IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 31, 2022, accepted February 18, 2022, date of publication March 4, 2022, date of current version March 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3156928

User Study Comparing Linearity and Orthogonalization for Polarimetric Visualizations

ANDREW W. KRUSE[1], DAMIEN J. MANNION[2], ANDREY S. ALENIN[1], AND J. SCOTT TYO[1], (Fellow, IEEE)

[1]School of Engineering and IT, University of New South Wales Canberra, Campbell, ACT 2600, Australia
[2]School of Psychology, University of New South Wales Sydney, Kensington, NSW 2052, Australia

Corresponding author: Andrew W. Kruse (andrew.w.kruse@gmail.com)

This work was supported in part by the Australian Government Research Training Program Scholarship, and in part by the Australian Government through the Australian Research Council (DP170100087 to DM).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the UNSW Human Research Advisory Panel C: Behavioural Sciences with application no. HC 3880 in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007).

ABSTRACT Traditionally, polarimetric imaging data is visualized by mapping angle of polarization, degree of polarization, and intensity to hue, saturation, and value coordinates of HSV color space. Due to possible perceptual uniformity issues in HSV, a method based on CAM02-UCS color space has been recently proposed. In this user study, the perceptual uniformity and nonlinear bias of the encoding of the degree of polarization parameter into the chromatic magnitude color channel is modeled by a power-law relationship between stimulus scale level and is estimated from responses to paired 2-alternative forced choice questions using Maximum Likelihood Difference Scaling. Estimated exponent and noise parameters for these methods are compared for same-hue and different-hue conditions to determine whether the chromatic magnitude channel can be used to orthogonality encode data parameters independently from the hue channel. Overall, the HSV condition displayed more nonuniformity, more nonlinear bias, and more non-orthogonality than the UCS condition. The results here indicate a lower bound for differences between methods since the intensity was chosen for the ''best case'' of HSV. These results further support the claim that the chromatic magnitude color channel of a uniform color space can be used to encode a data parameter independently of the hue channel in a multivariate colormapping visualization.

INDEX TERMS Polarimetry, optical polarization, data visualization, visualization, color, psychometric testing, user centered design.

I. INTRODUCTION

Polarization imaging is a field within the optical sciences in which the characteristics of the polarization of light are measured across a two-dimensional image plane. Measuring the polarization of light in a scene offers unique information not present in monochromatic or spectral imaging, particularly concerning the interaction of light with the surface of materials. A typical polarization imaging system involves pairing a camera system and a set of polarization filters, and the polarization parameters are derived by taking measurements at different orientations of the filters. Since humans are insensitive to polarization, the parameters derived from the measurements must be converted into the domain of human vision to be visualized. Conveniently, human color vision has a similar cylindrical coordinate system to polarization parameters. This observation, first made by Bernard and Wehner [1], has led to the development of several related colormapping strategies for polarization imaging. Walraven produced the first known formalism for this type of colormapping [2]. A few years later, Solomon defined a formalism for mapping the polarization parameters into a color space designed for perceptual uniformity [3]. Wolff and Mancini later defined a formalism using the

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo.

HSV colorspace for a digital polarization camera [4]. Most instances of this family of colormapping strategies use this formalism to this day. Tyo, *et al.*, added a modification to this method by amplifying dark, polarized regions [5]. Neumann, *et al.*, designed a colormapping strategy like Solomon's but instead using a more modern color space (CIE Luv) with display gamut-mapping considerations [6]. Kruse, *et al.*, designed a strategy using the CIECAM02-UCS color space [7], including display gamut-mapping as well as a method for improving visibility of dark, polarized regions [8]. Zhao, *et al.*, applied those strategies defined by Kruse, *et al.*, for the JzAzBz color space to improve the mapping of the achromatic channel [9].

While all of these colormapping strategies share a foundation in converting polarization to a cylindrical color space, they differ in the choice of color space among other significant algorithmic differences. Color has been studied in detail as a channel for data visualization. Ware described how color parameters affect reading data values and observing form in regards to color sequences [10]. Guidelines on how to use color to assist the interpretation of data are well documented by researchers such as Brewer [11], Rheingans [12], and Munzner [13]. Smith and van der Walt notably use these guidelines to design the default colormaps for the Python module matplotlib [14]. Research surveying the landscape of the types and uses of colormaps have been conducted by Zhou and Hansen for the one-dimensional case [15], and by Bernard, *et al.*, for the two-dimensional case [16]. The particular way in which color is used can have dramatic effects on the results of user studies [17], [18].

Previous work by Kruse, *et al.*, had been aimed at analyzing polarization colormapping strategies by using a perceptually uniform color space [8], [19]. The analysis primarily used the criteria of perceptual uniformity and parameter orthogonality to compare colormapping strategies. Perceptual uniformity is the characteristic of having the relative difference between data values produce proportional differences in the observer's perception of the differences in the visualization. Sometimes, this characteristic is referred to as perceptual *linearity*, especially when the color sequence follows a particular progression in an ordered visual encoding channel. For the purposes of this study, the terms uniformity and linearity are interchangeable. Parameter orthogonality is the characteristic of having the different measured data variables be mapped to color channels that are independently perceived by the observer; changes in one variable do not affect the perception of another variable. While uniformity is a characteristic that can be used to describe both univariate (single data variable) and multivariate (multiple data variables) colour sequences alike, orthogonality only applies to multivariate sequences.

The application of perceptually uniform color spaces for addressing perceptual (non)uniformity in colormaps is widely accepted as effective in visualization literature [11], [20]–[23]. This is not a trivial claim that ''uniform color has uniformity''. Rather, it is a claim about the transferability of models used to predict responses to stimuli *under certain conditions* to improving the performance of tasks on representations of data. In the case of uniformity, it indeed seems to be that the color models have this transferability. Perceptually uniform colorspaces have been effective for modeling some results of user studies on colormaps [24], and colormaps designed in uniform color spaces have better performance for data visualization tasks related to uniformity in some user studies [17], [18]. Moreland mathematically described a uniform color sequence as a path in 3D color space that has a constant derivative [25]. Bernard, *et al.*, defined a quality assessment metric of uniformity for multivariate maps as the standard deviation of a distribution created by randomly sampling the color sequence at many points and determining the ratio of the color difference (ΔE) between sampled colors (c_1, c_2) and the difference in data values (p_1, p_2) that the sampled colors represent [16]. Similarly, Bujack, *et al.*, defined a quality assessment metric for univariate maps that instead includes all the possible combinations of colors in the sequence rather than just a sample [22].

For orthogonality, the transferability claim has not been settled in visualization literature. Some researchers have expressed reservations on the effectiveness of multivariate colour sequences in general [16], [21], [26]. Despite this, the guidelines from Trumbo [27], Brewer [11], and Robertson and O'Callaghan [28] on bivariate colourmapping include orthogonality requirements, which inherently assume the transferability of using orthogonal coordinates of colour models to encode separate aspects of multivariate data. Experiments conducted by Wainer and Francolini in 1980 revealed some difficulty in reading specific bivariate colourmaps [26], but their conclusions should not be extended to claim that orthogonal perception of data values for any and all multivariate colourmaps is impractical or unattainable.

The purpose of this study is to experimentally verify claims made in the analysis by Kruse, *et al.*, [8] that a trivariate color sequences for polarimetric imaging data designed using a perceptually uniform color space would have greater perceptual uniformity and orthogonality than a similar color sequence designed using a non-perceptual color space. The claim of improved uniformity can safely be regarded as accurate, given the wide acceptance of using uniform color spaces for improving colour sequence uniformity. Still, experimental verification of this claim for the particular application to trivariate polarimetric imaging is warranted. The claim of improved orthogonality is less supported, so the primary focus of this study is to determine whether a method designed using perceptually uniform color space exhibits stronger orthogonality than a method designed using a non-perceptual color space. Specifically, the uniformity of the trivariate colour sequences with respect to one of the data variables is measured when the other variables are constant and when they are not constant. Thus, if the constancy of the other data variables has an effect on the measure of uniformity for a particular method, then the claim of orthogonality must be rejected.

II. THEORY

A. POLARIZATION IMAGING

Light is understood as transverse electromagnetic waves with fields that oscillate in a direction perpendicular to the direction of propagation. When the oscillation is deterministic in a particular orientation, the light is said to be polarized. The oscillation can remain fixed at a certain angle (linear polarized), or can rotate at a constant rate (circularly and elliptically polarized). In this study, we are only considering linear polarization. When the oscillation is in uniformly random directions, the light is unpolarized. Typically, light will not be purely polarized or purely unpolarized; the distribution of orientations is neither uniform nor only nonzero for a single orientation. When one orientation is more prominent than others, the light is said to be partially polarized. Partially polarized light can be modeled as a combination of polarized light and unpolarized light. To quantify the partial polarization, the intensity of the polarized part is divided by the total intensity to give a ratio.

$$P = \frac{I_p}{I} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{1}$$

where the ratio P is a measure of the degree of linear polarization (DoLP), I_p is the intensity of the linearly polarized portion, I is the total intensity, and I_{max} and I_{min} are respectively the maximum and minimum intensity values that could be measured through a linear polarizer as it is rotated through 180 degrees. All partially polarized light has a DoLP between 0 (unpolarized) and 1 (fully polarized). Besides the DoLP, the state of polarization of light is also described by the angle of polarization (AoP), which is the angle of the plane of oscillation of the light wave with respect to an arbitrary set of 2D axes perpendicular to the direction of propagation. Equivalently, this would be the angle of the polarizer achieving I_{max}, and would be perpendicular to the angle of the polarizer achieving I_{min}. When the light is unpolarized, the AoP is undefined since there is no plane of oscillation. Since the oscillation of the wave is symmetric about its axis, the AoP only needs to be defined on a half-circle (π) period rather than a full circle (2π). Note that a sensor that measures only linear polarization is unable to distinguish circular polarization from unpolarized light.

