Stimulus-specific variability in color working memory with delayed estimation

Gi-Yeul Bae Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA

Maria Olkkonen Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

Sarah R. Allred Department of Psychology, Rutgers—The State University of New Jersey, Piscataway, NJ, USA

Colin Wilson Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA

Jonathan I. Flombaum Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA

?1 **Working memory for color has been the central focus in an ongoing debate concerning the structure and limits of visual working memory. Within this area, the delayed estimation task has played a key role. An implicit assumption in color working memory research generally, and delayed estimation in particular, is that the fidelity of memory does not depend on color value (and, relatedly, that experimental colors have been sampled**
?2 **homogeneously with respect to discriminability). This assumption is reflected in the common practice of collapsing across trials with different target colors when estimating memory precision and other model parameters. Here we investigated whether or not this assumption is secure. To do so, we conducted delayed estimation experiments following standard practice with**
?3 **a memory load of one. We discovered that different target colors evoked response distributions that differed widely in dispersion and that these stimulus-specific response properties were correlated across observers. Subsequent experiments demonstrated that stimulus-specific responses persist under higher memory loads and that at least part of the specificity arises in perception and is eventually propagated to working memory. Posthoc stimulus measurement revealed that rendered stimuli differed from nominal stimuli in both chromaticity and luminance. We discuss the implications of these deviations for both our results and those from other working memory studies.**

Introduction

Color working memory has been the central focus in an ongoing and vigorous debate concerning the structure and limits of visual working memory. Initially, these limits were hypothesized to be discrete in nature, restricting the individual number of objects a person could store at once (Cowan, 2001; Luck & Vogel, 1997). More recently, the focus of much research has shifted to the quality of visual working memory—the precision with which an observer can store and report the specific value of an object feature. This shift in focus emerged to a large extent with the introduction of the delayed estimation paradigm (Wilken & Ma, 2004; Zhang & Luck, 2008). In this paradigm, an observer attempts to store the features of some number of objects and is then asked to identify the feature value of a probed item on a continuous scale. Because of the intuitive nature of continuous differences between colors, color working memory has enjoyed the lion's share of research with this paradigm (Figure 1; e.g., Anderson & Awh, 2012; Bays, Catalao, & Husain, 2009; Bays, Wu, & Husain, 2011; Emrich & Ferber, 2012; Fougnie & Alvarez, 2011; Fougnie, Asplund, & Marois, 2010; Fougnie, Suchow, & Alvarez, 2012; Gold et al., 2010; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004; Zhang & Luck, 2008, 2009,

Citation: Bae, G.-Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision, XX*(XX):X, 1–23, http://www.journalofvision.org/content/XX/XX/XX, doi:10.1167/XX.XX.XX.

doi: 10.1167/XX.XX.XX Received June 6, 2013; published Month 0, 2014 ISSN 1534-7362

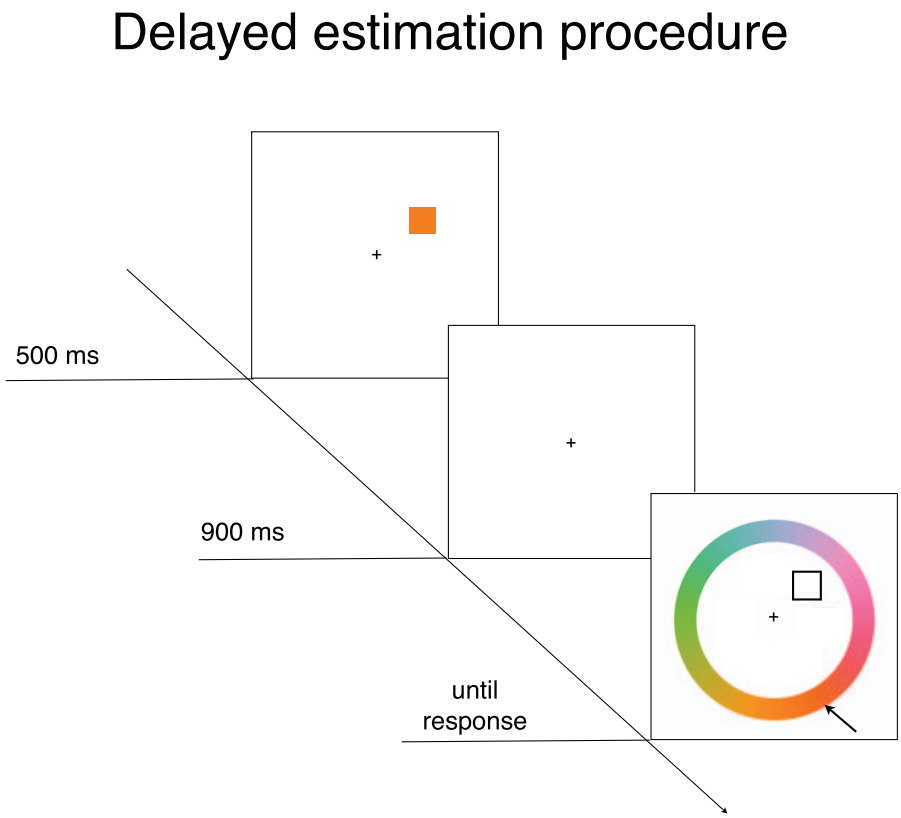

Figure 1. Schematic display of the delayed estimation procedure (with a memory load of one). Note that the display background, which is shown as white here, was a neutral gray in the experiments reported below, and its color has varied in the published literature.

?4 2011). While the debate concerning the limits of color working memory continues, there appears to be wide-ranging consensus that the working memory representation of color is noisy or probabilistic—that is, varying in fidelity—and that some of this variability is imposed by the structure of visual working memory.

An implicit assumption underlying this research has been that perceptual and memory fidelity are independent of particular stimulus values. This assumption is most clear in the method by which delayed estimation responses have been modeled. Typically, a mixture model is used to characterize responses as arising either from a noisy representation of a probed item or from unbiased guessing (possibly with a third component reflecting misbinding or misremembering of color position, e.g., Bays et al., 2009; alternative models remove the guessing component altogether and sometimes include other sources of variability such as motor imprecision, e.g., van den Berg et al., 2012). Because a single set of model parameters is estimated from all responses for a given memory load, the implication is that all color values are represented with the same fidelity on average (e.g., Bays et al., 2009; van den Berg et al., 2012; Zhang & Luck, 2008).

The assumption of value-independent fidelity is predicated on a more practical assumption: namely, that any value-*dependent* effects have been controlled for by sampling stimuli from a perceptually homogeneous set of colors. In fact, this assumption motivates the use of color and delayed estimation in many studies because, as Zhang and Luck (2011) put it, in these experiments "precision can be unambiguously operationalized" (p. 1434). In particular, researchers have most often used the CIELAB color space to sample values around a center point (defined by L*, a*, and b* values) with a fixed radius. This choice is presumably ?5 motivated by the assumption that the CIELAB space is perceptually uniform, with equal physical distances in the space corresponding to equal perceptual distances. However, there are at least three general reasons to doubt the validity of this assumption.

First, the nature of color perception could introduce stimulus-specific effects. Although CIELAB—the most commonly used space in studies of delayed estimation—was developed to be a perceptually uniform color

space, this is known to be only an approximation (e.g., Brainard, 2003; Fairchild, 1998; Wyszecki & Stiles, 1982). Furthermore, there are substantial individual differences in color perception between observers judged to be color normal by standard assessment techniques (Webster, Miyahara, Malkoc, & Raker, 2000a,b). Because two colors judged as equally discriminable by one color normal observer may not be judged so by another observer, no physically defined color space can be perceptually uniform across all observers. It is for this reason that studies seeking to link discrimination and color appearance measure discrimination thresholds directly rather than inferring them from standard color spaces (e.g., Bachy, Dias, Alleysson, & Bonnardel, 2012; Danilova & Mollon, 2012; Witzel & Gegenfurtner, 2013). Importantly, these studies have identified different just noticeable differences for different colors.

Second, the technical difficulty of rendering colors accurately could introduce stimulus-specific effects. This is because stimuli on emissive displays are specified in units that are device specific (e.g., RGB values) rather than in physical units (e.g., energy per wavelength). Because there is considerable variation in the hardware and software that govern stimulus display, two monitors requesting identical RGB values will likely produce different color signals and thus different CIELAB coordinates. Fortunately, there are standard calibration techniques that enable the reliable production of stimuli specified in a variety of color spaces. However, such calibration techniques have not been widely employed in the working memory literature. Some working memory studies report engaging in aspects of display calibration such as gamma correction (Zhang & Luck, 2008), but we were unable to find a single working memory paper that described utilizing the full calibration procedure that is standard in the literature on color perception (Allen, Beilock, & Shevell, 2012; Olkkonen & Allred, 2014; Witzel & Gegenfurtner, 2013; Xiao, Hurst, MacIntyre, & Brainard, 2012).

Third, memory itself could introduce stimulus-specific variability in precision. There is a literature on color memory that is largely distinct from the working memory literature. In this literature, it has been reported that memory for colors is shifted in systematic yet complex ways relative to presented colors (e.g., Burnham & Clark, 1954, 1955; Collins, 1931; Jin & Shevell, 1996; Ling & Hurlbert, 2008; Nemes, Parry, Whitaker, & McKeefry, 2012; Nilsson & Nelson, 1981; Prinzmetal, Amiri, Allen, & Edwards, 1998).

