nunez et al. [20,21,29].

100 The stimulus size measured 10×10 deg. The checkerboard had 32×32 checks, each of 101 dimensions 0.3141×0.3141 deg. The choice of the checkerboard and square sizes were chosen 102 according to the values used by Nunez et al. (2018). As they did, all stimuli were vignetted by 103 a stationary Gaussian window ($\sigma = 2$ deg). Thus, the checkerboard and the square faded 104 gradually to the background gray color moving away from the stimulus center. The gray 105 background had fixed chromatic coordinates (x,y = 0.30, 0.35; Y = 29.89 cd/m²), whereas 6 106 and 10 different chromatic values were used for the DE and for the MLCM procedures, 107 respectively. The CIE coordinates are provided in Table I (red stimuli) and II (white stimuli). 108 Note that for the red stimuli, contrast values differed from those used by Nunez et al. (2018) 109 but these values overlapped extensively. For the luminance conditions, values ranged from 36.16 cd/m^2 to 111.60 cd/m^2 . 110

111

112

113 114

Table 1. CIE color coordinates, cone contrast (the difference between the cone excitation of the stimulus and the cone excitation of background divided by the cone excitation of the background) and the contrast measured for each red color presented.

115

MLCM	CII	E coordinat	tes	(Cone values		Contrast
	x	У	Y	L	М	S	$\sqrt{\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta M}{M}\right)^2 + \left(\frac{\Delta S}{S}\right)^2}$
							$\sqrt{3}$
	0.3147	0.3440	30.36	19.9356	10.5941	0.2281	0.0189
	0.3290	0.3391	30.72	20.4461	10.4459	0.2279	0.0356
	0.3420	0.3325	31	20.9114	10.2650	0.2302	0.0531
	0.3421	0.3228	31.04	21.0366	10.1923	0.2445	0.0739
	0.3720	0.3229	32.14	22.3234	9.9991	0.2308	0.0971
	0.3848	0.3178	32.82	23.1016	9.9059	0.2336	0.1213
	0.3992	0.3141	33.86	24.1695	9.8825	0.2354	0.1521
	0.4121	0.3103	34.75	25.1287	9.8176	0.2370	0.1809
	0.4240	0.3087	36.1	26.3870	9.9134	0.2385	0.2164
	0.4369	0.3065	37.9	28.0368	10.0701	0.2424	0.2645
Direct Estimation	x	У	Y	L	М	S	$\sqrt{\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta M}{M}\right)^2 + \left(\frac{\Delta S}{S}\right)^2}$
							√3
	0.3147	0.3440	30.36	19.9356	10.5941	0.2281	0.0189
	0.34	0.33	30.62	20.6199	10.1729	0.2260	0.0473
	0.37	0.32	31.6	21.8173	9.9622	0.2286	0.0838
	0.39	0.32	33.09	23.4097	9.8684	0.2331	0.1302
	0.41	0.31	34.94	25.3081	9.8277	0.2359	0.1854
	0.4369	0.3065	37.9	28.0368	10.0701	0.2424	0.2645

116

117

118

121 122 123 124

Table 2. CIE color coordinates, cone contrast (the difference between the cone excitation of the stimulus and the cone excitation of background divided by the cone excitation of the background) and the contrast measured for each luminance level presented.