The underlying physical mechanisms that produce polarized light are varied, but typically involve an a unpolarized light source interacting with a medium to produce partially polarized light. The most common sources of polarized light in the visible spectrum one would come across under normal circumstances are the sky (Rayleigh scattering [29]), and light reflecting off of smooth surfaces (Fresnel reflections). Other sources of polarized light include rainbows, LCD monitors, and polarized sunglasses. Since the mechanisms that produce polarization are varied, the applications for measuring polarization are also varied. Polarimetric imaging can be used to study the clouds of nebulas [30], to detect man-made targets [31], for detecting cancer cells [32], to draw patterns in the distribution of particles in the sky [33], and to study the behavior of animals with polarimetric vision [34], among many others. More applications and discussions can be found in the review of polarimetric imaging applications from Snik, *et al.*, [35], and in the review of polarimetric imaging techniques by Tyo, *et al.* [36]

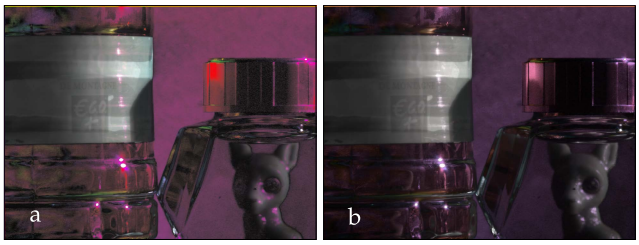

FIGURE 1. Example of mapping methods a: M1 and b: M2 on polarimetric camera data in the near-infrared. Data from multispectral polarimetric image database [37].

B. VISUALIZATION METHODS

In order to visualize the linear polarization parameters (intensity, DoLP, and AoP), the parameters need to be mapped into visual encoding channels. A single visualization technique that would be appropriate for each and every application of polarimetric imaging is not feasible. However, in the general case, there are choices for visualizing the intensity, DoLP, and AoP image data that have broad applicability. When a user wants to "see" the image data, there are certain implied tasks that the user would expect to be able to perform on the image data set that are shared for most applications. For the intensity image data, users should be able to detect form, textures, shapes, and everything else associated with monochromatic vision such that the users have a concrete reference for localizing the polarimetric information. For example, users should be able to recognize that there is a water bottle in the measurements visualized in Fig. 1, and be able to reference the other polarization parameters with respect to its features, e.g. the DoLP of the light reflecting off of the label. For DoLP, form detection is also a common task, since there can be spatial features in the DoLP that are not detectable in the intensity. Still, the primary task that is performed on DoLP image data is quantitative comparison of the strength of the DoLP of the features, e.g. determining that the light reflected off of one object is more polarized than the light reflected from another object. The ability to identify and look up DoLP values can be useful, but is usually not essential for understanding the important aspects of the polarimetric measurement. AoP, although a continuous parameter, is often described using more categorical terms such as "horizontal" or "vertical". Since it is a cyclic parameter, comparative judgements are not applicable. Instead, the primary tasks that a user would be performing on AoP image data are identification and look-up. That is, user should be able to determine whether features have similar AoP values as well as determine what those

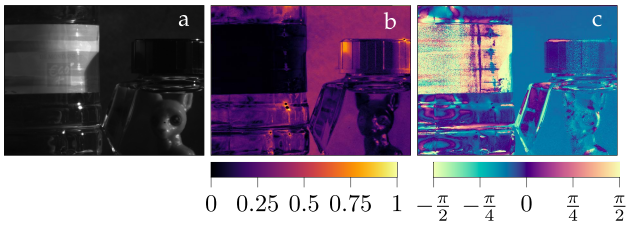

FIGURE 2. Example of visualizing polarization parameters a: intensity, b: DoLP, c: AoP using univariate color sequences. Data from multispectral polarimetric image database [37].

values are using a legend or key. Two features having a similar AoP value can be indicative that they share important physical properties such as surface orientation. Conversely, having different AoP values can be indicative of having different physical properties, even when the difference in value is small. Of all the polarization parameters, this property is unique to AoP. A longer discussion into why this is the case can be found in the article from Kupinski *et al.* [38]

Most often, each parameter is colormapped using univariate color sequences to produce a set of three individual pseudocolor images. An example of this is shown in Fig. 2. Because of the required tasks and the data structures of the three polarization parameters, the color sequences should be appropriately chosen to maximize the effectiveness of the visualization. Intensity can be colormapped using a perceptually linearized grayscale, DoLP using the ''spiral'' class of color sequences [10], and AoP using a cyclic color sequence such as a color wheel. Although these sets of images are easy to produce using available software packages, researchers using these images to investigate polarization phenomena have found difficulty in examining spatial correlations between the polarization parameters [39]. Methods for producing polarimetric imaging visualizations where the spatial attributes of each parameter coincide are generally either glyph-based or use a trivariate colour sequence [19]. In the glyph-based methods, glyphs, or markers, in the form of lines with their orientations and lengths determined by the polarization parameters of AoP and DoLP, respectively, are overlaid on an intensity image. This method has the advantage of having an intuitive association between the physical characteristics of the glyphs and the geometric description of polarization. The main drawback is that the glyphs must be significantly larger than the size of the pixels in order to be visible such that only sampled or spatially-averaged measurements of the polarization can be displayed.

The method for multivariate colormapping is done by mapping the polarization parameters into a three-dimensional color space. The justification for this method is that both the structure of linear polarization parameters and human color vision can be easily mapped into a cylindrical shape. Color vision can be modeled by an achromatic magnitude (e.g. brightness, lightness) as the longitudinal axis, a chromatic magnitude (e.g. saturation, chroma, colorfulness) as the radial axis, and the hue angle which describes the classification of color (e.g. blue, red, green) as the angular coordinate. For polarization parameters, the intensity can clearly be mapped to the achromatic magnitude. The DoLP as a radial component can thought of as the ''distance'' from unpolarized light, which makes it a clear candidate for mapping into the chromatic magnitude, which describes the radial distance from a neutral gray. Hue angle and AoP are both angular components, and can classify the polarization (e.g. horizontal, vertical) or color (e.g. red, blue). In this study we will compare two similar methods for trivariate colormapping that mainly differ in the color model in which they are designed.

1) METHOD 1 (M1)

Polarization parameters are mapped into the cylindrical coordinates of the HSV space. There are two main variations using HSV space. The older and more commonly used method maps polarization parameters intensity, DoLP, and AoP directly into the Value (achromatic magnitude), Saturation (chromatic magnitude), and Hue coordinates, with a factor of 2 multiplied to AoP to covert its π radian period into 2π [40]. The second method by Tyo *et al.* [5] is a slight adjustment to this method, which corrects for the lack of chroma for low levels of value. This is due to the fact that the measure of chromatic magnitude in the saturation coordinate as defined in HSV is relative to the value coordinate. The correction in this method is to increase the value coordinate when the DoLP is high. Thus, the chroma in the saturation is not limited by the value.

$$H(A) = 2\,A, \quad S(P) = P, \quad V(I,P) = max(I,P) \tag{2}$$

where I, P, and A are the quantities of intensity, DoLP, and AoP. This second method is chosen as M1 since it is more similar to M2 due to the correction. An example of this method applied to real polarimetric image data is shown in Fig. 1a.

2) METHOD 2 (M2)

Polarization parameters are mapped into the cylindrical coordinates of a uniform color space. This more recent method attempts to correct for issues in HSV-based methods that arise from the fact that the HSV space is not based on a perceptual model of human color vision. The reasoning goes that if the polarization parameters are mapped to a more perceptually uniform colorspace, then the tasks required of the resulting visualization will be performed more easily, accurately, and consistently. This reasoning has been used to develop color maps [14], and has evidentiary support from the results of user studies [17], [18]. The mapping functions for this method are defined by Kruse *et al.* [8].

$$\begin{aligned} H(A) &= 2A \\ C(P) &= gP \\ L(I,P) &= (L_1(P) - L_0(P))I + L_0(P) \\ L_0(P), L_1(P) &= \{L : f(L) = C(P)\} \end{aligned} \tag{3}$$

where H, C, and L are the hue, colorfulness (chromatic magnitude), and lightness (achromatic magnitude) coordinates of any uniform color space, f is a gamut-matching curve which defines the boundary of a rotationally symmetric subset, and g is a scaling factor such that $g = max(f)$. In this study, CAM02-UCS [7] is used as the uniform color space since that is what was used originally in the method proposed by Kruse *et al.* [8], although any uniform color space could be equivalently substituted. Like M1, the achromatic magnitude is dependent on the value of P. In this method, the curve c is used to find the values of lightness that allow for colorfulness C to be able to be displayed by a computer. I is then linearly mapped into this range $[L_0, L_1]$ of L. An example of this method applied to real polarimetric image data is shown in Fig. 1b.