For these reasons, we investigated whether memory fidelity for color is independent of stimulus value in delayed estimation. To do so, we employed two typical versions of the delayed estimation task and two commonly used color spaces for stimulus sampling, using procedures standard in the working memory literature (van den Berg et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008). In Experiment 1, contrary to the implicit assumption in color working memory research, we found that variability in memory was systematically dependent on color. In Experiments 2 through 4, we investigated the extent to which this dependence on stimulus generalized to perception and other memory loads. Posthoc measurements of color stimuli revealed significant deviations between intended and displayed stimuli. We explore the causes and likely effects of these deviations and discuss the reasons why such deviations are likely endemic in delayed estimation studies using color as the stimulus of interest.

Experiment 1: Color-specific variability with a memory load of one

The goal of this experiment was to determine empirically whether response variability in a standard delayed estimation task is stimulus dependent. Specifically, we sought to estimate response variability within and across observers for each of 180 colors in two of the sets of color samples typically utilized in delayed estimation experiments. We did so in a straightforward fashion: Using a minimal memory load of one, we collected measurements for all target colors from each observer. One group of observers each performed the experiment with colors nominally defined in CIELAB space and HSV space.

Method

Observers

In exchange for course credit, three Johns Hopkins University undergraduates participated in the experiment with stimuli constructed with the CIELAB color space. A different group of three observers participated with stimuli constructed with the HSV color space. All observers had normal or corrected-to-normal visual acuity and reported no known deficits in color perception. The Johns Hopkins University IRB approved the protocol for this experiment.

Apparatus

The experiment took place in a dark, sound-attenuated room. There was no light source except for a computer monitor. All stimuli were presented on a CRT monitor at a viewing distance of 60 cm such that

the display subtended approximately 28.64° by 19.09° of visual angle.

Stimuli and procedure

For the CIELAB stimuli we followed the methods of Zhang and Luck (2008) and others. We specified a nominal set of 180 evenly spaced colors from CIELAB color space centered on L* = 70, a* = 20, and b* = 38 and with a radius of 60 (Anderson & Awh, 2012; Gold et al., 2009; Zhang & Luck, 2009, 2011). These CIELAB coordinates were then converted into 180 RGB values via a color conversion algorithm (Image Processing Toolbox, Matlab). To do so, we used the ICC standard white point, with an xyY value of 0.35, 0.36, 1 (i.e., D50; Y value is normalized). For the HSV wheel, 180 equally spaced hues were sampled, with the intensities of saturation and value fixed at 80%.

Each trial began with a white fixation cross (0.5° × 0.5°) displayed in the center of a gray screen. After 500 ms, a colored square (2° × 2°) appeared at one of eight possible positions (4.5° from fixation). The square's color was one of the 180 stimuli. The square remained present for 500 ms, and observers were instructed to commit the color of the square to memory. After a 900 ms blank delay, a test display appeared with a black frame occupying the location of the square. A color wheel was drawn (8.2° radius and 2° thick) surrounding the space in which memory objects could appear. The color wheel consisted of all 180 stimuli, arranged so that each stimulus occupied 2°. The wheel was randomly rotated on each trial to avoid spatial encoding of the stimulus value. Observers were asked to click the target color on the color wheel as precisely as possible. After the observer's response, a black line superimposed on the wheel indicated the clicked position. Each observer completed 5 blocks of 360 trials, totaling 1,800 trials. Within a block, each of the 180 color values appeared twice in an order that was randomized by observer, producing 10 total measurements per color for each observer.

Stimulus measurement

Working memory studies generally use the same procedures described above: A stimulus set is specified in CIELAB, and device-specific RGB values are determined using an industry standard white point. However, because the hardware and software of laboratory displays vary, there may be substantial differences between intended and presented colors. Color perception research standardly employs monitor calibration to ensure that device-specific commands render stimuli that match the physical characteristics of the requested stimuli (for a discussion of display calibration, see Brainard, Pelli, & Robson, 2002). We did not perform monitor calibration in our study, nor are such calibration practices common in the working memory literature.

To assess the effects of incomplete calibration and color rendering, we made posthoc measurements of our stimuli using a PR-655 spectroradiometer (PR-655 SpectraScan, Photo Research Inc., Chatsworth, CA). The measurements were converted to CIE XYZ and CIELAB spaces using colorimetry routines in Psychophysics Toolbox (Brainard, 1997). The measured xyY values of our 180 stimuli (hereafter referred to as *rendered* stimuli) are reported in Appendix A, as are the intended (hereafter *nominal*) CIELAB values. In addition, we report the CIELAB values of the 180 stimuli when the background of the monitor during the experiment is used as the white point in the conversion between xyY coordinates and CIELAB coordinates. We do so because it is standard practice in color perception studies to take the color of the monitor background to be the white point in color conversions (e.g., Brainard, 1998; Giesel & Gegenfurtner, 2010; Wyscecki & Stiles, 1982).

As can be seen in Figure 2, there were large deviations between nominal and rendered stimuli. Several features are of note. First, rendered stimuli were not circularly arranged in terms of their a* and b* coordinates in CIELAB space (Figure 2a). This can occur for several reasons, but one common reason is that a particular device cannot physically display a requested value. In Figure 2c, for example, we plot the gamut of the monitor on the xy plane of the CIE xyY space as determined by posthoc measurements with the radiometer. Any point outside the solid triangle cannot be produced by the monitor; clearly, some of the nominal colors fell outside this gamut. These colors were automatically mapped to others that do lie within the gamut, resulting in a mismatch between nominal and rendered colors and luminances. The mismatch between requested and actual luminance is evident in Figure 2b, which shows the L* values of rendered colors. The sum of these effects can be seen by plotting rendered colors in xyY space (Figure 2d). The rendered values varied systematically in both luminance (Y value) and chromaticity (x and y values).

Because the rendered values, measured posthoc, deviated so substantially from nominal values and because standard analysis techniques depend heavily on the assumption that stimuli vary only in hue (i.e., that they are on a circle of constant L* in CIELAB space), we were faced with a difficult decision about how to report and interpret our data. On the one hand, our stimuli clearly did not differ from each other only in hue (and thus, angular distance). However, the entire analysis process as detailed below, and in the color working memory literature generally, rests on the

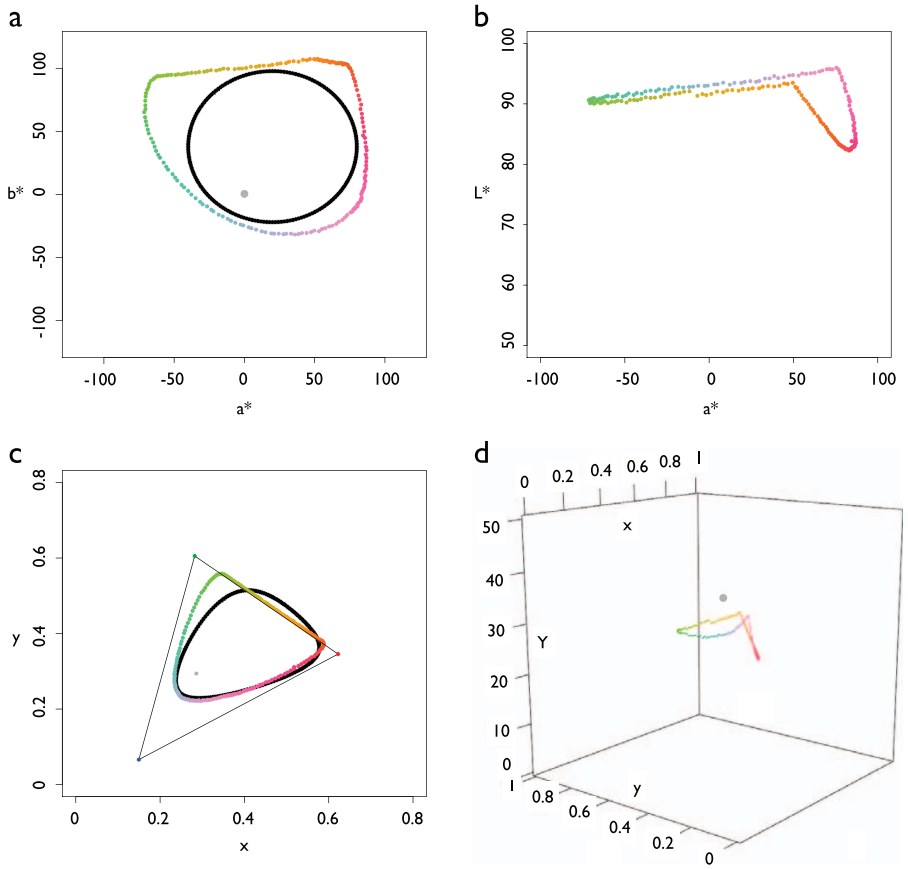

Figure 2. Comparison between intended and measured color values. (a) Distortions in a* and b* dimensions. Black circle represents a* and b* coordinates for intended color values. Colored outline represents measured color values. Gray dot in the middle represents the background color. Note that rendered colors were not circularly arranged and were not equally distant from the background. (b) Distortions in L* value. Plotted here is a* versus L* for measured (colored shape) colors. Rendered colors had different L* values, violating the isoluminance assumption. (c) Measured xy values (colored shape) and nominal xy values (black shape). Nominal xy values were calculated from nominal CIELAB values using the measured background xyY of the monitor as the white point in the conversion. The triangle in the figure represents the gamut of the monitor used in the study. Colors outside the triangle cannot be rendered correctly on the monitor, and some of the nominal colors lay outside this gamut. (d) Measured color values in xyY dimensions. Again, rendered colors were not isoluminant. Note: Although rendered colors deviated from nominal colors, the posthoc radiometer measurements demonstrate that the xy coordinates of all rendered stimuli were measurably different from each other.

assumption that stimuli can be described fully in terms of their angular distance from one other. Because of this, we proceed by analyzing data as though we had rendered stimuli correctly. We then carefully note the likely implications of rendering inaccuracies, which may be present in color working memory research quite generally, for the interpretation of our results.

