Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Warm/cool judgments as a function of hue, value, and chroma

Frédéric Devinck[1,†,*] and Kenneth Knoblauch[2,3,†]

[1]Université Rennes 2, LP3C, EA 1285, 35000 Rennes, France
[2]INSERM U1208 Stem-Cell and Brain Research Institute, Department of Integrative Neurosciences, Univ Lyon, Université Claude Bernard Lyon 1, 69675 Bron Cedex, France
[3]National Centre for Optics, Vision and Eye Care, Faculty of Health and Social Sciences, University of South-Eastern Norway, Kongsberg, Norway
[†]These authors contributed equally to this work.
*frederic.devinck@univ-rennes2.fr

Received 21 October 2024; revised 17 December 2024; accepted 18 December 2024; posted 18 December 2024; published 22 January 2025

Besides conventional perceptual attributes of hue, brightness, and saturation, colors are universally classified along a warm/cool dimension. Previous estimates of how warm/cool values are distributed across color space have relied on subjective ratings. Here we employed simple ordinal judgments between stimulus pairs using maximum likelihood conjoint measurement (MLCM) to assess the influence of Munsell hue, value, and chroma on warm/cool judgments. We also evaluated an identification task for single stimulus presentations. For the MLCM procedure, observers judged on each trial which of the two stimuli appeared warmer. For the identification task, observers classified individually presented color patches as cool or warm. The judgments were analyzed with probit regression to estimate the underlying perceptual scale values. The results confirm that the contributions of different dimensions to warm/cool variations in color space can be estimated using only ordinal judgments. While for most observers, warm/cool judgments depended on hue, there were individual variations in the extent to which value contributed to warm/cool, and little evidence for an effect of chroma.

https://doi.org/10.1364/JOSAA.545368

1. INTRODUCTION

Color appearance is conventionally defined using the perceptual attributes of hue, brightness, and saturation. Nevertheless, color is often additionally described by various attributes using opposing semantic words such as "soft–hard," "clean–dirty," "light–dark," etc. These attributes are thought to reflect the impact of affective or cognitive aspects of color, frequently in the description of art. A prominent example of one such dimension of color experience is based on warm/cool classifications. Warmness is generally attributed to red-orange hues or long-wavelength lights, while coolness with green-blue hues or short-wavelength lights [1].

This distinction is generally considered as a dimension of color appearance [1,2] and has been found to be culturally independent [3–5]. Studies on emotion have evoked the warm/cool distinction as one of the essential dimensions to define a single-color patch [4,6], and this dichotomy appears prominently when naïve observers rate colors on affective scales [7].

The origin of the warm/cool distinction and of its neural representation remains uncertain, however [8,9]. One hypothesis is that this classification is based on a cognitive process [10]. For example, an early stage of color lexicon evolution is defined by a two-term color naming system, as in the Dani language [11]. The two-term system constitutes a coarse division of color grouping that is marked by the non-basic, non-color-specific terms "warm" and "cool." This was shown by a concordance analysis of the World Color Survey indicating two major fault lines that reflect the psychological distinction between the categories warm and cool [3].

Alternatively, the warm/cool dichotomy might be related to fundamental visual mechanisms involved in sensory coding. In two separate experiments, Katra *et al.* [12] asked observers to rate NCS-specified color chips varying in hue, saturation, and lightness, along with lightness/darkness and warm/cool dimensions, respectively. Warm ratings increased somewhat when the percentage of saturation increased but did not change with the amount of lightness. They also demonstrated a relation between a weighted sum of the activation of opponent-color channels as defined by hue cancellation experiments and the warm/cool ratings. These results support that the warm/cool distinction is also constrained, at least to some extent, by sensory coding [13]. This raises the possibility that the warm/cool dimension may arise as a combination of sensory coding and cognitive constructs.

Most previous measurements of warm/cool values across color space relied on subjective ratings [12,14–16]. An important exception is the recent study of Koenderink *et al.* [17], who used ordinal judgments of stimulus pairs of adjacent hues on a color circle to make relative warm/cool judgments. We have previously shown that ordinal judgments of stimulus pairs using maximum likelihood conjoint measurement (MLCM) [18,19] suffice to measure hue responses reliably and generate results that agree well with subjective rating methods, thus circumventing some of the issues often raised with direct scaling methods [20]. This method is typically used to estimate perceptual scales associated with the integration of information along multiple dimensions [21,22]. Here we used MLCM to assess the influence of Munsell hue, value (lightness), and chroma (saturation) on warm/cool judgments, initially employing a three-way design and subsequently a two-way design when one of the stimulus dimensions was found to influence the judgments negligibly.

Another advantage of this method is that the stimulus factors influencing the ordinal judgments between stimulus pairs can be modeled within a signal detection framework [19,23]. A drawback of the MLCM approach in the current experiments is that it only allowed the estimation of the relative shape of the warm/cool psychological response. To address this deficiency, we also employed an identification task for single stimulus presentations. As the experimental design required a large number of models to be evaluated, an information criterion was used in each experiment for model selection instead of multiple testing of nested models that could have increased bias from Type I error [24].