MLCM	CIE coordinates			(Cone values		Contrast
	х	У	Y	L	М	S	$\sqrt{\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta M}{M}\right)^2 + \left(\frac{\Delta S}{S}\right)^2}$
							$\sqrt{3}$
	0.3023	0.3501	36.16	23.4569	12.9026	0.2717	0.2063
	0.3024	0.3502	42.53	27.5904	15.1741	0.3192	0.4184
	0.3025	0.3470	49.27	32.0006	17.5469	0.3766	0.6529
	0.3026	0.3457	56.68	36.8331	20.1689	0.4364	0.9060
	0.3028	0.3444	65.01	42.2725	23.1099	0.5040	1.1911
	0.3027	0.3430	72.74	47.3185	25.8422	0.5687	1.4587
	0.3031	0.3424	82.16	53.4753	29.1614	0.6438	1.7790
	0.3032	0.3414	91.91	59.8476	32.5992	0.7242	2.1146
	0.3034	0.3407	100.2	65.2729	35.5148	0.7923	2.3995
	0.3035	0.3400	111.6	72.7236	39.5340	0.8858	2.7911
Direct	Х	у	Y	L	М	S	
Estimation		-					$\sqrt{\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta M}{M}\right)^2 + \left(\frac{\Delta S}{S}\right)^2}$
							$\sqrt{3}$
	0.3023	0.3501	36.16	23.4569	12.9026	0.2717	0.2063
	0.3024	0.3470	47.89	31.1018	17.0580	0.3662	0.6068
	0.3028	0.3546	61.52	40.0003	21.8716	0.4764	1.0727
	0.3031	0.3430	75.74	49.2862	26.8916	0.5915	1.5589
	0.3032	0.3414	92.1	59.9714	32.6666	0.7257	2.1211
	0.3035	0.3400	111.6	72.7236	39.5340	0.8858	2.7911

125

126 Stimuli were specified in the DKL color space [22-24]. This is a three-dimensional cone-127 opponent space based on the Smith and Pokorny (1975) cone fundamentals [25]. The first axis 128 represents a luminance mechanism based on the sum of L and M cone excitations (L+M). The 129 second axis represents the difference between the L and M cone excitations (L-M). Finally, the 130 third axis defines lights that selectively modulate the S cones as L-M and L+M are constant 131 along this axis. Hue direction is specified by the azimuth. Luminance is scaled between -1 and 1, where the value of 1 corresponds to the maximum of the monitor luminance and the value of 132 133 -1 to the minimum. Chromatic strength ranges from 0 to 1 and refers to the maximum value along the axis that may be displayed on the monitor. Note that a value of 0 corresponds to the 134 origin in DKL color space; increasing the distance from 0 in the equiluminant plane leads to 135 136 increases in the perceived saturation of the stimulus.

137 The stimuli were defined with the gray background and red colors at azimuth of 0 deg. The 138 saturation levels of the reddish colored stimulus varied from 0.1 to 1 along the L-M axis in 139 DKL color space while the gray background was fixed at 0. In separate sessions, the red-gray 140 checkerboard and the red square were replaced by a white-gray checkerboard and a white 141 square. Here, the azimuth is fixed at 0 deg and the luminance elevations of the white color 142 ranged from 0.1 to 1 along the L+M axis in the DKL color space.

143 The contrasts of the colored stimuli are presented in Table I and II. In both tables, the L, M 144 and S cone contrasts were defined as the difference between the maximum cone excitation of 145 the stimuli and the background and divided by the cone excitation of the background. Then,

Simulus contrast was calculated as
$$\sqrt{\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta M}{M}\right)^2 + \left(\frac{\Delta S}{S}\right)^2}$$

146 the stimulus contrast was calculated as $\frac{\sqrt{L}}{\sqrt{3}}$

147

149 2.4 Procedure

Each observer completed the two experimental sessions on separate days. Observers were adapted to the grey background for 3 min at the beginning of each session. There was a practice session of 5 trials, followed by the experimental session, if the observer felt comfortable with the task; otherwise, additional practice trials were run. No feedback was provided during any part of the experiment.

155 *Direct estimation task*

156 Each stimulus was presented four times for each condition, tested in a random order. Each 157 session consisted of 48 trials (2 patterns \times 6 saturation levels \times 4 repetitions). On each trial, a black fixation cross was presented in the center of the screen for 500 ms. At its offset, the 158 stimulus appeared for 500 ms, followed by a blank screen. The participants were asked to rate 159 the saturation level as a percentage indicating the chromatic sensation. Thus, the absence of 160 chromatic sensation is indicated by 0% of saturation, and a maximal sensation corresponded to 161 162 100% of saturation. The subsequent trial was activated by pressing the appropriate button. A 163 1 s pause was interleaved between each trial.

164 Maximum Likelihood Conjoint Measurement task

165 In the MLCM experiments, 10 contrast levels and 2 stimulus patterns were used, generating a 166 2×10 grid of 20 stimuli. Figure 1 shows an example of the stimulus grid from which stimulus 167 pairs were selected for patterns displayed along the L-M axis. On each trial, two different 168 stimuli were chosen randomly from the 2×10 grid (see Figure 1); there are a total of $(20 \times 19)/2$ 169 = 190 non-identical pairs. A session consisted of the random presentation of all test pairs. It 170 was repeated two times, yielding 380 trials for each observer.