Despite this "worst case" scenario of large deviations between rendered and nominal stimuli, we believe our data to be useful for at least two reasons. Most importantly, for reasons we outline later, we believe that the conclusions we draw survive these deviations and make an important contribution to the literature. In addition, since we followed standard stimulus practices in the delayed estimation literature, it is likely that many other working memory studies fall prey to similar inaccuracies. Indeed, the authors of several influential reports were generous enough to discuss their methods with us in detail, confirming the absence of monitor calibration, the absence of nominal and rendered stimulus comparison, and the use of a default white point instead of a measured white point. Moreover, since we used the nominal CIELAB samples employed by several related reports, and because many of the stimuli fell outside our CRT monitor gamut, it is likely that at least some stimuli were out of gamut in those studies as well.

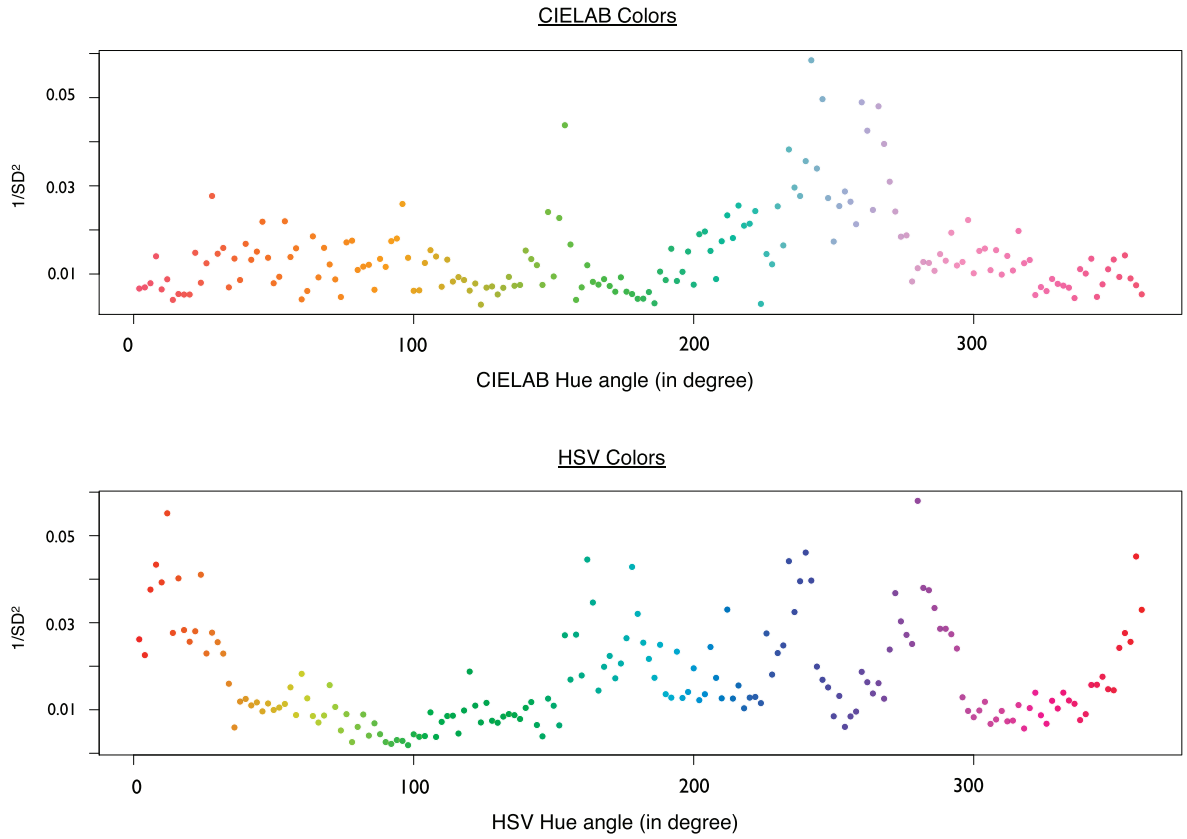

Figure 3. Reciprocal of circular SD^2 of target-response differences for each color value in CIELAB (top panel) and HSV (bottom panel) color spaces. Note that the x-axis of the CIELAB plot is hue angle centered on the intended color wheel (i.e., centered on a* b* = 20, 38, not on the origin of CIELAB).

Analysis

The raw data generated in these experiments were distributions of response positions elicited by each target stimulus. We characterized the spread of these response distributions in terms of angular precision (inverse variance), with interest in whether some stimuli elicited more concentrated response distributions than others. In keeping with the literature in this area, we also quantified response precision with a model-based measure. With only a single memory item, these two methods are very similar (see Figures 3 and 4b). But the working memory literature is typically interested in memory for several items (see Experiment 3), and the analysis of multiple-item data is thought to require a model in which representational precision combines with other sources of variability (e.g., lapse trials).

To characterize these results in terms typical of previous studies, we fit a mixture model comprising a von Mises distribution and a uniform distribution (Zhang & Luck, 2008). The von Mises distribution is a circular analog of the standard normal distribution. The model that we fit included two free parameters: the proportion of target-based (as opposed to lapse) responses (Pm; $0 \leq \mathrm{Pm} \leq 1$) and the concentration (κ; $0 \leq \kappa$) parameters of the von Mises distribution (equivalent to $1/SD^2$ of the circular normal):

$$\mathrm{p(Y|X)} = \mathrm{Pm} \times \text{von Mises}(\mathrm{Y}; \mu, \kappa) + (1 - \mathrm{Pm}) \times \frac{1}{2\pi}. \quad (1)$$

Here, Y denotes a response made to a particular stimulus X. The first term in the equation denotes the probability density of a specific response given the parameters (mean and precision) of the von Mises distribution, multiplied by the mixture coefficient Pm. The von Mises precision parameter κ is typically interpreted as the precision of a memory representation since larger values of κ generate narrower distributions (i.e., κ is the inverse of the distribution's variance). The parameter μ, denoting the mean of the von Mises density, was set to the target value (X) in the fitting. The second term in the equation refers to the density of