2. GENERAL METHODS

A. Observers

A total of 18 observers were tested in the experiments and only one (an author) completed the full set of conditions. All observers except the author remained naïve over the course of the experiments. Each observer returned for several sessions to complete the conditions in which he or she participated. We tested four participants for each MLCM task and 12 participants for the identification task. Ages ranged from 20 to 49 years. All observers had normal color vision as assessed by a Farnsworth Panel D15 and had normal or corrected-to-normal visual acuity. All experiments were performed in accordance with the principles of the Declaration of Helsinki for the protection of human subjects.

B. Apparatus

The stimuli were presented on a 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, Rochester, United Kingdom). The monitor had a diagonal screen size of 22 in., a resolution of 1024×768 pixels, and a refresh rate of 120 Hz. The screen was calibrated using a SpectroCal spectroradiometer with the calibration routines of Cambridge Research Systems. The experimental software was written to generate all stimuli, execute the experimental procedures, and collect responses in MATLAB 7.9 (MathWorks [25]), using the CRS Toolbox extensions (Cambridge Research Systems, Rochester, United Kingdom). A Cedrus RB540 response pad was used to collect observer responses (Cedrus Corporation; San Pedro, CA, USA). The observer position was stabilized by a chinrest so that the screen was viewed binocularly at a distance of 80 cm. Experiments were performed in a dark room.

C. Stimuli

Each stimulus was defined by a circular disk of 4° diameter. The same spatial pattern was used for the MLCM and for the identification task. All stimuli were displayed on a gray background with fixed chromaticity coordinates (CIE x, $y = 0.3128$, 0.3289; $Y = 49.68\ \mathrm{cd/m^2}$).

Stimuli were specified in the Munsell system and then translated to coordinates into CIE xyY using the open-source software R [26] and the **munsellinterpol** package [27]. The main step was to specify our stimuli in the Munsell system using color chips from 5R to 10 RP. To that end, stimuli were defined from chromaticity coordinates in CIE Lab color space and monitor RGB values using the CRS Color Toolbox extensions (Cambridge Research Systems, Rochester, United Kingdom) to present them on the screen. Then, we adjusted these chromaticity coordinates to obtain the Munsell chip specifications using functions from the **munsellinterpol** package. The package functions automatically interpolate the Munsell notations to provide fractional levels of chroma and value that do not correspond to actual Munsell chips.

For the three-way MLCM experiment, 20 evenly spaced levels of hue with two levels of value and two of chroma were pre-selected. Because of gamut limitations, in the first experiment, value and chroma were equated in the low- and high-level conditions for all hues except for four in the high chroma condition (5B, 5PB, 10PB, and 5P). Chroma and value, however, were not equated across hue. The Munsell coordinates along with Lab values for the first experiment are shown in Table 1. In the second experiment using a two-way MLCM design, the same 20 levels of hue were used with two levels of value (7.5 and 5.4) and chroma was held constant at a level of 7.

D. Procedure

Each observer completed the experimental sessions over several days. The observers were adapted to the screen for 3 min at the beginning of each session. There was a practice session of five 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.

1. *Maximum Likelihood Conjoint Measurement Task*

In the three-way MLCM experiment, the 20 hues $\times$ 2 values $\times$ 2 chromas design, generated 80 stimuli. On each trial, two different stimuli were chosen randomly from the 80-element grid, resulting in a total of $(80 \times 79)/2 = 3160$ possible, non-identical pairs. The observers were tested over six sessions of 125