III. METHODOLOGY

The purpose of this study is to experimentally measure and compare the uniformity and orthogonality of the two visualizations methods M1 and M2. Specifically, this study is limited to measuring those two aspects with respect to a user's ability to make comparative judgments between DoLP values represented by the chromatic magnitude channel. To measure the uniformity, the comparative judgments made by the users are modeled to determine the consistency of the judgments as well as whether there is a nonlinear bias in the way the comparative judgments are made. To measure the orthogonality, the uniformity was measured both when the judgments were made between DoLP values with same value of AoP, and between DoLP values with different AoP values. Since AoP is mapped into the hue coordinate, the former condition is referred to as the same-hue condition (E1), and the latter is referred to as the different-hue condition (E2).

A. PSYCHOMETRIC FUNCTION

In these visualizations, the DoLP is mapped into the chromatic magnitude channel. Since this is an ordered channel, the primary basic task that can be best accomplished is comparison. That is, observers should be able to accurately judge the relative difference between represented values as well as the order of the values. Similarly, interval comparison tasks are used to determine perceived scaling of stimuli in the Maximum Likelihood Difference Scaling (MLDS) technique [41]. In this technique, observers are presented with two pairs of stimuli and must choose which pair appears to have a greater difference (interval) between them. The pairs can be created from a set of three (triplets) or four (quadruplets) stimuli. The two intervals for triplets are defined using a common midpoint stimulus and a stimulus on each side of the stimulus scale. Quadruplet intervals are two independent pairs of stimuli that do not share stimulus values. Most often, perceptual scaling of stimuli are determined using stimuli with difference intervals around the just-noticeable difference (JND) threshold. The scale produced by these methods may only be accurate for small differences, while large differences are of more importance in data visualization. Thus the MLDS technique, which uses supra-threshold stimulus intervals, may be more accurate for determining scale values that predict large-difference perceived intervals.

The perceived scale can be constructed by parametrizing the stimulus interval. The set of n stimulus values x_0 to x_{n-1} correspond to perceived scale values ψ_0 to ψ_{n-1}, with both x and ψ normalized to the interval $[0, 1]$. For each trial, the observer will be presented with four stimuli, x_i, x_j, x_k, x_l corresponding to scale values $\psi_i, \psi_j, \psi_k, \psi_l$. The observer chooses which pair $(i, j$ or $k, l)$ has the larger perceived interval between them. This choice can be modeled by a decision variable D such that when $D < 0$, the difference in the first pair is judged to be greater, and the second pair when $D > 0$. In the absence of noise, the decision variable is determined just by the differences in scale value and would produce consistent responses for every observer.

$$D(i, j, k, l) = (\psi_l - \psi_k) - (\psi_j - \psi_i) \tag{4}$$

For a real observer, the decision variable must include a noise parameter.

$$D(i, j, k, l) = \Delta_t + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \sigma^2\right) \tag{5}$$

where $\Delta_t = (\psi_l - \psi_k) - (\psi_j - \psi_i)$, with i, j, k, l representing the indices of the quad for trial t, and ϵ is a Gaussian random variable from a zero-mean normal distribution with variance σ^2. For N trials at different quadruplets, the probability of the set of responses $R = \{R_0, \ldots, R_{N-1}\}$ for the set of difference-of-differences $\Delta = \{\Delta_0, \ldots, \Delta_{N-1}\}$ depends on the unfixed scale values $\{\psi_1, \ldots, \psi_{n-2}\}$ and noise parameter σ.

$$\mathbf{P}(R|\psi_1 \ldots \psi_{n-2}, \sigma) = \prod_{t=0}^{N-1} \Phi\left(\frac{\Delta_t}{\sigma}\right)^{1-R_t} \left(1 - \Phi\left(\frac{\Delta_t}{\sigma}\right)\right)^{R_t} \tag{6}$$

where $R_t = 0$ corresponds to choosing the first pair (i, j) as more different, and $R_t = 1$ corresponds to choosing the second pair (k, l) as more different. Φ is the standard normal cumulative distribution function. The derivation for this can be found in the original publication for the MLDS method [41]. Eq. 6 represents the product of Bernoulli-distributed probabilities of the responses in each trial, which is simply the probability that the normally distributed sample ϵ is greater than the difference-of-differences Δ_t. The probability of responses given the parameters is equivalent to the likelihood of these parameters given the responses.

$$\mathcal{L}(\psi_1 \ldots \psi_{n-2}, \sigma|R) = \mathbf{P}(R|\psi_1 \ldots \psi_{n-2}, \sigma) \tag{7}$$

Maximizing this function for scale values $\psi_1 \ldots \psi_{n-2}$ and noise σ gives $n - 1$ maximum likelihood estimators $\hat{\psi}_1 \ldots \hat{\psi}_{n-2}, \hat{\sigma}$.

While testing this model in preliminary runs, we discovered a difficulty with determining the number of stimulus

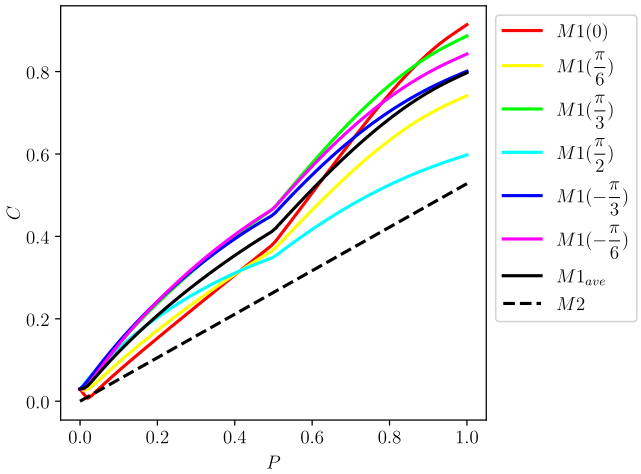

FIGURE 3. Colorfulness (C CAM02-UCS) curves for both methods as a function of DoLP at 0.5 intensity. Colored solid curves are for M1 (Eq. 2) mapped using HSV at 6 different values of AoP corresponding to the primary and secondary display colors of sRGB, with the thick black curve for the average. Dotted curve is for M2 (Eq. 3) mapped using CAM02-UCS, which does not have different curves for different values of AoP since they are by design equal.

levels to use in our trials. On the one hand, too few stimulus levels would not provide enough resolution to get useful information on the shape of the scaling. On the other hand, too many levels created too many parameters, since the expected number of trials would either create scale value parameters with very high uncertainty or would not be sufficient for maximum likelihood convergence. Generally, MLDS uses a large number of trials in order to converge with small errors. Given the speed at which observers performed this task, the number of expected runs would not be sufficient.

In order to have many stimulus levels as well as convergence on estimators, the number of estimators can be reduced by assuming the scale values fall along a parametric curve [42]. Given that the curves of the mapped colorfulness shown in Fig. 3 are roughly exponential, we assume that the scale values ψ will follow a power-law relationship with the stimulus levels x. Although this parametric form will not be able to show small nonuniformities that are expected from the curves in Fig. 3, the major differences between mapping methods can still be measured. We modeled

$$\psi(x) = x^b \tag{8}$$

where b is the exponent parameter. This exponent represents the nonlinear bias of how the user makes comparative judgments on the DoLP values. The left side of Eq. 7 simplifies to $\mathcal{L}(b, \sigma | R_0 \dots R_{N-1})$.

Additionally, in order to describe the probability of user responses, the psychometric function should also include the potential for users to have momentary lapses in judgement. Including this accounts for responses that may be quite inconsistent with the rest of their responses which are not representative of the parameters being determined. This, along with the power-law parametrization, forms the final psychometric function.

$$\begin{aligned} f(b, \sigma, R, \Delta, \lambda) &= \mathcal{L}(b, \sigma | R) \\ &= \prod_{t=0}^{N-1} \lambda + (1 - 2\lambda)\Phi\left(\frac{\Delta_t}{\sigma}\right)^{1-R_t} \\ &\quad \times \left(1 - \Phi\left(\frac{\Delta_t}{\sigma}\right)\right)^{R_t} \end{aligned} \tag{9}$$

where λ is the lapse rate. The lapse rate for each participant is measured as the proportion of responses to a set of 24 quadruplets with maximal difference-of-differences Δ_t that are inconsistent with reasonable values of the two parameters b, σ. The correction in Eq. 9 effectively sets a minimal and maximal probability for each response given the likelihood for a participant to lapse or guess. This set of responses are also used for excluding participants as described in Sec. IV-A.