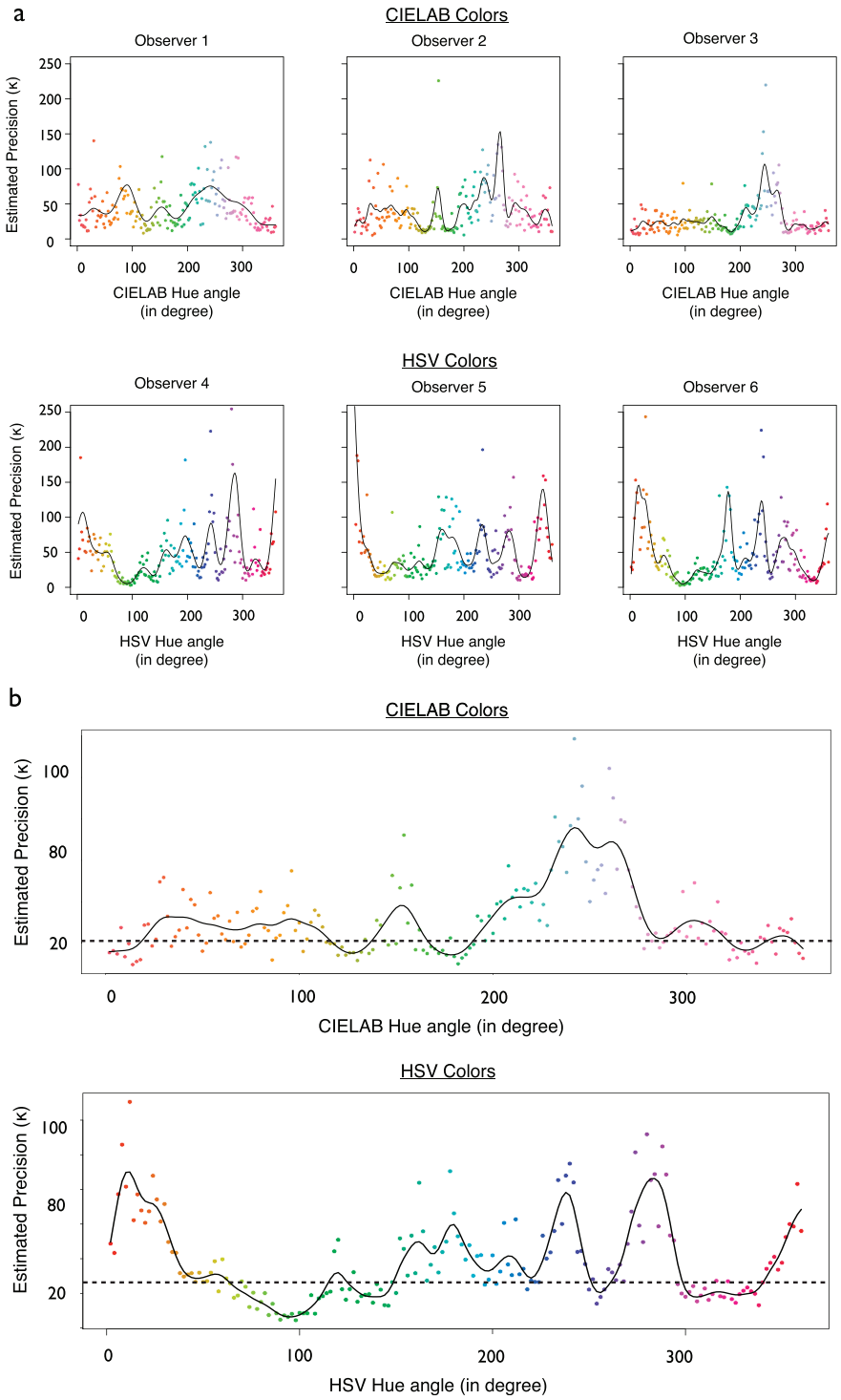

Figure 4. Estimated κ values fitted by color and observer for each of the two color spaces (a), and estimated κ values fitted by color, collapsed across observers. Solid black curves reflect spline smoothing, showing a pattern of κ variability by color, and dotted black lines show estimated κ when the model was fitted to all responses.

?22

the response according to random guessing (i.e., a uniform density on the circle), weighted by (1 − Pm). The Pm parameter gives the probability of responses that are based on working memory, while (1 − Pm) is the probability of random guesses. We expected guessing rates to be very low in our model fits given the minimal memory load (i.e., Pm was expected to be close to 1). All model fitting was performed by maximum likelihood inference. Parameters were initialized to multiple starting values in an attempt to avoid local maxima.

With respect to the methods employed by Zhang and Luck (2008) and subsequent work, the only deviation in this study involved fitting the model to the responses elicited by each stimulus separately rather than aggregating responses across all stimuli. The model was also fitted separately to each individual observer's data. In sum, we fitted the model 180 times for each observer, allowing us to identify differences in the von Mises precision (κ) values of the response distributions elicited by different color stimuli.

Results and discussion

If working memory performance were independent of color value, as is widely assumed in the working memory literature, precision should be comparable for all color values employed. Contrary to this assumption, we found that response precision varied with stimulus value. This is evident in both the direct estimate of precision as $1/SD$ (Figure 3) and the values of precision estimated from the model (Figure 4a). For each observer (Figure 4a), the fitted precision values obtained for each stimulus clearly varied with hue. This remained true when we collapsed across observers (thus including 30 observations for each individual color). These fitted precision values varied widely from the precision obtained by the typical procedure of fitting the model to all of the colors together (dashed line in Figure 4b).

Most importantly, stimulus-specific variations in estimated precision show similar patterns across observers, suggesting that our findings are due to stimulus properties rather than random fluctuations. To quantify interobserver similarity, we calculated correlations between the estimated stimulus-specific κ values for the three participants in each condition. Correlations across all pairs of observers were significant for both the CIELAB and HSV wheels, as shown in Figure 5 [$t(178) > 3.0$, $p < 0.01$ for all correlations]. The correlations across observers suggest that some target stimuli systematically elicited wider response distributions than others.

One might worry that the response precision estimate for each stimulus is unreliable due to the relatively small number of observations (10 for each color) included in each stimulus- and participant-specific model fit. Crucially, if model fits had excessive variance, then *relative* differences among estimates would not be expected to correlate across observers. To make this point quantitatively, we applied a Monte Carlo permutation test (Higgins, 2004, chapter 5). Specifically, for each pair of observers we calculated a null distribution of correlations as follows. Holding the order of the von Mises precision values for one observer fixed, the precision values for the other observer were randomly permuted 10,000 times; a correlation coefficient was calculated for each permutation. The resulting distribution of coefficients indicates that the empirical correlations between observers are highly unlikely to have arisen by chance ($p < 0.001$): None of the 10,000 simulations yielded a correlation as large as those observed, and none of the simulated correlations were both significant and in the right (positive) direction.

To further investigate the reliability of our precision estimates, we performed a simulation in which our stimulus-specific model was fit to data generated according to the null hypothesis that the true response precision is independent of color. For each of three simulated observers and each of the 180 stimuli, we generated 10 random responses according to the null hypothesis. Each response was obtained from a von Mises distribution with a κ value that was estimated by collapsing across all stimuli in Experiment 1 (indicated by the black dotted line in Figure 4b). Note that estimating precision from aggregated data is the standard way of fitting the mixture model of Equation 1. We then computed stimulus-specific fits to the responses of each of the simulated observers and calculated pairwise correlations of estimated κ values as above. This entire process was repeated 1,000 times.

If the correlations between observers in our experiments arise simply as a byproduct of random noise, this simulation should produce correlations similar in magnitude to those of Figure 5. But only one of the simulated across-observer precision correlations was larger in magnitude than the smallest of the correlations found in our experimental data. Thus the probability of obtaining correlations as strong as those found empirically, if the null hypothesis were true, is estimated to be smaller than 1/3,000 ($p < 0.001$).

Given both the strong correlations among observers and the simulation results, it seems unlikely that the stimulus-specific variations in precision are an artifact of unreliable estimates. However, the accuracy of the estimates is uncertain because of the large deviations between rendered and nominal stimuli. In later sections we return to this question. For now we note that, regardless of the ultimate cause, stimulus-specific

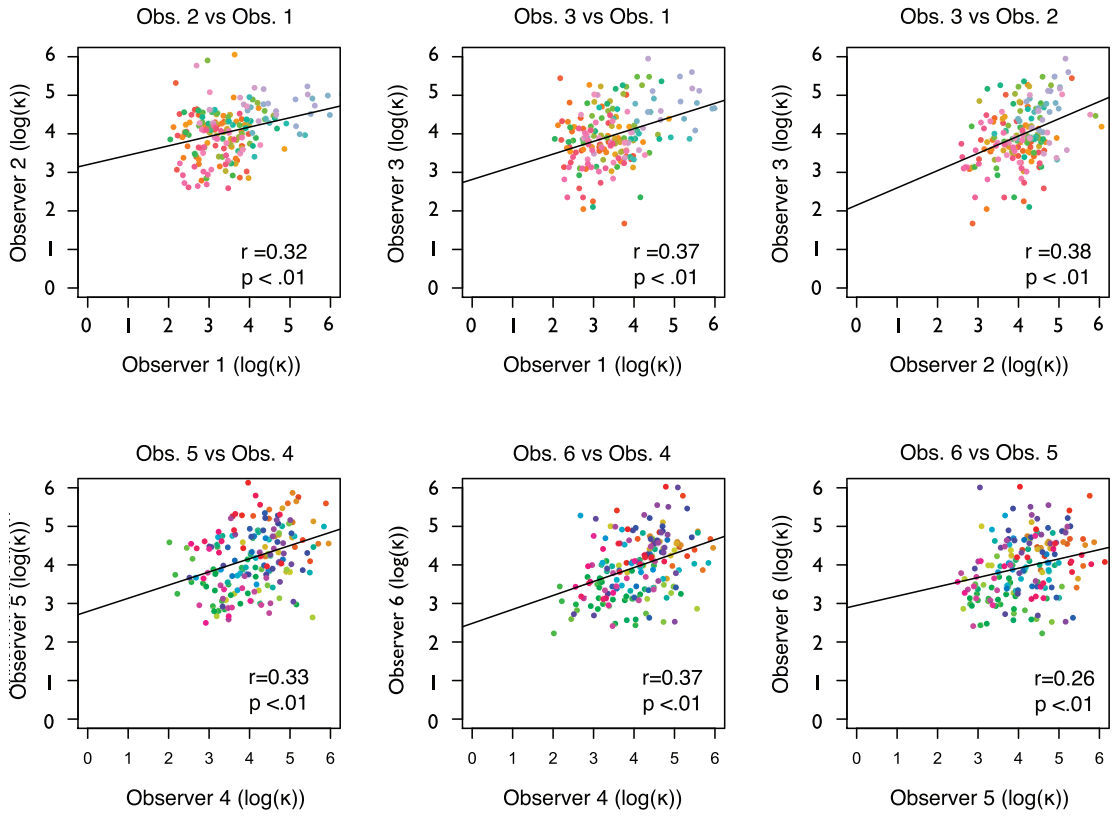

Figure 5. Correlation of stimulus-specific κ estimates across observers. CIELAB color space is shown in the top row and HSV color space is shown in the bottom row.

variation in precision is likely to be widespread in color working memory studies, as we used rendering techniques that are standard in the working memory literature.