Table 1. Munsell and CIE Lab Values for the 80 Colors Used in the Three-Way MLCM

	High Saturation (High Lightness/Low Lightness)				Low Saturation (High Lightness/Low Lightness)			
Hue	Value (cd/m^2)	Chroma	a^*	b^*	Value (cd/m^2)	Chroma	a^*	b^*
5R	5.4/4.1 (22.37/13.19)	14.2/14.2	70.03/56.97	38.99/30.24	5.4/4.1 (21.68/13.14)	12.2/12.2	50.03/51.11	27.55/27.12
10R	5.9/4.1 (29.64/14.24)	13.4/13.4	54.17/48.65	66.02/54.68	5.9/4.1 (32.35/14.98)	10.9/10.9	38.68/40.8	41.86/41.62
5YR	6.9/5 (48.65/24.31)	13.4/13.4	32.86/28.57	77.6/61.64	6.9/5 (52.25/24.88)	9.9/9.9	24.36/25.78	51.24/52.1
10YR	7.4/5.1 (62.56/27.27)	9.7/9.7	14.18/12.7	81.04/61.6	7.4/5.1 (66.25/27.79)	7.5/7.5	8.48/10.52	47.52/47.89
5Y	7.2/5.2 (62.42/29.95)	9/9	−2.72/−0.4	70.93/62.26	7.2/5.2 (64.17/30.02)	5.6/5.6	−3.28/−1.22	40.31/40.58
10Y	7.5/5.2 (70.92/30.98)	10/10	−15.84/−12.07	80/61.68	7.5/5.2 (71.76/30.98)	7.3/7.3	−12.63/−11.25	52.97/53.11
5GY	7.2/5.3 (67.47/33.90)	10/10	−33.89/−27.95	77.8/60.63	7.2/5.3 (67.64/33.33)	6.7/6.7	−21.45/−21.46	42.51/41.79
10GY	7.2/5.3 (67.78/34.51)	10/10	−54.64/−47.08	48.84/40.54	7.2/5.3 (67.98/33.97)	6.7/6.7	−31.74/−31.4	27.74/26.46
5G	7.2/5.3 (67.48/33.80)	9.8/9.8	−51.43/−40.42	17.04/12.9	7.2/5.3 (68.30/33.63)	6.8/6.8	−35.98/−35.62	12.71/11.64
10G	7.5/5.3 (71.46/33.27)	7.4/7.4	−38.88/−35.55	5.37/3.77	7.5/5.3 (71.39/32.25)	5.4/5.4	−23.97/−23.1	3.66/2.74
5BG	7.5/5.5 (71.64/35.34)	9/9	−44.05/−33.67	−4.68/−5.44	7.5/5.5 (71.85/35.33)	5.6/5.6	−27.96/−26.97	−2.54/−4.18
10BG	7.5/5.4 (70.08/32.99)	7/7	−31.78/−33.67	−13.51/−15.11	7.5/5.4 (70.26/32.96)	5.3/5.3	−19/−17.63	−7.72/−8.99
5B	7.1/5.2 (58.71/29.17)	9/9	−27.84/−21.53	−30.05/−25.88	7.1/5.2 (60.14/29.45)	6.8/6.8	−20.46/−18.86	−20.54/−22.03
10B	6.7/4.8 (49.09/22.77)	10/10	−13.35/−11.61	−35.61/−38.85	6.7/4.8 (51.16/23.87)	7.2/7.2	−10.68/−9.02	−26.16/−27.56
5PB	6.4/4.6 (40.60/18.17)	10/10	2.26/5.82	−39.62/−48.18	6.4/4.6 (43.72/20.79)	7.6/7.6	1.34/2.83	−29.56/−30.89
10PB	6.4/4.6 (39.82/16.45)	11/11	22.06/36.11	−40/−53.32	6.4/4.6 (42.40/18.91)	7.5/7.5	13.91/16.52	−27.19/−28.47
5P	6.4/4.6 (36.67/16.06)	13.8/13.8	42.05/49.31	−40.7/−43.32	6.4/4.6 (40.38/19.06)	9.1/9.1	26.57/29.02	−26.66/−27.26
10P	6.4/4.6 (35.14/16.45)	14/14	64.43/53.45	−31.87/−27.6	6.4/4.6 (36.68/17.32)	12.3/12.3	45.25/47.2	−22.22/−24.29
5RP	5.8/4.1 (25.78/13.26)	12.5/12.5	68.95/53.28	−7.83/−8.19	5.8/4.1 (30.87/14.24)	10.7/10.7	43.21/45.74	−4.02/−6.72
10RP	5.6/4.1 (24.36/12.63)	12.4/12.4	69.32/54.43	13.05/8.04	5.6/4.1 (27.88/13.14)	11.8/11.8	49.48/51.91	10.11/7.82

trials, yielding a total of 750 trials per observer (23.73% random sampling with replacement from the total of all possible pairs).

A two-way MLCM experiment was also performed varying hue and value dimensions with chroma fixed. In each trial, a pair of different stimuli from the 2×20 grid was chosen at random and presented to the observer. There are $(40 \times 39)/2 = 780$ such non-identical pairs. For each observer, all 780 pairs were presented in random order over six sessions of 130 trials each.

For both MLCM experiments, two stimuli were presented, and the observers were instructed to judge which of the two appeared warmer. An inter-stimulus interval between each pair was fixed at 500 ms. Responses were recorded as right or left button presses. The next pair was presented following the observer's button press response.

2. *Identification Task*

The observers classified individually presented color patches as warm or cool. The dimensions varied were hue and value with chroma fixed as in the two-way MLCM. Each stimulus was presented 18 times for each condition and tested in random order. Six sessions consisting of 120 trials each were run for a total of 720 trials (20 hues $\times$ 2 values $\times$ 18 repetitions). On each trial, a color patch was presented in the center of the screen, and the subsequent trial was activated by pressing the appropriate button. At its offset, a blank screen appeared for 500 ms.

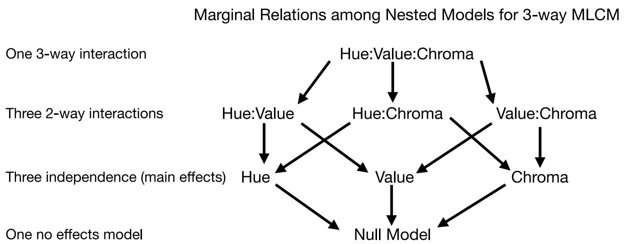

Fig. 1. All possible model terms and their marginality relations. For a given term, the tree of arrows emanating from it indicates the marginal terms that must be retained for nested model comparisons. For example, a model including the hue:value interaction must also include the hue and value main effects terms. The ":" between dimension names indicates an interaction. The null model corresponds to a model in which the observer responses are independent of any of the dimensions.