171 Initially, a black fixation cross was presented in the center of the screen for 500 ms. When it 172 disappeared, two stimuli were presented during 500 ms, followed by a blank screen. Observers 173 were instructed to judge which pattern appeared more reddish or brighter, depending on the 174 stimulus set. Responses were recorded as right or left button presses and initiated a 1 s pause 175 before the initiation of the next trial. Overall, the MLCM sessions, including a rest between 176 repetitions, lasted around 25 minutes while the DE task required about 15-20 minutes.

177



- 178
- 179 180 181

182

Fig. 1. Examples of stimulus set used for the Maximum Likelihood Conjoint Measurement experiment. Each row corresponds to a stimulus pattern, a checkerboard on the top and a square on the bottom. Each column represents different saturation levels displayed along the L-M axis.

183 2.5 Statistical Analyses

184
185 All statistical analyses were performed in the R statistical environment [26]. Data were plotted
186 with functions from the lattice package [27]. Random effects analyses were performed using
187 the *lmer* function in the lme4 package [28].

189 3. Results

190 3.1 Saturation levels displayed along the L-M axis

191 Figure 2 shows each observer's mean saturation estimates for the DE task as a function of cone 192 contrast for the square (white circles) and the checkerboard (black circles) stimuli. The column 193 labels identify the observers; average functions are presented in the rightmost panels. The 194 abscissa cone contrast levels of the uniform square were scaled to take into account that the 195 space-averaged cone contrast levels of the red-gray checkerboard are one-half those of the 196 uniform square of the same cone contrast levels [10]. The observers were ordered with respect 197 to their estimations at the maximum contrast, and then their identifications were assigned so 198 that Observer 1 gave the highest ratings and Observer 5 the lowest. The top row shows the 199 results displayed with a linear abscissa while the bottom row displays the same data on a 200 logarithmic abscissa. For both the chromatic checkerboard and uniform square, estimates 201 increased monotonically with the cone contrast. The estimates for the chromatic checkerboard increased more steeply than those for the uniform square over the same cone contrast levels. 202 203 On the log abscissa, the difference in steepness appears as a lateral shift of a roughly invariant 204 shaped function. Informally, after the measurements were completed, observers reported the 205 checkerboard to appear more saturated than the square at equal cone contrast, when presented 206 side-by-side. These results replicate those found by Nunez et al. [20].

207 The data displayed individual variability in the range of estimated saturations and in the relative 208 separation of the two curves. For example, the saturation estimated for the maximum contrast 209 for Obs. 5 was 40% of that of Obs. 1. In addition, the two curves for Obs. 5 tend to overlap for 210 low contrasts, whereas this does not occur for the other observers. Obs. 5 also tended to show 211 more variable estimates.

212

(a)





Fig. 2. Results obtained with a direct estimation procedure. Mean estimated scale presented as a function of cone contrast for the square (white circle) and the checkerboard (black circle) patterns displayed along the L-M axis. Data are presented with a (a) a linear scale and (b) a logarithmic scale. Error bars show +/- 1 SEM for estimates across the four repetitions.

218

213 214

215

For the MLCM procedure, the data were analyzed as a decision process within the framework
of a signal detection model and fitted by maximum likelihood [5,6]. Perceptual contributions
of each dimension were estimated using the software package MLCM [29] in the open-source
software R [26].