We also explored alternative options for analyzing the data, including drawing a new circle in CIELAB space that had minimum distance in color space to the rendered stimuli. This analysis also showed considerable stimulus-specific variability in κ, though the pattern of variation differed from that we found using the nominal stimuli. Importantly, however, the observer-by-observer correlations for κ across colors remained significant (observer 1 vs. observer 2, $r^2 = 0.49$; observer 2 vs. observer 3, $r^2 = 0.35$; observer 1 vs. observer 3, $r^2 = 0.31$; $p < 0.001$ for all correlations), again suggesting stimulus-specific variability. Here we focus only on the analysis of the nominal stimuli. Given the device specificity of the rending process and the fact that other working memory studies also likely suffer from rendering errors, we emphasize that the important result is not the particular pattern of stimulus-specific variability but the fact that stimulus-specific variability exists and is correlated between observers.

Possible effects of response method and exposure duration

While much of the relevant literature utilizes procedures like those described here, two important experimental variables that differ across studies are the time of exposure to memory displays and the method for selecting a response. We used 500 ms exposures and a color-wheel response method in the experiments just described. But we also sought to determine whether our findings hold for other exposure durations and a different way of making responses. A new group of two observers was tested in a nearly identical experiment with the CIELAB color space, with two key differences: (1) Memory samples were displayed for only 100 ms and (2) we used a scrolling method for collecting responses (Figure 6; as in van den Berg et al., 2012). At test, the response was made with a single square rather than a wheel. The response square initially appeared in a randomly chosen color. Observers used two keys to scroll through the available colors in either angular direction, and pressed a button when the displayed color matched their memory of the target.

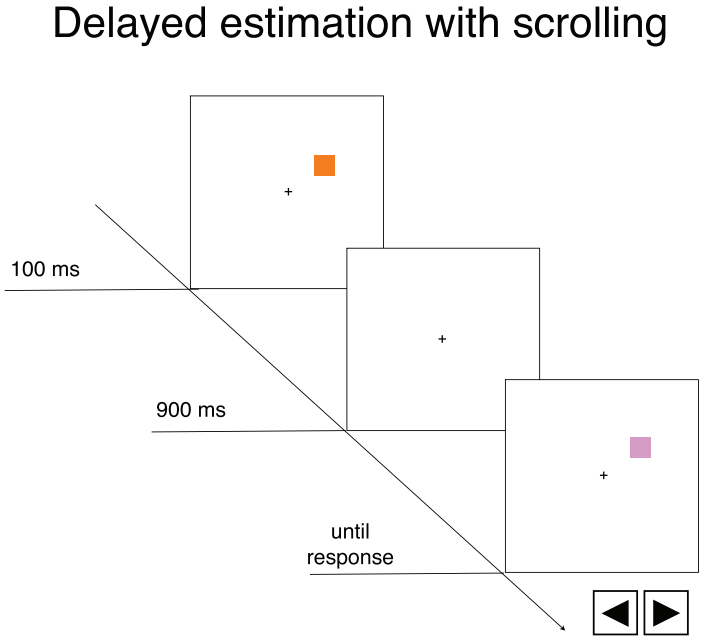

Figure 6. Delayed estimation with the scrolling method.

After fitting the model of Equation 1 in the same way described above, we found a significant correlation between color-specific precision estimates from this experiment and those from the previous one [$t(178) = 4.46$, $r = 0.35$, $p < 0.001$; Figure 7). These results demonstrate that systematic color-specific variability arises for reasons independent of the exposure duration and response method.

Summary

Experiment 1 identified stimulus-dependent response effects in a working memory experiment with standard methods. This undermines the assumption of uniform response variability across colors.

Experiment 2: Estimation without delay

Experiment 1 showed that the response distributions in a standard delayed estimation task contained stimulus-dependent variability. Next, we investigated the source of this variability. Since there were differences between nominal and rendered colors, it is likely that there was perceptual inhomogeneity in the stimuli. To investigate the relationship between precision in memory and precision in perception, we replicated Experiment 1 but without a memory delay. Stimuli appeared simultaneously with the response color wheel, and the task was simply to identify the color of a cued item by clicking on the wheel.

Methods

Observers

A new group of 14 Johns Hopkins University undergraduates participated in exchange for course credit. Each observer had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus and stimuli

The apparatus and stimuli were identical to those used in Experiment 1.

Procedure

This experiment was identical to Experiment 1 with the following exceptions. The color target and response wheel were presented simultaneously. The condition with the HSV color space included set sizes of one, two, four, or six objects in a display, distributed randomly over the course of a session; 12 observers completed 60 trials for each set size (240 trials total) in this space. In the CIELAB condition, two observers completed 1,800 trials each, 10 trials for each color presented singly. In the HSV condition, where multiple objects could be

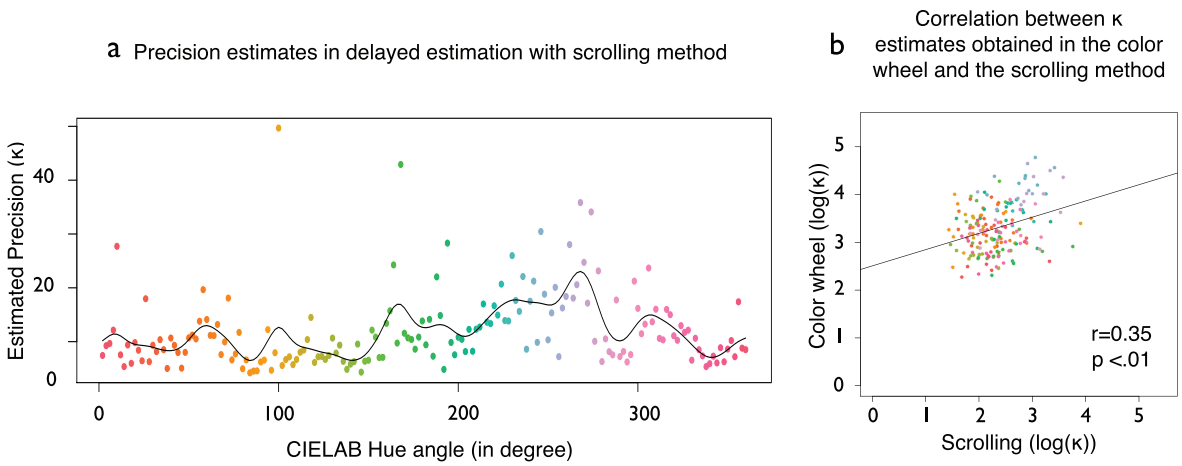

Figure 7. (a) Color-by-color κ estimates in the scrolling experiment. (b) Color-by-color correlation between κ estimates obtained in the color wheel and scrolling experiments (collapsed across observers).

presented, one square included a bold black frame identifying it as the item on which the response should be based.

Analysis

A stimulus-specific mixture model (including precision as a free parameter for each color) was fitted to the results of Experiment 2, collapsing across the responses of all observers in each color space. Although the HSV experiment employed multiple set sizes, we did not expect any specific effect of set size because no delay was introduced and the display remained on the screen until the response. For these reasons, we collapsed the HSV data across set size in the subsequent analyses. (Multiple set sizes were tested for unrelated reasons pertaining to other ongoing research.)

Results and discussion

Overall, responses were significantly more precise in this experiment than in Experiment 1 [with delay vs. without delay; CIELAB mean κ: 44.10 vs. 66.10, $t(179) = -7.17$, $p < 0.001$; HSV mean κ: 52.21 vs. 64.34, $t(179) = -2.49$, $p = 0.017$]. This is consistent with findings using calibrated stimuli that discrimination thresholds tend to increase with the addition of a delay (Nemes, Perry, & McKeefry, 2010; Nilsson & Nelson, 1981; ?15 Olkkonen & Allred, 2014).

Importantly, estimated precision continued to vary by stimulus, even in the absence of an explicit memory demand (Figure 8a). Estimated precision without a delay was significantly correlated with the estimates obtained with the delay in Experiment 1 [CIELAB κ: $r = 0.37$, $t(178) = 5.29$, $p < 0.001$; HSV κ: $r = 0.49$, $t(178) = 7.57$, $p < 0.001$; Figure 6b].

Summary

The main finding in this experiment is that stimulus-dependent variability is present even in the absence of a memory delay and that this variability is correlated with that found in the with-delay experiment. Minimally, these results suggest that at least some of the stimulus-dependent variability observed in working memory is already present in perceptual color estimation. Because of the differences between nominal and rendered colors, we cannot be certain of the extent to which this particular pattern of variability would persist even with stimuli rendered properly on a CIELAB circle. Note, however, that color discrimination is known to be inhomogeneous across color space even with careful display calibration, suggesting that at least some of the inhomogeneity found here may not be due to lack of calibration or other rendering issues (Bachy, Dias, Alleysson, & Bonnardel, 2012; Danilova & Mollon, 2012; Witzel & Gegenfurtner, 2013). Determining to what extent the stimulus-dependent variability in working memory is present in perceptual estimation, and to what extent it arises from memory processes, is an important question for future research.