B. PREDICTIONS

To test the validity of using the colorfulness coordinate in CAM02-UCS for analyzing these visualizations, we can establish a set of predictions for the results we would see if the scale values were purely determined by the colorfulness of the stimuli. Using the color mapping functions in Eqs. 2 and 3, the maximum likelihood estimators for an ideal observer can be simulated for both methods (M1, M2) and hue conditions (E1, E2). The scale values $\{\psi_1, \dots, \psi_{n-2}\}$ for this ideal observer are directly proportional to the colorfulness coordinate C such that their decision variable for each trial in Eq. 5 depends solely on the colorfulness of the quadruplet stimuli after colormapping and the decision noise (Gaussian random variable). Using a set of stimulus quadruplets, the decision variables can be generated for a given noise parameter, leading directly to a set of corresponding responses. The maximum likelihood estimators $(\hat{b},\ \hat{\sigma})$ for the sets of responses and quadruplets can be estimated and recorded. An iteration of an experiment involving a given number of ideal participants can be simulated, and the mean values can be recorded. Running many iterations of simulated experiments, distributions for the means of the estimators can be generated for different levels of decision noise for both methods and hue conditions. These distributions can be used to determine if the experimental means are unlikely to occur under the condition that the scale values for the participants are determined by the colorfulness of the stimulus quadruplets.

Fig. 4 shows the results of 10,000 iterations of these simulations when using the same sample sizes as the real experiment. For the estimator $\hat{b}$, the mean value of the distribution of simulated experiment means is mostly independent of the input decision noise. In contrast, the confidence intervals for both estimator distributions widen as the input decision noise increases, with a more pronounced effect for E2 due to its smaller sample size. Additionally, the confidence intervals are much larger for E2 for the same reason. The confidence intervals here indicate the

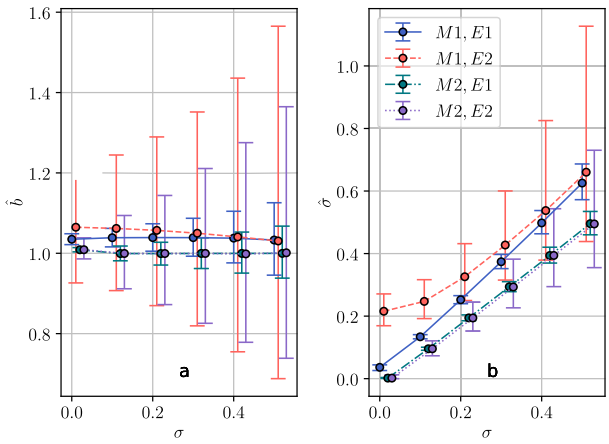

FIGURE 4. Mean values and 95% confidence intervals of the means of 10,000 iterations of simulated experiments for several levels of input decision noise for parameters A: b, B: σ. Data points are offset on the x-axis to make error bars more visibly distinguishable.

range of values expected for an individual experiment where participants behaved like the ideal observer. The means indicate the asymptotically "true" value of the parameters, which are extremely precise due to the large number of iterations. These statements also hold for the simulated effects and interaction terms, which are shown in Table 4.

The mapping method M2, which has a strictly linear colorfulness mapping (Fig. 3), unsurprisingly has an estimated exponent mean centered around 1. The mapping method M1, which does not have a linear colorfulness mapping, has estimated exponent means greater than 1, with a larger value for E2 than E1. For the noise parameter, the exactly linear mapping of M2 leads to estimators equal to the input decision noise. For M1, the estimated noise parameters are greater than the input decision noise. With the single-hue condition, E1 shows a smaller increase in estimated noise compared to the multi-hue condition E2. The difference here can be attributed to the effect of hue on the colorfulness mapping for M1 (see Fig. 3). This effect would be more pronounced in the multi-hue condition, which can explain the difference in E1 and E2 for M1. That is, the hue-colorfulness interaction causes deviations from a strict power-law scaling, which manifests as an increased noise estimator.

C. PARTICIPANTS

Participants were undergraduates fulfilling requirements for a first year psychology course. Participants gave informed and written consent in accordance with the experiment protocols approved by the Human Research Ethics Advisory Panel in the School of Psychology, UNSW Sydney. In E1, there were 55 participants, of which 47 were included for the final analysis. Due to social distancing restrictions for COVID-19, the collection for E2 was cut short. This resulted in 27 participants for E2, with 21 for final analysis. The gender and age statistics can be found in Table 1.

TABLE 1. Number of participants with age and gender statistics. "All" : total dataset. "Included" : dataset after exclusion criteria (see Sec. IV-A).

	Total		Female		Male		Med. Age	
	E1	E2	E1	E2	E1	E2	E1	E2
All	55	27	38	9	17	18	19	19
Included	47	21	34	7	13	14	19	18

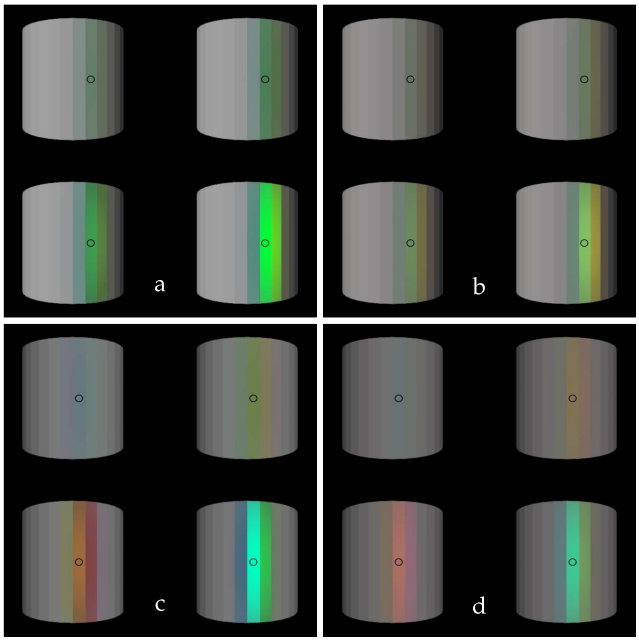

FIGURE 5. Example stimuli quadruplets for A: M1/E1, B: M2/E1, C: M1/E2, D:M2/E2. Circle indicates area of interest.

D. STIMULI

Individual stimuli consist of three computer rendered layers of a faceted cylinder mapped into each of the linear polarization channels. The normalized intensity coming from diffuse reflection of the object rendered with a rough surface was used for the intensity channel, while the degree of polarization channel was modeled by the normalized intensity from specular reflection of the object when rendered with a smooth material. The normal vector angles of the facets were used for overall pattern for the values of the angle of polarization, and the specific values in the area of interest for each stimulus is achieved by adding a scalar offset to the entire pattern. The center of the specular reflection layer (i.e. the brightest part of the layer) was used as the location of the area of interest. The area of interest was indicated by a black circle, and the participants were instructed to only consider the area within the circle when making their decision on each stimulus quadruplet.

Each cylinder has 32 facets, comprising 0.5 scene units in height and 0.5 scene units in diameter, and is positioned at the origin. The cylinder has a Phong surface model with an exponent parameter of 5. The 'diffuse' renders have the diffuse reflectance set to 0.5 (and zero specular reflectance)

and the 'specular' renders have the specular reflectance set to 0.05 (and zero diffuse reflectance). The scene is viewed from a perspective camera at 1 scene unit distance with a horizontal field of view of 76.6 degrees. The scene is lit by a directional light source, rendered at 5 degree increments between -45 and $+45$ degrees. Rendering was performed using Mitsuba (ver. 0.5.0) [43]. Only the center 5 facets were considered for the areas of interest, as the others would be too small due to the projection. The angle of the specular light source has to be set such that the area of interest falls in the center of one of the 5 facets, and the diffuse light source must be set such that the area of interest has the desired intensity. From the set of possible combinations of angles that match this criteria, one combination was chosen at random for each stimulus quadruplet. Each cylinder in a quadruplet share the same illumination layers, with the specular layer linearly scaled for each cylinder to match its DoLP stimulus level x.

During preliminary trials, the stimulus levels x corresponded directly to DoLP levels such that the full range of DoLP was used. Some participants had difficulty determining how to interpret difference-of-differences Δ when one of the cylinders appeared gray, having no colorfulness/saturation (DoLP $= 0$). That is, the difference between cylinders that were gray and cylinders that had a discernable color could be seen as categorical, whereas the difference between cylinders that both had discernable colors could be seen as quantitative. To eliminate the levels of DoLP that produce gray cylinders, the DoLP range was set as [0.2, 1.0] (20%-100%), which now correspond with stimulus levels x [0, 1].