223 Figure 3 shows the average estimated scales for each pattern obtained from fitting the additive 224 model to the data of each of the 5 observers and the average functions. As in the previous 225 figure, the abscissa values indicate the cone contrast values, with linearly spaced abscissa in 226 Fig 3a and a logarithmically spaced abscissa in Fig. 3b. Filled circles represent the estimated 227 checkerboard scale while open circles indicate the uniform square scale. The data follow the 228 same qualitative pattern as in Figure 2, using the DE method. The contribution of the 229 checkerboard pattern increases more steeply with cone contrast than the square pattern. On a 230 logarithmic abscissa, curves of similar shape are simply translated along the abscissa. As the 231 observers are ordered as in Fig. 2, the extent to which the two methods agree with respect to 232 the response range can be evaluated. The response ranges of the estimated scales tend to follow 233 those obtained using the DE method except for Obs. 5, who rather than showing the weakest 234 responses as with DE, shows one of the strongest responses with MLCM. 235



236

Fig. 3. Estimated scale obtained with MLCM procedure. Average estimates for the square (white circle) and the checkerboard (black circle) patterns displayed along the L-M axis as a function of cone contrast for five observers. Data are presented with a (a) a linear scale and (b) a logarithmic scale. Error bars show +/- 1 SEM for estimates across the two sessions.

241

242 3.2 Luminance levels displayed along the L+M axis

Results of the DE task for five observers are displayed in Figure 4 along with the average curves
and are presented as a function of cone contrast for the checkerboard (black circles) and for the
square (white circles). Results for both patterns show a qualitatively similar pattern as that
along the L-M axis with estimates increasing monotonically with contrast, except for Obs. 7.
The estimated contrast of the checkerboard was higher than that of the uniform square and on

the logarithmic abscissa, the curves appear to differ only by a lateral translation. There are individual differences in the range of estimates that the observers used. Obs. 7 generated less consistent estimates as indicated by the larger error bars and the estimates varied nonmonotonically with contrast. Again, the ordering in the figures is with respect to maximum estimates along the scales in order to facilitate comparison with the MLCM measures.



Fig. 4. Estimated scale obtained with direct estimation procedure. Mean estimated scale presented as a function of cone contrast for the square (white circle) and the checkerboard (black circle) patterns displayed along the L+M axis across all observers. Data are presented with a (a) a linear scale and (b) a logarithmic scale. Error bars show +/- 1 SEM for estimates across the four repetitions.

254

259 Scales estimated from the MLCM procedure are displayed in Figure 5, using the same format 260 as the previous figures. Results show the average estimated scales under the additive model 261 for each observer. Filled circles indicate the estimated checkerboard scales and open symbols the square scales. The values on the abscissa indicate the 10 cone contrast values for each 262 263 dimension as displayed along the L+M axis in the DKL color space. The MLCM results agree 264 qualitatively with those from the DE method in the relative forms and spacing of the two curves. 265 The range of responses for DE and MLCM agree across methods except for Obs. 6, whose MLCM responses are relatively smaller compared to those for DE. Also, the response curves 266 267 of Obs. 7 are more orderly for MLCM than for DE.



268 269 270

271

272 273

Fig. 5. Estimated scale obtained with MLCM procedure. Additive model average estimates for the square (white circle) and the checkerboard (black circle) luminance patterns as a function of cone contrast for five observers. Data are presented with a (a) a linear scale and (b) a logarithmic scale. Error bars show +/- 1 SEM for estimates across the two sessions.

274

290

275 3.3 Comparison between Direct Estimation and MLCM methods.

276 Since the DE and MLCM methods yield response curves that are on different scales, we 277 compared variability using the coefficient of variation (CV), defined as the standard deviation 278 divided by the mean and that provides a measure of relative variability. Figure 6 compares the 279 CV across method for each pattern and color axis, using box-and-whisker plots. The medians (black points) are similar across methods and the interquartile ranges (grey boxes) overlap, 280 281 suggesting that there is little difference in variability between the two methods. This is 282 confirmed by fitting the data with a linear model with contrasts chosen to compare the means within each panel of Figure 6 (L+M/Check: t(278) = 0.422, p = 0.673; L+M/Square: t(278) = 0.422, t283 284 0.985, p = 0.325; L-M/Check: t(278) = 0.779; p = 0.437; L-M/Square: t(278) = 1.213, p = 0.985, p 285 0.226). Because all of the explanatory variables are factors, this is technically an analysis of 286 variance, but we only concentrate on the pre-planned contrasts of interest here, which are 287 distributed as t-statistics. None of the contrasts were significant so that corrections for multiple 288 tests would not affect the (non-)significance of the results. These plots, however, collapse the 289 data over subject and contrast level.