Experiment 3: Color-by-color fits for varying memory loads

The canonical effect in research on visual working memory is the degradation of performance with increas-

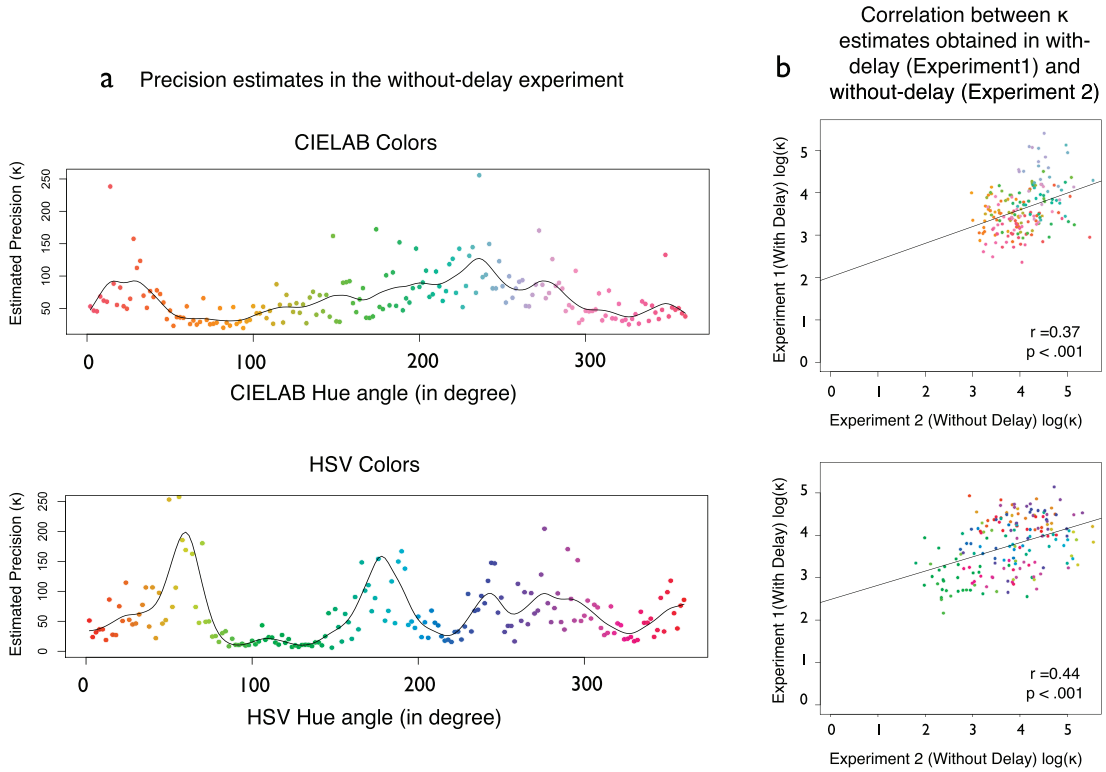

Figure 8. (a) Estimated κ values for CIELAB and HSV color wheels in Experiment 2, and (b) κ correlations between Experiments 1 and 2.

ing memory load, a pattern thought to reveal the limited storage capacity of the system (Cowan, 2001; Luck & Vogel, 1997; Sperling, 1960). With delayed estimation, visual working memory quality—and that of color in particular—has been shown to decline with increasing memory load (though a debate continues about whether it declines further beyond a load of about three or four objects; Anderson & Awh, 2012; Bays et al., 2009; van den Berg et al., 2012). Given the importance of this issue, we sought to examine whether stimulus-dependent differences in response variability persist with increasing memory loads. If they do not, this would surely not settle any debates in working memory research. But we thought it would be important to at least determine whether effects of memory load wash away all stimulus-driven effects. If they do, then the stimulus-dependent variability we identified in Experiments 1 and 2 need not be of central concern to the working memory community. Similarly, if the memory load effect is substantially larger than the stimulus-dependent variability, then it is not functionally important to resolve the methodological problems that led to the differences between nominal and rendered stimuli in our lab (and likely other working memory labs).

A new group of observers participated in a typical delayed estimation experiment with memory load varying on each trial. Again, we fit a mixture model to response distributions by stimulus. If this experiment also produced stimulus-dependent variability in precision estimates, and this variability is correlated with that found in Experiment 1, this would suggest that the effects of memory load do not completely eliminate stimulus-driven effects.

Crucially, although the stimuli for Experiment 3 were nominally the same as those in Experiment 1, they were presented on a different monitor. They were generated using the same procedure as in Experiment 1, and the monitor was not calibrated. Thus, the rendered stimuli in Experiment 3 likely differ both from the nominal values and from the rendered stimuli in Experiments 1 and 2. If any stimulus-dependent variability in Experiment 3 remains correlated with that in Experiment 1, this would suggest that the stimulus-specific effects are large enough to survive failures to appropriately calibrate the display.

Method

Observers

A new group of 24 Johns Hopkins University undergraduates participated in exchange for course

credit. Each observer had normal or corrected-to-normal visual acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus and stimuli

Stimuli were presented on an iMac computer (Apple, Inc., Cupertino, CA) with a liquid-crystal display monitor. The stimuli were generated using the same procedures as in Experiment 1. However, since the stimuli were displayed on a different monitor, and because this monitor was not calibrated, the stimuli likely differed from both the nominal stimuli and those rendered in Experiment 1. These stimuli were not measured posthoc.

Procedure

Experiment 3 was identical to Experiment 1 except as noted below. This experiment utilized the CIELAB color space from Experiment 1. The experiment included memory loads of one, two, four, and six distributed randomly over the course of a session. For half of the observers, colors were sampled randomly on each trial. For the other half, colors were sampled with the restriction that no two colors were allowed to be closer than 20° (e.g., Fougnie, Asplund, & Marois, 2010). Each observer completed 60 trials for each memory load (240 trials total).

Analysis

As in Experiments 1 and 2, a stimulus-specific mixture model was fitted to the results. There were no obvious differences in parameters when the stimuli in a trial were sampled with restrictions and without, so data from the two conditions were collapsed for all subsequent analyses. Because the number of trials for each target stimulus was not equal, a smoothing algorithm was applied to the estimated stimulus-specific precision (κ) values for each memory load; estimates were smoothed by a moving average of ± 1 adjacent values weighted by the number of observations for each stimulus.

Results and discussion

The correlations between stimulus-specific precision estimates in Experiments 1 and 3 are shown in Figure 9. Overall, when collapsing across memory load, the stimulus-specific effects on precision in Experiments 1 and 3 were correlated. In addition, each memory load in Experiment 3 was independently correlated with the single memory load in Experiment 1 [$t(178) > 4.53$, $p < 0.01$ for all correlations]. The stimulus dependency of precision estimates survived the effect of higher memory loads.

To be certain that these results are not due to artifacts from data analysis (e.g., our smoothing procedure), correlations were applied over the data with randomly permuted stimuli as in Experiment 1. The Monte Carlo simulations confirmed that the correlations we observed are unlikely to have arisen by chance (zero out of 10,000 empirical correlations were significant in the right direction). Thus properties of individual stimulus values produce large variability in subsequent memory responses, and these differences between stimuli persist as memory load increases.

Summary

Overall, this experiment makes it clear that the stimulus-dependent effects identified in Experiment 1 persist as memory load increases (at least within the range most typically studied in related experiments). In addition, several specific aspects of the results are worth emphasizing. First, there is a strong interexperiment correlation between stimulus-specific estimates with a memory load of one. Because two different non-calibrated displays were used in the two experiments, it is very likely that the rendered stimuli in the two experiments were somewhat different from each other. Despite this, the stimulus-specific correlation between experiments survived. This suggests both that the stimulus-specific effects are not particular to one display and that they are large enough to be seen despite inaccuracies in color rendering. In light of this, it seems likely that previous working memory experiments have also obtained systematic but unmeasured variation in color precision.

Second, we note that although stimulus-specific precision estimates were still significantly correlated at higher memory loads of four and six, these correlations were considerably weaker. This could mean that stimulus-specific effects are less important at higher memory loads, but it could also be an artifact of the model. This model has an inherent correlation between the von Mises part (i.e., the estimates of precision, κ) and the uniform part (i.e., guessing, $1 - \mathrm{Pm}$) of the equations (see also Suchow et al., 2013). If the estimate of guessing rate is low, then the precision estimate must also be lower to accommodate extreme responses; conversely, if a higher guessing rate is estimated, then the von Mises component will have high responsibility for accurate responses only, leading to larger precision estimates. High rates of guessing at higher memory loads could lead the mixture model to overestimate precision for stimuli that tend to elicit noisy responses. Indeed, the specific stimuli that deviated most from the expected correlations seem to be those that were