While the choice to use a specular layer for the DoLP and the normal vectors for the AoP have some basis in the physics of Fresnel reflections, this method of mapping the rendered illumination channels to the polarization parameters directly does not employ a physics-based polarization rendering model. The resulting polarization properties are still physically possible and reasonable for an object in a polarization imaging scene. The choice to not use a physical model had two justifications: 1) ability to change polarization parameters independently and 2) processing time for rendering images. First, the experimental design requires the ability to freely change the degree of polarization without affecting the other polarization parameters. Using a physical model would make this much more difficult since changing the physical properties of the light source or object would affect all of the polarization channels simultaneously as well as the geometry of the object which could be an additional source of experimental bias or error. Second, polarization rendering methods are computationally expensive, and—with thousands of images needed to be rendered—impractical.

E. TASKS

The MLDS method involves participants answering a series of two-alternative forced-choice questions (2AFC). At each trial, participants are shown a quadruplet containing two pairs of stimuli, which have been colormapped according to one of the methods M1 or M2. Participants must make a judgement on which pair appears to have a greater difference in chromatic magnitude (colorfulness for M2 and saturation for M1). To indicate their choice, they key in the up or down arrow.

In order to test the orthogonality assumption, two conditions of AoP/hue similarity were tested. In the first experiment E1 (same-hue condition), each quadruplet stimulus was mapped using the same layers of intensity and AoP, with only DoLP changing within the quadruplet. The value of intensity at the area of interest was set at a constant level of 0.5. This does not mean that the brightness of this area is constant, since both color mapping methods scale the brightness with DoLP. The value of 0.5 was chosen since it produces the most linear colorfulness curve for M1, so that the comparison between the methods are made at the "best case scenario" for M1. The AoP was set based on the surface normals with a random offset constant added. In the second experiment E2 (different-hue condition), the offset constant for AoP was set at a different random value for each stimulus, with a minimum difference in AoP set to 30 degrees.

The two experiments E1 and E2 were conducted at different times and with different participants (one participant completed both experiments on separate occasions). For each experiment, the mapping methods were not mixed together, with participants performing a session on each mapping method fully before starting the other mapping method. The order of methods was counterbalanced such that half of the participants performed M1, then M2, and half with M2, then M1. In the results section, the group order G indicates participants who completed M1 first as G1, and M2 first as G2.

The quadruplets were chosen using a modified psi method [44], which attempts to maximize the amount of information gained during each trial. For each quadruplet, the observer's response would reveal information on the parameters (b, σ) with the amount of information varying by quadruplet. The adaptive psi method is a Bayesian technique for choosing the next stimulus that minimizes the expected information entropy in the posterior probability distribution based on the prior probability distribution. In our preliminary studies, this method would tend to repeat certain quadruplets many times within a trial. While the minimum was often stable at a certain quadruplet, other quadruplets had values of expected entropy very close to the minimum due to having a large number of stimulus quadruplets with similar level pairings. In order to avoid potential issues in regards to biases and to broaden range of quadruplets selected, the next stimulus quadruplet was taken as a random choice from the set of quadruplets with an expected entropy within 0.03% of the minimum value.

For each method, participants would respond to a set of 38 quadruplets before taking a 30 second break. After completing a total of 4 sets for the method, participants would take a self-paced break of at least 30 seconds. Then, the participants would repeat the instruction portion and set

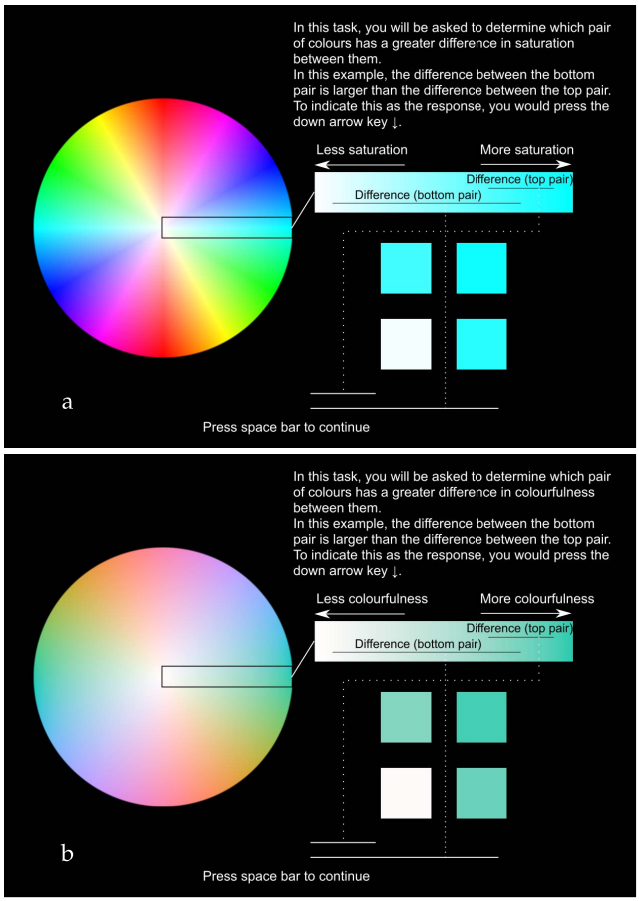

FIGURE 6. One slide of the participant instruction which describe the color terms of a) "saturation" for M1, and b) "colorfulness" for M2.

of stimuli for the other method. In total, the participants responded to 304 stimulus quadruplets (152 per method).

F. INSTRUCTIONS

Prior to conducting the session for each method, participants were briefly informed on the color terms relating to the chromatic magnitude channel (''saturation'' for M1, ''colorfulness'' for M2) using a color wheel typical in polarization visualization. The chromatic magnitude increased with the radial component, and the hue varied along the angular component. Participants were further informed on the other color terms ''hue'' and ''brightness'' and their orthogonality by showing how colors with different hues or different brightnesses can have the same chromatic magnitude (see Fig. 6 for one example). participants were then instructed on how to interpret the difference in chromatic magnitude by using visual aids marking the radial length of two colors (see Fig.6). The instructions continued to describe the task of comparing differences of two pairs of colors before showing a sample rendered stimulus quadruplet.

G. EQUIPMENT

Experiments were conducted in one of three similar dark rooms with a calibrated monitor (Display++ from Cambridge Research Systems Ltd). A custom linear RGB colorspace was developed for each monitor based off of the colorimetric measurements of the color primaries. Stimuli were rendered and saved as arrays of CIE XYZ tristimulus values, which would be converted to the RGB colorspace of each monitor during each experiment run. The following formulas (modified from Lindbloom [45]) were used to convert the XYZ values to and from custom RGB spaces.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = M \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad \begin{bmatrix} R \\ G \\ B \end{bmatrix} = M^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{10}$$

where M is a 3×3 display calibration matrix that depends on the colorimetric measurements of the display.

$$M = \begin{bmatrix} S_r X_r & S_g X_g & S_b X_b \\ S_r Y_r & S_g Y_g & S_b Y_b \\ S_r Z_r & S_g Z_g & S_b Z_b \end{bmatrix} \tag{11}$$

where X_c, Y_c, Z_c refer to luminance-corrected colorimetric measurements of the display at one of the three fully saturated monitor primary colors $c \in \{r, g, b\}$, and S constants are white point coefficients such that the white point $RGB = [1, 1, 1]$ corresponds to the measured monitor white point $[X_w, Y_w, Z_w]$.

$$\begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix}^{-1} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}. \tag{12}$$

The colorimetric measurements are normalized such that $Y_w = 100$ in order to set the relative luminance of the monitor to the same scale as the colormapping methods which assume a standard D65 white point. Computers running PsychoPy were used to supply stimuli to monitors and to record responses [46]. Figure 7 shows the benefits of using a custom RGB colorspace rather than the linear-sRGB space. Measurements from the monitor at multiple intensities agree with XYZ tristimulus values converted from the custom RGB colorspace significantly more than values converted from linear-sRGB.

IV. RESULTS

A. EXCLUSION CRITERIA

After conducting both methods, participants completed the Ishihara color test to determine whether their color vision may be deficient enough to warrant exclusion. Four participants were excluded from the analysis of E1 due to scoring too low on the Ishihara plate test. In addition, during each run twelve catch trials were included. These trials appeared exactly as the rest, so that the participant would not know they were any different from the rest of the trials. The quadruplet pair was chosen such that the difference in differences Δ_t would be quite large for any observer with reasonable parameters. Provided that participants understand the task and are answering honestly, the responses to these catch trials should be consistent. The responses to these trials

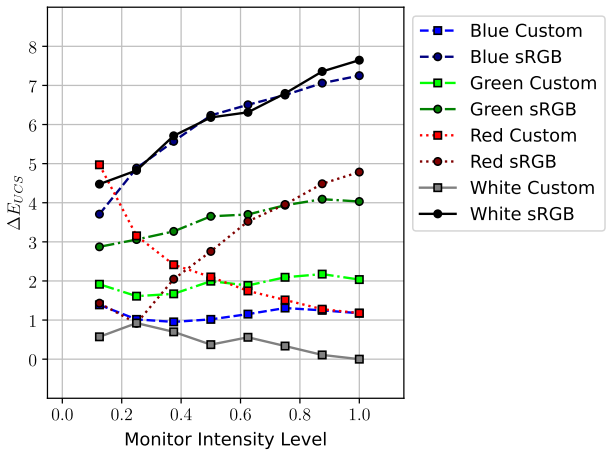

FIGURE 7. Comparison of linear sRGB and custom RGB for predicting tristimulus XYZ color measurements for one monitor. Color difference ΔE_{UCS} [7] between predicted and measured XYZ values is shown for different monitor intensity levels (RGB values).

were not included in parameter estimation. The choice of cutoff catch response rate was set at the fifth percentile of the distribution of all correct catch responses. With twenty-four catch trials between the two runs for each participant, the fifth percentile of correct catch responses was eighteen, which corresponds to a catch response rate of 75%. This cutoff rate applied to the number of catch trials per run, gave a cutoff rate for each run of 9/12. In order to maintain the accuracy of the data, participants were excluded from analysis if their number of correct catch trial responses was below nine for either run. Using this criteria, four participants were excluded from E1, and six from E2. Including the four participants testing too low on the Ishihara plates, a total of eight participants were excluded from E1.