Fig. 6: Comparison of DE and MLCM estimated with the coefficient of variation (CV) for the checkerboard (left side) and the square (right side) displayed along the L+M (top row) and L-M axes (bottom row) as a function of cone contrast. Filled symbols represent the medians, and the grey boxes indicate the interquartile ranges (IQR). The whiskers are defined to extend to the most extreme data points within 1.5 times the IQR from the box. Any data points outside the whiskers are plotted as unfilled symbols.

297

298 Figure 7 shows the CVs for each individual plotted as a function of contrast for both methods 299 for both patterns and for both axes in color space. Within observer, the CV appears to be more 300 stable for the MLCM than for the DE method. The MLCM values do not appear to depend 301 systematically on contrast, stimulus or color axis except for a few observers along the L - M 302 axis, who show a jump at the lowest contrast. High variability for small means (e.g., at low 303 contrasts) will inflate the CV. For some observers, the DE values fluctuate unsystematically 304 with contrast. There appears to be greater variability of the CV across observers for MLCM 305 than DE, however. This is supported by a random effects analysis for each method. The within 306 observer variance for DE accounts for 90% of the variance while only 71% for MLCM. The 307 across observer variance for MLCM is 4.7 times greater than that for DE.



Fig. 7. The CVs for each observer plotted as a function of cone contrast for each procedure and pattern. Panels on the left side display the results for the L-M axis, and panels on the right side indicate the results for the L+M axis. Symbols assigned to each observer are indicated in the legends.

317 4. Discussion

310 311 312

313

314

315 316

318 DE methods have been widely used to measure color appearance as they are relatively fast and 319 simple to implement [30]. However, the task has been criticized as not necessarily measuring 320 color appearance alone but also reflecting the observer's perception of numbers, especially for 321 naïve observers. The qualitative similarity in the shapes of the functions obtained for the two methods provides some validation of the DE approach for color scaling. In most cases, 322 323 individual differences in ordinal strength of the estimated appearance correlate between 324 methods, but some glaring exceptions occur. It is possible that the variability of some of the 325 observers for the DE results reflects uncertainties in the task rather than aspects of color 326 appearance per se. Such uncertainties might be reduced by presenting observers at the 327 beginning of a DE session with examples of the highest and lowest contrasts and indicating that 328 these should receive the most extreme ratings. This procedure, however, would force the 329 functions for different patterns to attain the same maximum value, even if the actual 330 appearances are not equal. The MLCM procedure is based on a more intuitive response that 331 requires the observer only make an ordinal judgment, i.e., to choose which of a pair of stimuli 332 appears more reddish or brighter. The appearance scale is then derived from a model of the 333 observer's decision process using a maximum likelihood criterion [6].

Here, we repeated the DE task 4 times at 6 contrast levels which required 48 judgments.
Repeating the MLCM task twice using 10 contrasts required 3.5 times as many trials.
Nevertheless, the MLCM sessions on average only lasted about 5 minutes longer than those
using DE. Thus, given that the DE sessions lasted about 20 minutes, it might be expected to
require at least an hour to acquire as many trials from DE as from the MLCM task.

In general, the two tasks yielded similar results. As previously reported using DE for stimuli modulated along the L-M axis [20], the contribution of the checkerboard pattern increases more steeply with cone contrast than the uniform pattern. We report, here, the same behavior for stimuli modulated along the L+M axis. We found comparable results along both color axes using MLCM, thereby recommending it as an alternative procedure that is more intuitive for naïve observers. 345 When the results are plotted as a function of log contrast, the scales for the two patterns appear 346 invariant in shape except for a lateral translation along the abscissa. This behavior is consistent 347 with a single mechanism that is less sensitive to uniform fields than checkerboard stimuli. 348 Evidence for multiple mechanisms is supported by results from VEP studies [11,20,21,31] that 349 show, for example, different contrast dependencies of the response latency across patterns. The 350 similar behavior along the L+M axis would be expected given the low-frequency fall-off of the 351 luminance contrast sensitivity function and the evidence for multiple spatially tuned channels 352 in the visual system [32]. These results are also supported by recent evidence indicating separate mechanisms for luminance coding of surface and patterns in cortical area V1 [33,34]. 353

There was a trend across most observers for scale strengths to agree across methods, i.e., observers with high ratings at maximum contrast using DE also displayed strong perceptual responses at maximum contrast as estimated by MLCM. There were some notable exceptions, however, in which low response on DE was matched with high response on MLCM (Obs 5 along the L-M axis, Figs 2 and 3) and vice versa (Obs 6 along the L+M axis, Figs 4 and 5). We suspect that such differences reflect individual biases in how these observers assign numbers to perceptual events.