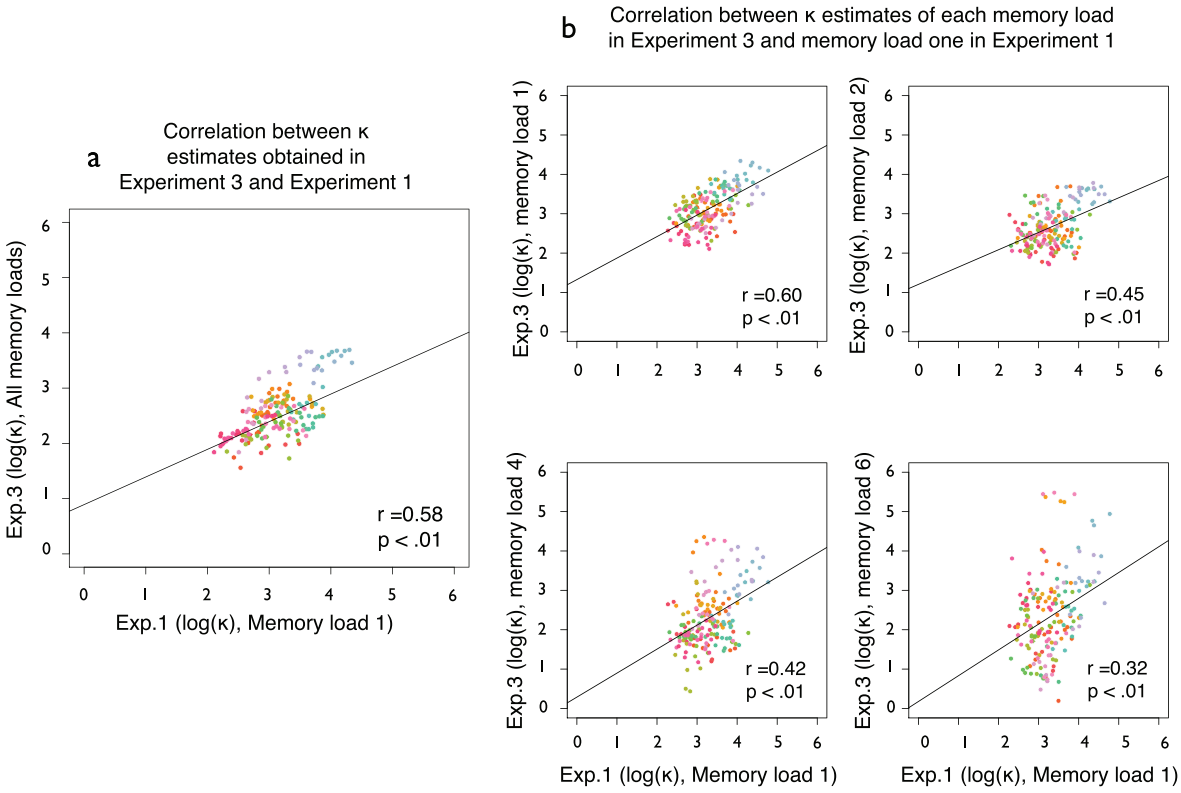

Figure 9. Correlations between color-specific κ values obtained in Experiment 1 ($N = 3$) and Experiment 3 ($N = 24$; CIELAB color wheel). (a) Correlations between the two experiments, collapsing across all memory loads in Experiment 3. (b) Correlations between each individual memory load in Experiment 1 and Experiment 3.

imprecise with a memory load of one; we suspect that the response precision associated with these stimuli at higher memory loads was overestimated. (Note that nontarget responses would be expected to have the same effect on precision as guesses with respect to estimates of precision; consult Bays, Catalao, & Husain, 2009.)

Experiment 4: A relationship between labeled color regions and response precision

Experiments 1 through 3 demonstrated that the response distributions elicited by some stimuli were more variable than those elicited by other stimuli. Experiment 2, in particular, demonstrated that stimulus-specific effects in a perception experiment correlated with those in the memory experiments. This suggests that the memory effects of color are likely mediated by the perceptual representation of color. To investigate whether this pattern of variability is common to different perceptual tasks, we conducted a color labeling experiment. Observers were shown the same rendered colors that served as the response wheel in Experiments 1 and 2, and they were asked to identify the best example of each of seven color terms. If the observers are consistent in their responses, and if their responses are systematically related to the stimulus-specific precision estimates, it could shed some insight on the origin of the variability. Furthermore, in light of the deviations between nominal and rendered stimuli, a sensible relationship between color-labeling and stimulus effects would provide additional reassurance that the effects we measured survive the inaccuracies of color rendering.

Method

Observers

A new group of eight Johns Hopkins University undergraduates participated in exchange for course credit. Each had normal or corrected-to-normal visual

acuity. The protocol for this experiment was approved by the Johns Hopkins University IRB.

Apparatus and stimuli

Apparatus and stimuli were identical to those used in Experiment 1.

Procedure

At the start of a trial, the 180 stimuli rendered in Experiment 1 were presented in the center of the screen organized around a wheel. To the right of the wheel the seven color terms *orange, yellow, lime, green, blue, purple,* and *pink* were vertically presented in a random order. We did not use the basic color terms (Berlin & Kay, 1969) because we did not expect this set of color samples to be labeled by all and only basic terms. Recall that the stimulus set was constructed with constraints unrelated to color terminology. Specifically, we excluded the terms *red* and *brown* because of the high luminance values of the sampled colors, and we included the term *lime* because of the relatively wide spectrum of samples that appeared green to us.

Observers were asked to find the best example of each of these seven colors by clicking on the color wheel. The color wheel stayed on the screen until an observer made seven responses. The color wheel was rotated randomly on each trial, and each observer completed 20 trials.

Results and discussion

Combining data from eight observers yielded 160 responses for each color term. Histograms of the results ?16 are given in Figure 11a. Response distributions for each term appeared approximately normal. In other words, certain stimuli were noisy attractors for the color labels, with responses for a given label diminishing rapidly with distance from the corresponding attractor. We emphasize that our intent was not to identify definite category boundaries or focal points. More sophisticated methods are available for this purpose and could be employed in future research (e.g., Witzel & Gegenfurtner, 2013), and they would be more appropriate with a set of color samples designed for this purpose and faithfully rendered. Instead, we merely sought to determine whether observers shared even loose intuitions about how to label this stimulus set and whether the labeling structure was related to the stimulus-specific precision effects observed in Experiments 1 through 3.

Toward this end, inspection of the histograms in Figure 10a reveals several important features. First, the fact that response distributions are relatively clustered indicates that observers shared intuitions about labels for the color stimuli. Second, the amount of variability in selection of the best instance of a label differs across color space, as seen by the varying width of the peaks in Figure 10a. For example, observers were more consistent in their labeling of purple than lime. The variability could be within observer, between observers, or both. Third, it appeared that there was a relationship between a label's consistency in Experiment 4 and the response precision elicited by those colors in the other experiments. For example, the stimuli labeled as the best exemplars of purple also elicited response distributions with the highest precision in Experiments 1 and 2.

To explore this observation quantitatively, we fit seven von Mises distributions to the histograms in Figure 10a, comparing likelihood values to identify operational boundaries in this spatial arrangement of the rendered colors. This in turn afforded a measure of a label's spread: namely, the number of individual colors between boundaries (orange: 40; yellow: 20; lime: 18; green: 23; blue: 19; purple: 15; pink: 45). To relate these findings to the response precision associated with these colors, we averaged together the κ values obtained in Experiment 1 for all of the stimuli falling within each of the seven bounded regions. There was a significant negative correlation between the number of stimuli in a region and the average precision of stimuli within that region [$t(5) = -2.87$, $r = -0.79$, $p = 0.035$; Figure 10b].

We emphasize again that these data were not collected for the purpose of understanding color categorization. Rather, they provide some traction in understanding the stimulus-specific effects in perception (Experiment 2, Figure 8) that are propagated to working memory (Experiment 1, Figure 4). Empirically, we have demonstrated that stimuli falling in labeled regions that are narrow tend to produce narrower response distributions. This is of general interest, but it is particularly important in the context of the present experiments because it provides added verification that the stimulus-specific variability observed in perception and memory is not solely an artifact of the deviation between nominal and rendered stimuli.

General discussion

This study sought to investigate assumptions built into an influential and rapidly growing literature that utilizes the delayed estimation paradigm with color to draw conclusions about the structure and limits of visual working memory (Anderson & Awh, 2012; Bays et al., 2009, 2011; Emrich & Ferber, ~~2011~~; Fougnie & Alvarez, 2011; Fougnie et al., 2010, 2012; Gold et al.,

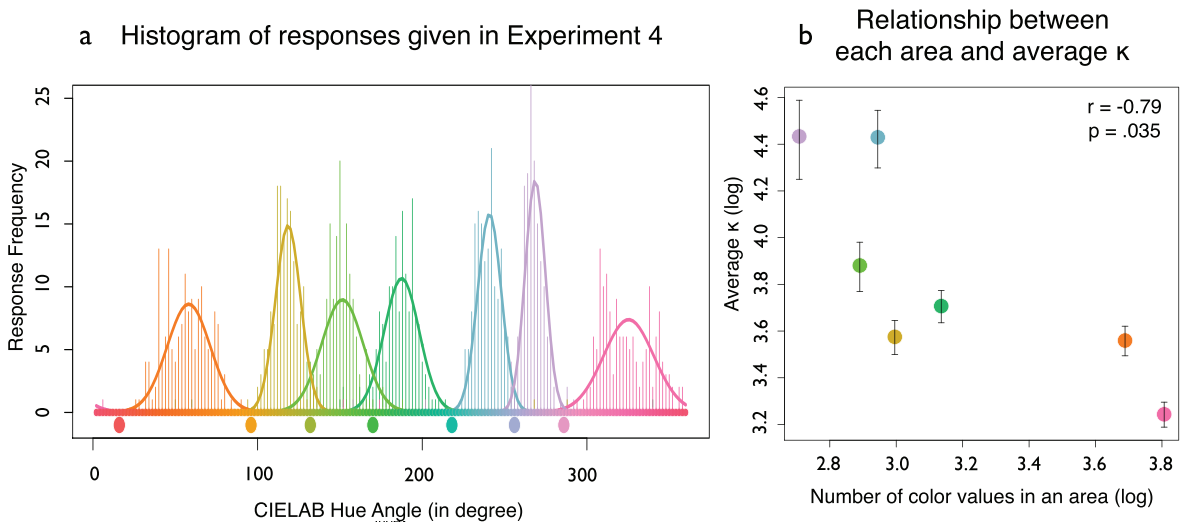

Figure 10. Results of Experiment 4. (a) Histogram of responses given as the best example of each of seven color terms, along with seven von Mises distributions fitted to these responses. Overlap points between the distributions designate operational boundaries between regions on the wheel. (b) The relationship between the size of each identified region (in terms of the number of individual colors) and the average κ value obtained for the colors in that region in Experiment 1.