E. Data Analyses

All analyses were performed using the open-source software R [26] with functions from the MLCM package [28]. In analyzing the data from an MLCM experiment with n dimensions, there are 2^n possible models (including the null model in which the observer's choices are independent of all dimensions) since each term in the model can be included or excluded. Among these, the nested model comparisons are between pairs for which the terms of one of the models are a proper subset of the other. A simple counting argument leads to the result that for n dimensions there are

$$\sum_{k=0}^{n-1} \binom{n}{k} (2^{n-k} - 1)$$

possible models that conserve the marginality of lower-order terms. These are the models that can be used in nested comparisons with adjacent models, i.e., that make comparisons to models that eliminate just the highest-order terms(s). While for two dimensions, five such nested models are possible, for three the number grows to 19. Figure 1 shows the marginal relations of the terms of the three-way model. To avoid performing multiple statistical tests that can lead to an increase in Type 1 error, all possible models were fit (five in the case of two-way MLCM and 19 for three-way MLCM), and the best model was chosen as the one that minimized the Akaike information criterion (AIC) [29]. The AIC is defined as minus twice the log-likelihood plus twice the number of free parameters estimated in the fit so that models with too many parameters are penalized at the expense of goodness of fit. This procedure, however, selects a model that minimizes prediction error [30].

3. RESULTS

A. Results of Three-Way MLCM (Hue, Value, and Chroma)

Data were analyzed as a decision process within the framework of a signal detection model and fitted by maximum likelihood, as described elsewhere [17,18]. Results indicated that the minimum AIC models were "H + V" for Observer 1, "H + V + C + H:V + C:V" for Observer 2, "H + V" for Observer 3, and "H + V + C" for Observer 4. The models vary among observers, but all include both hue and value main effects. Observers 1 and 3 show additive main effects of hue and value, so the value does not change the *relative* contribution across hue, i.e., it only causes a constant vertical shift across hue (see Fig. 2). For Observer 4, the main effect of chroma also does

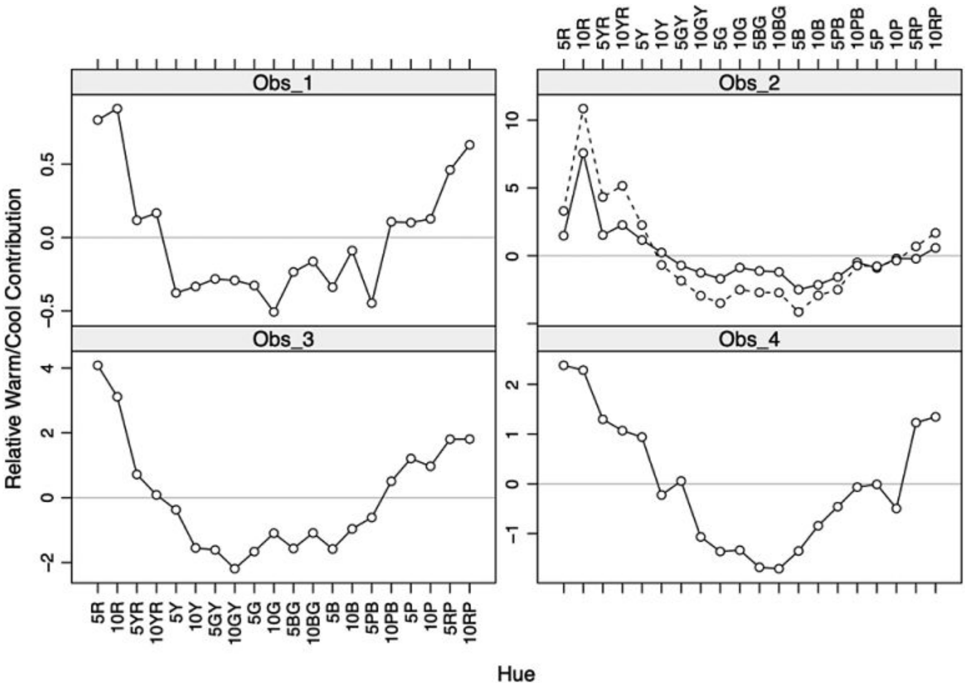

Fig. 2. Estimated scale obtained for the three-way MLCM procedure (hue, value, and chroma). Each panel shows the average estimates for warm versus cool responses as a function of the hue in Munsell color space. The two curves for Observer 2 indicate the change of shape of the hue curve with value due to the interaction of these two stimulus features. The solid line corresponds to the lower level of value and the dashed to the higher.