B. MINIMIZATION

Parameters b, σ were estimated for each participant within each method using maximum likelihood estimation. That is, for each set of responses, the estimators were the minimum of the logarithm of likelihood function Eq. 9. The minimum of 9 has no closed-form expression, so numerical minimization must be used. The log-likelihood function has a single minimum that is smooth, so any number of numerical minimization methods could be used with virtually identical results. We used the Nelder-Mead method with a parameter and function tolerance of 10^{-4}. Supplying the starting parameters as the minimum value of the psychometric function evaluated at 400 points in parameter space, the minimization had rapid convergence at roughly 20 to 150 iterations.

C. SUMMARY

The means for each parameter can be seen in Table 2. Due to unequal sample sizes for the group order after excluding some participants' data, a de-weighted mean was used. A de-weighted mean is the mean of the means of each subgroup so that subgroups with larger sample sizes do not weight the

TABLE 2. De-weighted means and 95% CI for parameters exponent b and noise σ.

		M1		M2	
Param.		mean	95% CI	mean	95% CII
b	E1	1.16	[1.087, 1.23]	1.06	[0.97, 1.15]
	E2	0.91	[0.80, 1.030]	0.96	[0.87, 1.06]
σ	E1	0.20	[0.18, 0.21]	0.19	[0.17, 0.21]
	E2	0.23	[0.20, 0.27]	0.18	[0.16, 0.20]

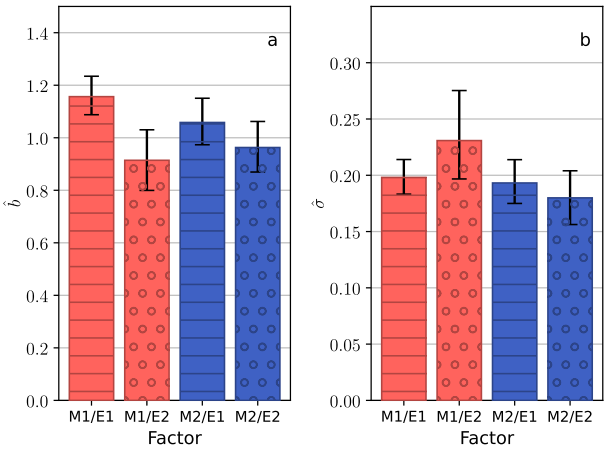

FIGURE 8. De-weighted mean values for both methods and hue conditions. Error bars depict the [2.5, 97.5] percentile range for means of 10,000 bootstrapped resampled data.

mean toward the average of the subgroup. The de-weighted means were calculated on log-transformed data and then exponentiated to get the results in Table 2. Additionally, to get confidence intervals for these means, the data were resampled using bootstrapping. For each combination of group (G1/G2), hue condition (E1/E2), and method (M1/M2), the data were resampled with replacement for the number of participants within that combination. Then, the de-weighted mean was calculated on each set of resampled data to get a distribution of the mean values. This distribution can be used to find a confidence interval for the mean value by taking the $[100\frac{\alpha}{2}, 100(1-\frac{\alpha}{2})]$ percentiles. The results in Table 2 uses $\alpha = 0.05$.

Unsurprisingly, the estimated exponent parameters are mostly distributed log-normally, so the hypothesis tests were performed on the log-transformed data. In contrast, the estimated noise parameters often did not have a detectable distribution. To make the least number of assumptions about the distributions of the data, the permutation test was used instead of Student's t-test for hypothesis testing. The permutation test determines how likely the difference between means in two samples is due solely to random variation. First, the contrast between two de-weighted sample means is calculated. Second, sample labels under test are permuted within the other labels which are held constant. For example, if the effect of experiment is under test, the labels E1 and E2 are shuffled within each group label and redistributed

TABLE 3. Main effects and interaction terms of log contrast and scaling from bootstrap sampling, and p-values from permutation tests.

			log contrast		scaling factor	
Factor	Param.	p	mean	95% CI	mean	95% CI
M1-M2	b	0.55	0.018	[-0.037, 0.076]	1.019	[0.96, 1.079]
	σ	**0.018**	0.14	[0.039, 0.24]	1.15	[1.040, 1.27]
E1-E2	b	**0.012**	0.17	[0.045, 0.29]	1.18	[1.046, 1.34]
	σ	0.59	-0.041	[-0.18, 0.10]	0.96	[0.83, 1.11]
G1-G2	b	**0.040**	-0.14	[-0.26, -0.012]	0.87	[0.77, 0.99]
	σ	0.36	-0.078	[-0.23, 0.066]	0.92	[0.80, 1.068]
M x E	b	**0.040**	0.14	[0.027, 0.25]	1.15	[1.028, 1.29]
	σ	**0.050**	-0.22	[-0.43, -0.030]	0.80	[0.65, 0.97]
M x G	b	0.23	-0.070	[-0.18, 0.046]	0.93	[0.83, 1.047]
	σ	0.075	-0.24	[-0.45, -0.05]	0.78	[0.64, 0.95]
G x E	b	0.90	-0.041	[-0.29, 0.20]	0.96	[0.74, 1.22]
	σ	0.079	0.22	[-0.069, 0.51]	1.25	[0.93, 1.68]
E1-E2 (M1)	b	**0.00082**	0.24	[0.098, 0.38]	1.27	[1.10, 1.46]
	σ	0.086	-0.15	[-0.35, 0.027]	0.86	[0.71, 1.028]
E1-E2 (M2)	b	0.21	0.094	[-0.036, 0.23]	1.099	[0.96, 1.26]
	σ	0.46	0.071	[-0.091, 0.24]	1.073	[0.91, 1.27]
M1-M2 (E1)	b	**0.029**	0.089	[0.016, 0.17]	1.093	[1.016, 1.18]
	σ	0.67	0.025	[-0.080, 0.13]	1.026	[0.92, 1.14]
M1-M2 (E2)	b	0.25	-0.052	[-0.14, 0.034]	0.95	[0.87, 1.035]
	σ	**0.0091**	0.25	[0.086, 0.430]	1.28	[1.090, 1.54]

TABLE 4. Comparison of scaling across factors for the experimental results in Table 3 against the simulated ideal observer distributions (input decision noise = 0.2).

Factor	Param.	Exp. Mean	Sim. Mean	Sim. 95% CI
M1-M2	b	**1.019**	1.05	[0.93, 1.18]
	σ	**1.15**	1.48	[1.23, 1.78]
E1-E2	b	**1.18**	0.99	[0.88, 1.12]
	σ	**0.96**	0.88	[0.73, 1.058]
M x E	b	**1.15**	0.98	[0.77, 1.25]
	σ	**0.80**	0.77	[0.53, 1.12]
E1-E2 (M1)	b	**1.27**	0.98	[0.80, 1.20]
	σ	**0.86**	0.77	[0.58, 1.015]
E1-E2 (M2)	b	**1.099**	1.000	[0.87, 1.15]
	σ	**1.073**	1.001	[0.79, 1.28]
M1-M2 (E1)	b	**1.093**	1.04	[1.00, 1.085]
	σ	**1.026**	1.30	[1.21, 1.39]
M1-M2 (E2)	b	**0.95**	1.06	[0.84, 1.34]
	σ	**1.28**	1.68	[1.18, 2.43]

key: **agrees with sim.** **more extreme** **less extreme**

to the participants in that group. Then, the contrast between means is calculated for the permuted data. This is repeated for some large number of resamplings, and the p-value is calculated as the ratio of samplings with a contrast more extreme than the observed contrast. For interaction terms, the contrast is calculated for the first term labels within one of the labels of the second term, and then for the other label of the second term. Then, the contrast between contrasts is calculated and compared to resampled contrast between contrasts to determine a p-value.