361 There were also differences in within-subject variation between methods. On the one hand, DE 362 methods were relatively more variable within observer, perhaps reflecting observer uncertainty 363 in assigning ratings. On the other hand, MLCM scales showed greater relative variability across 364 observers. Given the lower within subject variability of MLCM, we believe these across 365 observer differences are indicative of actual individual differences in the appearance of the stimuli, i.e., individual differences in how they are perceived. Perhaps, additional practice 366 367 sessions prior to data collection would further minimize variability as well as differences 368 between methods.

In conclusion, modeling of decisions based on simple ordinal judgments of pairs of stimuli
 largely reproduces the results from direct estimation of perceptual attributes of individual
 stimuli. Occasional differences between the two methods may reflect level of observer practice
 or biases in the assignment of numbers to perceptual magnitudes.

- 373
- 374
- Funding, K. Knoblauch was supported by LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon (ANR-11-IDEX- 0007) and (ANR-19-CE37-0000, DUAL_STREAMS) operated by the French National Research Agency.
- 377 Acknowledgments. We thank John S. Werner for critical comments of an earlier draft of the manuscript.
- **378 Disclosures.** The authors declare no conflicts of interest.

379 Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

381 382

383 References

- L.M. Hurvich, and D. Jameson, "Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation, and hue in normal and dichromatic vision," J. Opt. Soc. Am. 45, 602-616 (1955).
- J.S. Werner, and B.R. Wooten, "Opponent chromatic mechanisms: Relation to photopigments and hue naming," J. Opt. Soc. Am. 69, 422–434 (1979).
- J. Gordon, I. Abramov, and H. Chan, "Describing color appearance: hue and saturation scaling," Percept. Psychophys. 56, 27-41 (1994).
- C.N. Matera, K.J. Emery, V.J. Volbrecht, K. Vermuri, P. Kay, and M.A. Webster, "Comparison of two method of hue scaling," J. Opt. Soc. Am. A 37(4), A44-A54 (2020).
- 392 5. Y.X. Ho, M.S. Landy, and L.T. Maloney, "Conjoint measurement of gloss and surface texture," Psychol. Sci. 19, 196–204 (2008).
- 6. K. Knoblauch, and L.T. Maloney, *Modeling Psychophysical Data in R* (Springer, 2012).
- 395 7. D.M. Green, and J.A. Swets, Signal Detection Theory and Psychophysics (Wiley, 1966).