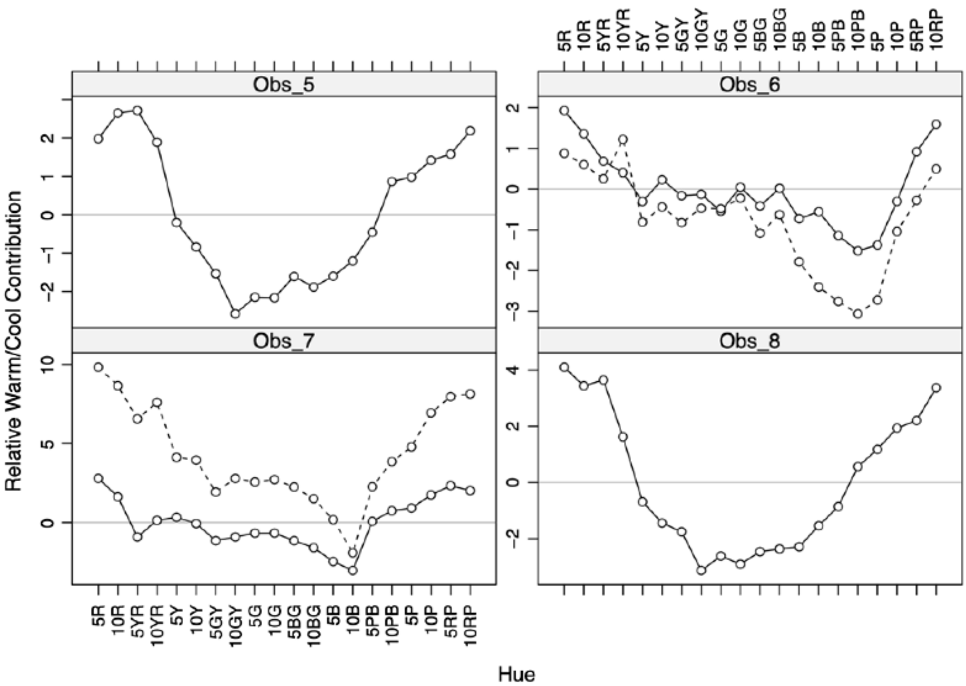

Fig. 3. Estimated scale obtained for the two-way MLCM procedure (hue and value). The solid curves show the average estimated scales under the additive model for each observer. The dashed lines for Observers 6 and 7 indicate the estimated interaction of value on the shape of the hue curve. The values on the abscissa indicate the hue values as displayed along the Munsell color system. The two curves for Observers 6 and 7 indicate the change of shape of the hue curve with value due to the interaction of these two stimulus features. The solid line corresponds to the lower level of value and the dashed to the higher.

not change the relative contribution across hue. Only Observer 2 shows an interaction between hue and value, indicating that the hue contribution depends on the level of the value. While the observers showed individual differences in the model that best predicted their judgments, all showed a systematic effect of hue and a small effect of value. Effects of chroma, if present, were also systematically small.

Since the shape of the hue component does not depend on the level of value and chroma for Observers 1, 3, and 4, it is possible to display its shape independent of these dimensions. MLCM provides only the relative contribution of each dimension, so the warm/cool transition level is unknown. Panels in Fig. 3 show the average estimated scales for each of the four observers, centered vertically on the zero level (gray line). Note that the two curves for Observer 2 indicate the slight dependence of hue on value due to the interaction term. For all observers, the dependence on hue is a U-shaped curve, suggesting that the warm/cool dimension divides color space into two regions. These results are based only on ordinal judgments between stimulus pairs, but the individual curves and averages are consistent with curves obtained by direct estimation [12,15].

B. Results of Two-Way MLCM (Hue and Value)

In the previous experiment, the results showed that chroma made a negligible if any contribution to the warm/cool judgments. In the two-way MLCM design, there were 20 levels of hue and two levels of value with chroma fixed across conditions. With two dimensions, the possible number of nested models is reduced to five. The results of model selection for all observers, using the AIC criterion, are presented in Table 2 with the

Table 2. AIC Values of All Observers for Each of the Five Models from the Two-Way MLCM Procedure

Model	Obs. 5	Obs. 6	Obs. 7	Obs. 8
Null	1081.0	1064.1	839.9	1080.4
V	1014.0	991.4	766.2	1017.5
H	535.4	947.5	911.0	492.9
H + V	**361.0**	834.1	423.7	**288.8**
H * V	367.8	**740.9**	**344.5**	297.4

minimal AIC values in bold. Data indicate that the additive model is best for Observers 5 and 8, while the full or saturated model produces the lowest AIC value for Observers 6 and 7.

Scales estimated from the MLCM procedure are displayed in Fig. 3, using the same format as the previous figures. For Observers 6 and 7, the curves estimated for the lower value are centered vertically on the zero level (gray line) as in the *three-way MLCM*, because the ordinal judgments leave the vertical offset undetermined. For Observers 5 and 8, no interaction with value was necessary (additive model), and a single curve is shown in each panel. The two curves for Observers 6 and 7 indicate the effect of value on the shape of the hue curve. The results support that observers agree generally on the assignment of warm/cool judgments across hue while differing considerably on whether and how value contributes to warm/cool appearance.

C. Results of Identification Task (Hue and Value)

MLCM permits assessing the contributions of color dimensions to warm/cool judgments but only allows the relative