The p-value results in Table 3 were calculated using 300,000 random permutations. For the exponent parameter, the experiment factor is the most significant as a main effect, with a slightly less significant effect seen in the group order. The method factor is not significant for exponent alone, even though the mean values in Table 2 suggest a significant difference between the two methods. This can be explained by the interaction effect between method and experiment seen in Table 3. For the noise parameter, the method was a significant main effect, with a possible additional interaction effect from method and experiment. While the significance for this interaction term is marginal, the fact that such an interaction can be predicted from the color model simulations means we are less inclined to be strict on reducing a Type I (false positive) error here. The potential interaction effect between method and group shows some significance in the bootstrap confidence intervals for the noise parameter, but with less significance from the permutation test ($p = 0.075$). Without strong significance nor a model for this effect, there is not convincing evidence to reject the null hypothesis that there is no effect. The other interaction terms were not seen as strongly significant for either parameter.

Due to the possible interaction between method and experiment, the effect of experiment can be investigated for each method, and the effect of method for each experiment. Table 3 shows the results of the permutation test and the contrasts from bootstrap sampling for investigating the interaction effects between method and experiment. Most significantly, the effect of experiment was very significant for M1 but not significant for M2. Additionally, the effect of method was more significant in E1 for the exponent b.

V. DISCUSSION

The results of the parameter means and main effects somewhat agree with the simulated results of an ideal observer. Most of the effects are inside the 95% confidence intervals of the simulated results. In the factors where they disagree, there is a consistent trend: scaling values for the exponent b in the experimental results are more extreme (further from 1) than the simulated results, and less extreme (closer to 1) for the noise σ. In particular, the results show a strong effect of hue condition (E1 - E2) on the exponent for M1, which is not predicted in the simulated results. In contrast, there is no significant effect of hue condition in M2 for either simulated or experimental results for either parameter. Additionally, the predicted effect of an increase

in estimated noise parameter for the multi-hue condition E2 can be seen in the scaling of M1 relative to M2 (mean scale factor of 1.28). The simulated results also predict a small decrease in noise parameter for E1 relative to E2 for M1, but those samples in the experimental data only indicate a weakly significant difference between them. A connection could be made that since there is significant difference in the noise parameter between M1/E2 and M2/E2, yet no significance between M2/E1 and M2/E2 nor between M1/E1 and M2/E1, that the weak significance between M1/E1 and M1/E2 could indicate a real, but small, effect. This would be consistent with the distinctions seen in the simulated data, although the effect sizes are smaller in the experimental data.

The difference in sample size between E1 and E2 is accounted for in the permutation test discussed in Sec. IV-C. However, the difference in sample size does affect the power of the test for agreement between experimental and simulated results. Due to the larger confidence intervals for E2 from the smaller sample size, it is more likely to find a significant difference between experimental and simulated results for the measure of contrast between methods in E1 than for E2. Therefore, it is not surprising that the experimental results agreed with the simulated results even though there was a significant difference for E1.

The group order factor was not included in the ideal observer modeling, as the mechanism for such an effect is not known to us. From the participant feedback, there was some confusion regarding the distinction between the chromatic magnitude terms ''saturation'' and ''colorfulness'' that were used in the instruction portion of the experiment for M1 and M2, respectively. There may be some connection between using different terms and the group factor, but this is just a conjecture. In reality there may not be any consistent effect at all, or the effect may have completely different causes. Significantly larger sample sizes would be necessary to conclude that this is a consistent effect.

Regardless of the whether the experimental data match the simulations of ideal observers, the results display a consistent trend of M1 exhibiting more nonlinearity and less hue-colorfulness orthogonality that agree with the previous analysis by Kruse *et al.* [8] Conversely, M2 did not show any sign that hue-colorfulness were non-orthogonal, and a linear exponent ($b = 1$) was well within its 95% confidence intervals of the means. Additionally, M1 displayed a greater amount of decision noise compared to M2. In no measure was there any advantage of M1 over M2 found in this study. Furthermore, since the intensity level was set such that the mapping of colorfulness C for M1 was on average close to linear, these results should be considered a conservative estimate for the differences between the two methods with respect to the full range of intensity levels. The effect of parametrising the scales levels ψ as an exponential function could be obscuring some fine-grain nonuniformities of the visual encoding in the HSV method that would be expected from the nonuniformities of the saturation channel of HSV as seen in Fig. 3. Thus, the parametric model could be considered as slightly biasing the results in favor of making the HSV method appear more uniform. Again, this indicates that the estimation of the advantages of the UCS method over the HSV method is on the conservative end.

Although the trivariate colormapping methods are specifically made for visualizing polarimetric image data, the results of this study have further implications for the effectiveness of multivariate colormapping methods in general. The results here support long-held guidelines advocating for using uniform color spaces for the visual encoding of data into color. First, the results support the claim that using a uniform color space is more effective than a non-perceptual color space for supporting the user task of quantitative comparison. Second, the results also support the claim that quantitative comparative judgments can be made using the chromatic magnitude channel for visual encoding independent of the hue, provided the chromatic magnitude channel corresponds to a perceptually uniform color space coordinate. Thus, a multivariate colormapping system can be used to encode one variable using the chromatic magnitude channel and another using the hue channel without the hue-encoded variable affecting the ability for users to perform quantitative comparisons on the chromatic magnitude-encoded variable.

VI. FURTHER WORK

This study only investigated users' ability to make quantitative comparative judgments using the chromatic magnitude channel as an visual encoding channel in a multivariate colormapping system independent of the hue channel. However, it cannot be assumed that the converse would be true, that users would be able to perform identification and look-up tasks on a hue-encoded variable independent of the chromatic magnitude-encoded variable. Certainly these tasks would be impossible when the chromatic magnitude is zero, so the hue channel could not be completely independent. However, at sufficient levels of chromatic magnitude the ability to perform tasks on hue-encoded variables could potentially be independent of the chromatic magnitude.

It should also be noted that the effects of size and color constancy were not measured in this study. Since the regions of interest of the stimuli (see Fig. 5) were quite large and surrounded by colors of the same hue, these two effects likely had very little impact on the results of the study. For images created from real polarimetric measurements, these effects could have a significant impact on the ability to perform comparative judgments on the DoLP. Further study of these effects for this type of visualization is warranted.

ACKNOWLEDGMENT

This research was a part of Andrew W. Kruse's Ph.D. thesis. As such, portions of this article were included in one of the chapters of this thesis [47].

REFERENCES

[1] G. D. Bernard and R. Wehner, ''Functional similarities between polarization vision and color vision,'' *Vis. Res.*, vol. 17, no. 9, pp. 1019–1028, Jan. 1977.