396	8.	L.T. Maloney, and K. Knoblauch, "Measuring and modeling visual appearance," Annu. Rev. Vis. Sci. 6, 519-
397		537 (2020).
398	9.	G. Aguilar, and M. Maertens, "Toward reliable measurements of perceptual scales in multiple contexts," J. Vis.
399		20 , 4:19 (2020).
400	10.	H.C. Sun, D. St-Amand, C.L. Baker, and F.A.A. Kingdom, "Visual perception of texture regularity: Conjoint
401		measurements and a wavelet response-distribution model," PLoS Comp. Biol. 17, 10:e1008802 (2021).
402	11.	R. Shapley, V. Nunez, and J. Gordon., "Cortical double-opponent cells and human color perception," Curr.
403	10	Opin. Behav. Sci. 30 , 1-7 (2019).
404	12.	E.N., Johnson, M.J. Hawken, and R. Shapley, "The spatial transformation of color in the primary visual cortex
405	12	of the macaque monkey," Nature Neurosci. 4, 409-416 (2001).
400	13.	P. Lennie, J. Krauskopi, and G. Sciar, Chromatic mechanisms in strate cortex of macaque, J. Neurosci. 10,
407	14	049-009 (1990).
400	14.	M.S., Livingstone, and D.H. Huber, Anatomy and physiology of a color system in the primate visual cortex, L Nauracci A 200 356 (1084)
410	15	J. Neurosci. 4, 303-330 (1304). D. Shanlay and M. I. Hawkan "Color in the cortay: single and double opponent cells." Vision Pag. 51, 701
411	15.	A snapely, and M.J. Hawken, Color in the cortex, single- and double-opporter certs, vision Res. 51, 701- 717 (2011)
412	16	R M Shapley M I Hawken and F R Johnson "Color in primary visual cortex" in <i>The New Visual</i>
413	10.	Neurosciences, I.S. Werner and L.M. Chaluna, eds. (MIT Press, 2014), np. 569–586
414	17.	D. Schluppeck, and S.A. Engel. "Color opponent neurons in V1: A review and model reconciling results from
415		imaging and single-unit recording," J. Vis. 2 , 480-492 (2002).
416	18.	L.G. Thorell, R.L. De Valois, and D.G. Albrecht, "Spatial mapping of monkey V1 cells with pure color and
417		luminance stimuli," Vision Res. 24, 751-769 (1984).
418	19.	E.N. Johnson, M.J. Hawken, and R. Shapley, "The orientation selectivity of color-responsive neurons in
419		macaque V1," J. Neurosci. 28, 8096-8106 (2008).
420	20.	V. Nunez, R. Shapley, and J. Gordon, "Cortical double-opponent cells in color perception: Perceptual scaling
421		and chromatic visual evoked potentials," i-Perception 9, 1-16 (2018).
422	21.	V. Nunez, J. Gordon, and R.M. Shapley, "A multiplicity of color-responsive cortical mechanisms revealed by
423		the dynamics of cVEPs," Vision Res. 188 , 234-245 (2021).
424	22.	D.I. MacLeod, and R.M. Boynton, "Chromaticity diagram showing cone excitation by stimuli of equal
425		luminance," J. Opt. Soc. Am. 69, 1183-1186 (1979).
426	23.	J. Krauskopf, D.R. Williams, and D.W. Heeley, "Cardinal directions of color space," Vision Res. 22, 1123-
427	24	1131 (1982).
428	24.	A.M. Derrington, J. Krauskopt, and P. Lennie, "Chromatic mechanisms in lateral geniculate nucleus of menory," J. Directal (J. 1941) 257–2041 205 (1004)
429	25	macaque, J. Physiol. (Lond.) 35 7, 241-265 (1984).
450	25.	v.c. smith, and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm,"
431	26	Vision Res. 15(2), 101-171 (1973).
432	20.	_K Cole Team, K. A language and environment for statistical comparing. Vienna, Austria. K Foundation for Statistical Computing (2021) LIPL https://www.R-project.org/
434	27	D. Sarka Lattice: Multiveriate Data Visualization with R (Springer 2008)
435	28	D. Bates M. Maechler B. Bolker and S. Walker "Fitting Linear Mixed-Effects Models Using Ime4." Journal
436	20.	of Statistical Software. 67(1), 1-48 (2015).
437	29.	K. Knoblauch, and L.T. Maloney, and G. Aguilar, MLCM: Maximum Likelihood Conjoint Measurement, R
438		package version 0.4.3 (2019). https://CRAN.R-project.org/package=MLCM
439	30.	I. Abramov, and J. Gordon, "Color appearance: on seeing red, or yellow, or green, or blue," Annu. Rev.
440		Psychol. 45, 451-85 (1994).
441	31.	V. Nunez, R. Shapley, and J. Gordon, "Signals from single-opponent cortical cells in the human cVEP," J.
442		Neurosci. 42(21), 4380-4393 (2022).
443	32.	F.W. Campbell, and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," J. Physiol.
444		(Lond.) 197 , 3:551-66 (1968).
445	33.	Y. Yang, T. Wang, Y. Li, W. Dai, G. Yang, C. Han, Y. Wu, and D. Xing, "Coding strategy for surface
446	~ (luminance switches in the primary visual cortex of the awake monkey," Nature Commu. 13 , 1:286 (2022).
44/	34.	G. Zurawel, I. Ayzenshtat, S. Zweig, R. Shapley, and H. Slovin, "A contrast and surface code explains complex
448 440		responses to black and white stimuli in $V1$," J. Neurosci. 34, 43:14388-402 (2014).
449		