Table 3. Results of the AIC for the Identification Task

Model	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5	Obs. 6	Obs. 7	Obs. 8	Obs. 9	Obs. 10	Obs. 11	Obs. 12
H	473.9	**163.2**	903.8	**340.9**	703.4	**361.1**	652.3	326.9	**249.1**	333.1	663.0	436.1
H + V	**348.6**	165.2	373.8	342.8	510.0	363.0	**549.6**	**312.7**	250.2	**298.7**	464.5	**300.0**
H * V	369.2	194.6	**368.5**	354.2	**383.4**	370.4	570.9	325.6	271.8	317.7	**461.5**	332.6

dependence of the hue curve shape to be determined. To circumvent this limitation, we employed an identification task in which observers are presented with individual stimuli, varying independently in hue and value, and are asked to classify each stimulus as warm or cool. The decision rule for the identification task can be modeled as

$$\Delta_{\mathrm{HV}} = \psi_{\mathrm{WC}}\left(\phi_{\mathrm{HV}}\right) + \epsilon,$$

where the decision variable, Δ_{HV}, is a function of the perceptual response, ψ_{WC}, that depends on the physical hue and value of the stimulus, ϕ_{HV}, and is perturbed by Gaussian noise, ϵ, of mean 0 and variance σ^2. Arbitrarily, we assume that when the decision variable is greater than 0, the observer identifies the stimulus as warm, otherwise cool. The perceptual response is estimated by maximum likelihood using probit regression to estimate the probability of choosing warm, where the physical stimulus is expanded into separate columns of indicator variables in a model matrix for the main effects of the hue and value levels and their interaction, depending on the model fit to the data. Continuing with the logic of the previous experiments, we used AIC for model selection. Three nested models were fitted to each observer's data. The first model depends on hue, the second model is additive with respect to hue and value, and the last model includes an interaction term.

The AIC values are shown in Table 3 with the lowest values in bold. The model with lowest AIC for Observers 3, 5, and 11 included a hue and value interaction. AIC is the lowest for the additive model for Observers 1, 7, 8, 10, and 12, while for Observers 2, 4, 6, and 9, the hue contribution is independent of the value. Thus, for 9 out of 12 observers, there is no interaction of value with hue.

Figure 4 shows the comparison of the average hue response for the model best fit by AIC. Here, Observers 3, 6, and 8 were left out of the average because of the interaction of hue with value.

The results from the identification task provide absolute estimates of the hue response. Maximal warm appearance falls between 10R and 5YR and maximal cool over a relatively flat region around 5 GB. The curve crosses zero at two hues, 5Y and between 5P and 10P. Although the tendency is for reddish and yellowish hues to be judged as warm and bluish and greenish as cool, purplish (i.e., with a red component), yellowish, and bluish hues are found in both warm and cool regions.

Figure 5 shows the estimated warm/cool responses as a function of hue for the three observers whose judgments were best predicted by a model including a hue:value interaction. The two curves in each panel indicate how the response curves were affected by the value level of the stimuli. Not only do the curves depend on the value level, but they suggest that the observers do not classify colors along a warm/cool continuum in a manner similar to the other nine observers.

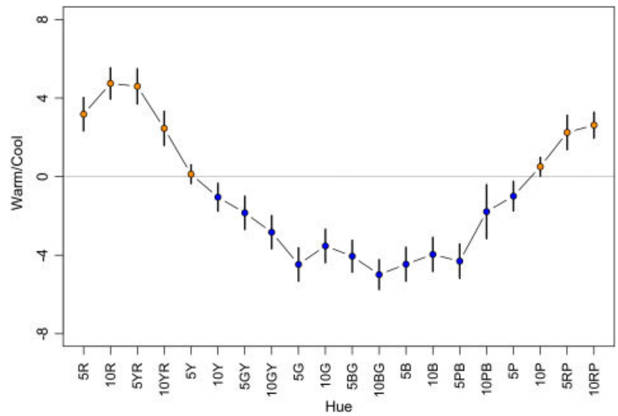

Fig. 4. Average responses for the model best fits by AIC, excluding Observers 3, 5, and 11 whose judgments revealed a hue:value interaction, so the hue curve shape depended on the level of value. Orange points are those classified as warm and blue cool. The abscissa indicates the hue values as displayed along the Munsell color system. Error bars represent SEs.

4. DISCUSSION

Several previous studies have estimated warm/cool scales by using subjective ratings [7,12,14–16]. Our results are consistent with these in showing that, on average, hues between 5PB/10P and 5Y were estimated to be warm and hues between 10Y and 10B/5P cool. As we have argued elsewhere [20], the ordinal judgments involved using MLCM are generally easier for subjects to perform and are likely to be less impacted by subjective biases related to culture, linguistics, and internal representations of how subjective magnitude translates to numerical values. In addition, Vincent *et al.* [22] have demonstrated using simulation that the scales obtained by MLCM accurately estimate the shape of internal perceptual scales, while matching procedures do not lead to a unique estimate.

An advantage of the MLCM procedure is that the influence of and potential interactions between several dimensions on warm/cool judgments can be estimated simultaneously. The perceptual scales generated through the MLCM procedure, however, are only unique up to an additive constant and a multiplicative factor [19]. Thus, the MLCM procedure only permits estimating the relative shape of the warm/cool curve with respect to the dimensions probed. It is for this reason that we also performed an identification task from which we could also estimate the influence of multiple factors.