[2] R. Walraven, "Polarization imagery," *Proc. SPIE*, vol. 112, pp. 164–167, Oct. 1977.
[3] J. E. Solomon, "Polarization imaging," *Appl. Opt.*, vol. 20, no. 9, pp. 1537–1544, 1981.
[4] L. B. Wolff and T. A. Mancini, "Liquid crystal polarization camera," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Apr. 1992, pp. 120–127.
[5] J. S. Tyo, B. M. Ratliff, and A. S. Alenin, "Adapting the HSV polarization-color mapping for regions with low irradiance and high polarization," *Opt. Lett.*, vol. 41, no. 20, pp. 4759–4762, 2016.
[6] L. Neumann, R. Hegedüs, G. Horváth, and R. García, "Applications of high precision imaging polarimetry," in *Proc. 4th Eurographics Conf. Comput. Aesthetics Graph., Vis. Imag.*, Jun. 2008, pp. 89–97.
[7] M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," *Color Res. Appl.*, vol. 31, no. 4, pp. 320–330, Jul. 2006.
[8] A. W. Kruse, A. S. Alenin, I. J. Vaughn, and J. S. Tyo, "Perceptually uniform color space for visualizing trivariate linear polarization imaging data," *Opt. Lett.*, vol. 43, no. 11, pp. 2426–2429, 2018.
[9] F. Zhao, Y. Dong, and J.-L. Zhang, "Polarization visualization for low-irradiance regions by perceptually uniform color space," *Defence Technol.*, vol. 17, no. 2, pp. 505–511, Apr. 2021.
[10] C. Ware, "Color sequences for univariate maps: Theory, experiments and principles," *IEEE Comput. Graph. Appl.*, vol. CGA-8, no. 5, pp. 41–49, Sep. 1988.
[11] C. A. Brewer, "Color use guidelines for data representation," in *Proc. Amer. Statisc. Assoc. Sect. Statist. Graph.*, Alexandria, VA, USA, 1999, pp. 55–60.
[12] P. L. Rheingans, "Task-based color scale design," *Proc. SPIE*, vol. 3905, pp. 35–43, May 2000.
[13] T. Munzner, "Map color and other channels," in *Visualization Analysis and Design* (AK Peters Visualization Series), A. K. Peters, Ed. Boca Raton, FL, USA: CRC Press, 2014, pp. 218–241, doi: 10.1201/b17511-11.
[14] N. Smith and S. V. D. Walt, "A better default colormap for matplotlib," in *Proc. SciPy*, 2015, pp. 6–12.
[15] L. Zhou and C. D. Hansen, "A survey of colormaps in visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 8, pp. 2051–2069, Aug. 2016.
[16] J. Bernard, M. Steiger, S. Mittelstädt, S. Thum, D. Keim, and J. Kohlhammer, "A survey and task-based quality assessment of static 2D colormaps," *Proc. SPIE*, vol. 9397, Feb. 2015, Art. no. 93970M.
[17] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman, and H. Pfister, "Evaluation of artery visualizations for heart disease diagnosis," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2479–2488, Dec. 2011.
[18] R. M. Karim, O.-H. Kwon, C. Park, and K. Lee, "A study of colormaps in network visualization," *Appl. Sci.*, vol. 9, no. 20, p. 4228, Oct. 2019.
[19] A. W. Kruse, A. S. Alenin, and J. S. Tyo, "Review of visualization methods for passive polarization imaging," *Opt. Eng.*, vol. 58, no. 8, Apr. 2019, Art. no. 082414.
[20] P. K. Robertson and J. F. O'Callaghan, "The application of perceptual color spaces to the display of remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-26, no. 1, pp. 49–59, Jan. 1988.
[21] C. Ware, "Color," in *Information Visualization*. Boston, MA, USA: Morgan Kaufmann, 2013, pp. 95–138.
[22] R. Bujack, T. L. Turton, F. Samsel, C. Ware, D. H. Rogers, and J. Ahrens, "The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 923–933, Jan. 2018.
[23] P. Nardini, M. Chen, F. Samsel, R. Bujack, M. Bottinger, and G. Scheuermann, "The making of continuous colormaps," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 6, pp. 3048–3063, Jun. 2021.
[24] C. Ware, T. L. Turton, R. Bujack, F. Samsel, P. Shrivastava, and D. H. Rogers, "Measuring and modeling the feature detection threshold functions of colormaps," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 9, pp. 2777–2790, Sep. 2019.
[25] K. Moreland, "Diverging color maps for scientific visualization," in *Advances in Visual Computing*. Berlin, Germany: Springer, 2009, pp. 92–103.
[26] H. Wainer and C. M. Francolini, "An empirical inquiry concerning human understanding of two-variable color maps," *Amer. Statistician*, vol. 34, no. 2, pp. 81–93, May 1980.
[27] B. E. Trumbo, "A theory for coloring bivariate statistical maps," *Amer. Statistician*, vol. 35, no. 4, pp. 220–226, Nov. 1981.
[28] P. Robertson and J. O'Callaghan, "The generation of color sequences for univariate and bivariate mapping," *IEEE Comput. Graph. Appl.*, vol. CGA-6, no. 2, pp. 24–32, Feb. 1986.
[29] K. L. Coulson, *Polarization and Intensity of Light in the Atmosphere*. Mumbai, India: A Deepak Pub, 1988.
[30] L. Woltjer, "The polarization and intensity distribution in the crab nebula derived from plates taken with the 200-inch telescope by Dr. W. Baade," *Bull. Astron. Inst. Netherlands*, vol. 13, no. 478, pp. 301–311, 1957.
[31] T. J. Rogne, F. G. Smith, and J. E. Rice, "Passive target detection using polarized components of infrared signatures," *Proc. SPIE*, vol. 1317, pp. 242–251, Oct. 1990.
[32] A. Pierangelo, A. Nazac, A. Benali, P. Validire, H. Cohen, T. Novikova, B. H. Ibrahim, S. Manhas, C. Fallet, M.-R. Antonelli, and A.-D. Martino, "Polarimetric imaging of uterine cervix: A case study," *Opt. Exp.*, vol. 21, no. 12, pp. 14120–14130, Jun. 2013.
[33] M. V. Berry, M. R. Dennis, and J. R. L. Lee, "Polarization singularities in the clear sky," *New J. Phys.*, vol. 6, no. 1, p. 162, 2004.
[34] G. M. Calabrese, P. C. Brady, V. Gruev, and M. E. Cummings, "Polarization signaling in swordtails alters female mate preference," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 37, pp. 13397–13402, Sep. 2014.
[35] F. Snik, J. Craven-Jones, M. Escuti, S. Fineschi, D. Harrington, A. D. Martino, D. Mawet, J. Riedi, and J. S. Tyo, "An overview of polarimetric sensing techniques and technology with applications to different research fields," *Proc. SPIE*, vol. 9099, May 2014, Art. no. 90990B.
[36] J. S. Tyo, D. L. Goldstein, D. B. Chenault, and J. A. Shaw, "Review of passive imaging polarimetry for remote sensing applications," *Appl. Opt.*, vol. 45, no. 22, pp. 5453–5469, 2006.
[37] P.-J. Lapray, L. Gendre, A. Foulonneau, and L. Bigué, "Database of polarimetric and multispectral images in the visible and NIR regions," *Proc. SPIE*, vol. 10677, May 2018, Art. no. 1067738.
[38] M. K. Kupinski, C. L. Bradley, D. J. Diner, F. Xu, and R. A. Chipman, "Angle of linear polarization images of outdoor scenes," *Opt. Eng.*, vol. 58, no. 8, Jun. 2019, Art. no. 082419.
[39] Y. L. Gagnon and N. J. Marshall, "Intuitive representation of photopolarimetric data using the polarization ellipse," *J. Experim. Biol.*, vol. 219, no. 16, pp. 2430–2434, Jan. 2016.
[40] L. B. Wolff, "Polarization camera for computer vision with a beam splitter," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 11, no. 11, pp. 2935–2945, 1994.
[41] L. T. Maloney and J. N. Yang, "Maximum likelihood difference scaling," *J. Vis.*, vol. 3, no. 8, p. 5, Oct. 2003, doi: 10.1167/3.8.5.
[42] K. Knoblauch and L. T. Maloney, "MLDS: Maximum likelihood difference scaling in R," *J. Stat. Softw.*, vol. 25, no. 2, pp. 1–26, 2008.
[43] W. Jakob. (2010). *Mitsuba Renderer*. [Online]. Available: http://www.mitsuba-renderer.org
[44] L. L. Kontsevich and C. W. Tyler, "Bayesian adaptive estimation of psychometric slope and threshold," *Vis. Res.*, vol. 39, no. 16, pp. 2729–2737, 1999.
[45] B. Lindbloom. (Apr. 7, 2017). *RGB/XYZ Matrices*. [Online]. Available: http://www.brucelindbloom.com/
[46] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv, "PsychoPy2: Experiments in behavior made easy," *Behav. Res. Methods*, vol. 51, no. 1, pp. 195–203, Feb. 2019.
[47] A. W. Kruse, "Visualisation methods for polarimetric imaging," Ph.D. dissertation, School Eng. Inf. Technol., Univ. New South Wales Canberra, Campbell, ACT, Australia, 2021.

ANDREW W. KRUSE received the B.S. degree in optical engineering from The Institute of Optics, University of Rochester, in 2016, and the Ph.D. degree in electrical engineering from the School of Engineering and IT, University of New South Wales Canberra, in 2021. His main research interests include developing and analyzing visualization methods for polarimetric imaging data. He has published two other articles and several conference papers on this topic, and maintains a public repository for polarimetric imaging visualization software on GitHub: https://github.com/andrewwkruse

DAMIEN J. MANNION received the Bachelor of Psychology degree from The University of Newcastle, Australia, in 2005, and the Ph.D. degree in psychology from The University of Sydney, Australia, in 2010. He is currently a Senior Lecturer with the School of Psychology, University of New South Wales (UNSW) Sydney, which he joined in 2013 after holding postdoctoral positions at the Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA, from 2010 to 2011, and the University of Minnesota, from 2011 to 2013. His research interests include visual and auditory perception and their neural foundations, and he is an author of over 20 articles in these areas.

ANDREY S. ALENIN graduated from the College of Optical Sciences, The University of Arizona. His Ph.D. thesis titled "Matrix structure for information-driven polarimeter design" has provided the necessary formalism to enable design enhancement through a better understanding of the underlying measurement space. Since moving to the University of New South Wales (UNSW) Canberra, in 2015, he has applied his tools to various polarimeter architectures as well as branched out to other modalities, where carefully designed machine learning and/or statistical analysis are most necessary. He is currently a Senior Lecturer with the School of Engineering and IT, UNSW Canberra.

J. SCOTT TYO (Fellow, IEEE) has been a Professor of electrical engineering with the School of Engineering and IT, UNSW Canberra, since 2015. Before joining UNSW, he was with the College of Optical Sciences, The University of Arizona, and the ECE Department, The University of New Mexico. Prior to his career in academic, he served in the U.S. Air Force as a Research and Development Officer. He has been involved in important work in understanding the measurement, analysis, and interpretation of polarization imagery for almost three decades. His research interests include a broad range of applied optics and electromagnetics, especially in regard to applications around remote sensing and imaging.

He is a fellow of SPIE and the OSA. He was previously a member of the Administrative Committee of the IEEE Antennas and Propagation Society, and he is currently a member of the Board of Directors of SPIE. He has won the 2014 SPIE GG Stokes Award for his work in polarization imaging.