While we did not obtain warm/cool scales based on direct estimation, it is reassuring to find that the scales that we obtained based on identification judgments do largely correspond to those obtained via subjective ratings by others [12,13,15]. Our results also agree with the average results of Koenderink *et al.* [17] who used ordinal judgments to assess warm/cool relations

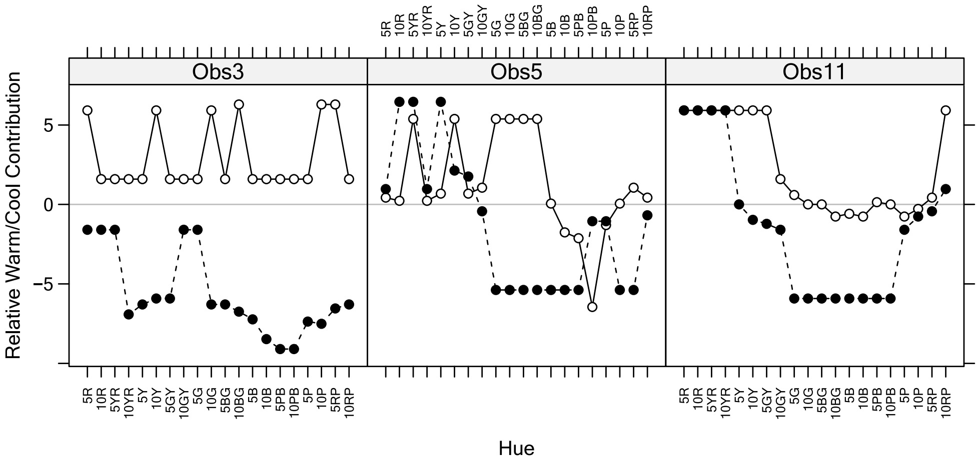

Fig. 5. Estimated warm/cool responses for each level of value (white filled circles: value = 5.4; black filled circles: value = 7.5) of the three observers who displayed a hue:value interaction in the identification task.

for adjacent hues around a color circle. Importantly, the ordinal judgments performed here involve comparing stimuli that covary independently along several dimensions and do not only involve stimulus pairs that are adjacent in color space. In addition, we demonstrate feasible model selection methods for analyzing MLCM paradigms more complex than two-way experimental designs, and that avoid having to correct for multiple testing [31].

While there were large individual differences between observers concerning what dimensions and their interactions best predicted the observers' judgments, we found that warm/cool judgments largely were little affected by chroma level and do depend on value for most observers. This contrasts with Katra *et al.* [12] who found a slight dependence of subjective ratings on saturation and Manalansan and Webster [15] whose results were generally independent of luminance level. In contrast, a recent study by Hammond *et al.* [16] displays a systematic effect of luminance on average warm/cool estimates for lights of different hue. A difference between our study and that of Katra *et al.* [12] is that they used stimuli specified in NCS coordinates and we specified our stimuli according to Munsell designations. While both spaces are, in principle, based on perceptual color spacing, they are not identical, and the discrepancies may have arisen from differences in the two spaces. In our first experiment, however, influences of chroma were slight even with variations between value between hues, and the dependence on value persisted for most subjects even when chroma was held constant in experiments 2 and 3.

With respect to Manalansan and Webster [15], we note that luminance is not the same as value. In their experiment, observers were confronted with a single luminance level within a run while in our experiments, observers had the opportunity to compare or experience different levels of value during the same run, which may have influenced their judgments differently.

We also note that the amplitude of the estimated warm/cool response varied considerably across observers. The same trend can be seen in the Katra *et al.* [12] study, in which about one-quarter of the observers display a nearly flat warm/cool response. Similar individual variations were found by Manalansan and Webster (personal communication) and occur in the data from individual observers of Hammond *et al.* [16] (available from them upon request). Koenderink *et al.* [17] report that almost half of their observers were unable to make coherent warm/cool judgments. For those that could, their results showed a correlation with a peak for warm responses at red or yellow and a peak for cool responses around a cyan color. Their results indicated that the warm/cool distinction is not systematically reflected in observer responses, and there is uncertainty in assessing adjacent hue steps.

Three of our observers in the identification experiment behaved similarly (Fig. 5). Anecdotally, two of our observers from experiment 3 (Observers 3 and 5) who displayed atypical warm/cool responses were classified as having a color synesthesia, based on an online test battery ([32]) [33,34]. Our sample is too small and the method *post hoc* to make a reliable association here, but the suggestion warrants further study as such interference in warm/cool judgments would imply a more cognitive basis for the dimension. Perhaps as Briggs [35] ([36]) has argued, the warm/cool dimension, itself, corresponds to a form of color/temperature synesthesia.

Funding. LABEX CORTEX (ANR-11-LABX-0042, ANR-11-IDEX-0007).

Acknowledgment. K. K. was supported by LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon (ANR-11-IDEX-0007) operated by the French National Research Agency. We thank Michael Webster for feedback and suggestions on an earlier draft of the paper.

Disclosures. The authors declare no conflicts of interest.

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.

REFERENCES

1. J. J. Koenderink and A. van Doorn, "Orange & teal," Art Percept. **9**, 134–136 (2021).
2. C. L. Hardin, "Red and yellow, green and blue, warm and cool: explaining color appearance," J. Consc. Stud. **7**, 113–122 (2000).
3. D. T. Lindsey and A. M. Brown, "Universality of color names," Proc. Natl. Acad. Sci. USA **103**, 16608–16613 (2006).

4. L. C. Ou, M. R. Luo, A. Woodcock, *et al.*, "A study of colour emotion and colour preference. Part I: colour emotions for single colours," Color Res. Appl. **29**, 232–240 (2004).
5. N. Zaslavsky, C. Kemp, N. Tishby, *et al.*, "Color naming reflects both perceptual structure and communicative need," Topics Cog. Sci. **11**, 207–219 (2019).
6. X. P. Gao, J. H. Xin, T. Sato, *et al.*, "Analysis of cross-cultural color emotion," Color Res. Appl. **32**, 223–229 (2007).
7. L. Albertazzi, J. J. Koenderink, and A. van Doorn, "Chromatic dimensions earthy, watery, airy, and fiery," Perception **44**, 1153–1178 (2015).
8. Y. Xiao, C. Kavanau, L. Bertin, *et al.*, "The biological basis of a universal constraint on color naming: cone contrasts and the two-way categorization of colors," PloS one **6**, e24994 (2011).
9. I. A. Rosenthal, S. R. Singh, K. L. Hermann, *et al.*, "Color space geometry uncovered with magnetoencephalography," Curr. Biol. **31**, 515–526 (2021).
10. B. R. Conway, S. Ratnasingam, J. Jara-Ettinger, *et al.*, "Communication efficiency of color naming across languages provides a new framework for the evolution of color terms," Cognition **195**, 104086 (2020).
11. E. R. Heider, "Universals in color naming and memory," J. Exp. Psychol. **93**, 10–20 (1972).
12. E. Katra, B. R. Wooten, and K. Knoblauch, "Perceived lightness/darkness and warmth/coolness in chromatic experience," PsyArXiv, (2023).
13. K. Knoblauch, J. S. Werner, and M. A. Webster, "Warm and cool reheated," Color Res. Appl. **48**, 814–817 (2023).
14. B. Wright, "The influence of hue, lightness, and saturation on apparent warmth and weight," Am. J. Psychol. **75**, 232–241 (1962).
15. J. Manalansan and M. A. Webster, "The warm-cool color dimension aligns with asymmetries in color perception implicit in uniform color spaces," J. Vis. **23**(9), 5572 (2023).
16. B. R. Hammond, C. R. Gardner, B. R. Wooten, *et al.*, "Increasing intensity directly increases the perceived warmth of primary colors," Sci. Rep. **14**, 26852 (2024).
17. J. J. Koenderink, A. van Doorn, and D. Braun, ""Warm," "Cool," and the colors," J. Vis. **24**(7), 5 (2024).
18. Y. X. Ho, M. S. Landy, and L. T. Maloney, "Conjoint measurement of gloss and surface texture," Psychol. Sci. **19**, 196–204 (2008).
19. K. Knoblauch and L. T. Maloney, *Modeling Psychophysical Data in R* (Springer, 2012).
20. F. Devinck and K. Knoblauch, "Color appearance of spatial patterns compared by direct estimation and conjoint measurement," J. Opt. Soc. Am. A **40**, A99–A106 (2023).
21. L. T. Maloney and K. Knoblauch, "Measuring and modeling visual appearance," Annu. Rev. Vis. Sci. **6**, 519–537 (2020).
22. J. Vincent, M. Maertens, and G. Aguilar, "What Fechner could not do: separating perceptual encoding and decoding with difference scaling," J. Vis. **24**(5), 5 (2024).
23. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, 1966).
24. R. J. MacCoun, "P-hacking," in *Research Integrity*, L. J. Jussim, J. A. Krosnick, and S. T. Stevens, eds. (Oxford University, 2022), pp. 295–313.
25. http://mathworks.com.
26. R Core Team, "R: A Language and Environment for Statistical Computing" (R Foundation for Statistical Computing, 2023), https://www.R-project.org/.
27. J. Gama, P. Centore, and G. Davis, "Munsellinterpol: Interpolate Munsell Renotation Data from Hue/Chroma to CIE/RGB, R Package Version 3.0-0" (2022), https://cran.rproject.org/web/packages/munsellinterpol/index.html.
28. K. Knoblauch, L. T. Maloney, and G. Aguilar, "MLCM: Maximum Likelihood Conjoint Measurement, R Package Version 0.4.3" (2019), https://CRAN.R-project.org/package=MLCM.
29. H. Akaike, "Akaike's information criterion," in *International Encyclopedia of Statistical Science*, M. Lovric, ed. (Springer, 2011), p. 25.
30. S. N. Wood, *Core Statistics* (Cambridge University, 2015), Vol. **6**.
31. H. C. Sun, D. St-Amand, C. L. Baker, Jr., *et al.*, "Visual perception of texture regularity: conjoint measurements and a wavelet response-distribution model," PLoS Comput. Biol. **17**, e1008802 (2021).
32. https://synesthete.ircn.jp/home.
33. D. A. Carmichael, M. P. Down, R. C. Shillcock, *et al.*, "Validating a standardised test battery for synesthesia: does the Synesthesia Battery reliably detect synesthesia?" Consc. Cog. **33**, 375–385 (2015).
34. D. M. Eagleman, A. D. Kagan, S. S. Nelson, *et al.*, "A standardized test battery for the study of synesthesia," J. Neurosci. Methods **159**, 139–145 (2007).
35. D. J. C. Briggs, "The Dimensions of Colour, Modern Colour Theory for Traditional and Digital Painting Media," http://www.huevaluechroma.com.
36. http://www.huevaluechroma.com/077.php.