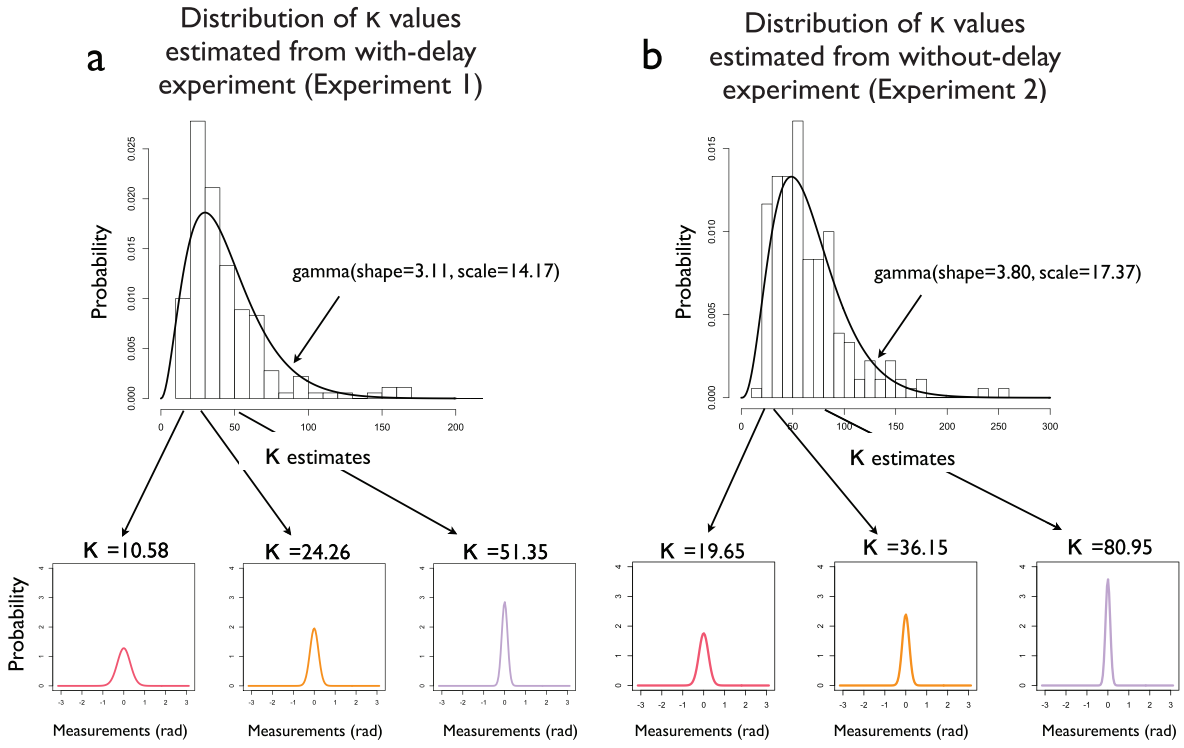

Figure 11. Distribution of estimated precisions (κ) obtained in Experiment 1 and Experiment 2. (a) Histogram of κ values for all colors estimated in the with-delay experiment (Experiment 1, CIELAB). The black curve represents the best fit gamma distribution. The smaller graphs show von Mises densities for three example colors. (b) The same graph with the data from the without-delay experiment (Experiment 2).

2010; van den Berg et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008, 2009, 2011). The assumptions are that the colors rendered in the relevant experiments have been sampled from a circle in CIELAB color space, that they are perceptually homogenous, that they have equal lightness values, and that they tend to elicit equally variable response distributions—assumptions that motivate value-independent analysis of response distribution properties.

In Experiments 1 through 3, we identified effects that undermine these assumptions. First, we found stimulus-specific response variability that correlated across unique observers (Experiment 1). This stimulus-specific variability was present even in an experiment without an explicit memory demand (Experiment 2) and persisted with larger memory loads (Experiment 3). Experiment 4 demonstrated that some of this stimulus-specific variability likely arose during interaction with the stimulus set at test.

Importantly, we determined posthoc that the colors rendered in Experiments 1, 2, and 4 were different from the nominal colors specified, differing in lightness and nonuniformly distributed in CIELAB space. In this regard, our methods reflect a kind of "worst case" scenario with respect to stimulus display. We did not apply the calibration procedures that ensure faithful rendering of color stimuli. We emphasize that although calibration procedures are standard in studies of color perception (Allen, Beilock, & Shevell, 2012; Olkkonen & Allred, 2014; Witzel & Gegenfurtner, 2013; Xiao, Hurst, MacIntyre, & Brainard, 2012), calibration procedures have not been implemented in delayed estimation reports (see, e.g., Suchow et al., 2013, which provides a tutorial for conducting such experiments and analyses without mention of calibration or color rendering validation). Indeed, we could not identify even one study that reported implementing a full calibration procedure (e.g., that described in Brainard, ?17 Robson, & Pelli, 2002). Neither could we find reports that confirmed the properties of color stimuli using posthoc radiometer measurements.

Taken together, our results undermine the assumption, endemic in the delayed estimation literature, that colored stimuli are represented homogenously in memory. This is not a minor issue: Stimulus homogeneity is central to the parameter fits used to make inferences about the structure of working memory and to compare alternative memory models.

The assumption of stimulus homogeneity is perhaps most clear in two recent studies, which argued that visual working memory varies stochastically on a moment-to-moment, trial-by-trial, or even item-by-item basis (Fougnie et al., 2012; van den Berg et al., 2012). These inferences are predicated on the assumption that there is no systematic reason for response distributions to vary trial by trial—in particular, no reason for the precision of the distribution to vary according to the trial-specific stimulus. Specifically, one previous model of working memory proposed that memory precision varies from moment to moment in a way that can be described as sampling from a gamma distribution (van den Berg et al., 2012). However, our stimulus-specific precision estimates obtained in Experiments 1 and 2 produced perfect gamma distributions (Figure 11a and b). An important future direction would be investigating how much variability in working memory precision should be attributed to trial-by-trial variability after stimulus-specific variability in perception and working memory has been taken into account.

Similar points extend to what is perhaps the central debate in this area: whether or not there is a discrete capacity limit in addition to a more continuous precision limit on memory for smaller memory loads. This debate has often hinged on whether the estimate of memory precision plateaus at larger memory loads and on whether guessing rates increase in a way that suggests frequent guesses when memory load exceeds some fixed quantity. But these parameter values are obtained as the best fits of models that assume that responses to a memory target (as opposed to guesses) are drawn from a distribution whose characteristics are independent of the target color. Similarly, some studies have suggested that observers make responses that are not random but that also are not drawn from a target representation. Instead, responses may be based on a nontarget display item (Bays et al., 2009; Emrich & Ferber, 2012). Correctly estimating the probability of such nontarget responses depends on accurate expectations about the response distributions for specific target and nontarget colors.

Conclusions

We draw two broad conclusions from our results. The first is prescriptive: Replications and extensions of previous work should be conducted with CIELAB specified color stimuli that are rendered faithfully following standardized calibration procedures (Brainard, Pelli, & Robson, 2002; Gegenfurtner & Kiper, 2003).

Second, it seems likely that the assumption of perceptual and memory homogeneity among stimuli in studies of color working memory is unwarranted. Our results indicate clearly that stimulus-specific variation in precision exists, even though the deviations between nominal and rendered stimuli do not allow us to definitively establish the cause of those stimulus-specific effects. More broadly, our results motivate examination of implicit homogeneity assumptions for other stimulus classes. For example, in the case of orientation

a variety of phenomena suggest differences in the fidelity of the representation of oblique and cardinal values (Girshick, Landy, & Simoncelli, 2011; Wolfe, Klempen, & Shulman, 1999)—differences that have not been incorporated into the modeling of delayed orientation estimation experiments (Fougnie & Alvarez, 2011; Keshvari, van den Berg, & Ma, 2013; van den Berg et al., 2012).

Ultimately, we suggest that a complete understanding of the structure of visual working memory and its capacity limits will require a stimulus-specific understanding of both perceptual and memory representations. For this reason, despite the caution provided here, color remains a good candidate for a stimulus class with which to investigate working memory. A great deal is known about color perception, including its neurophysiological basis, the computations that support color adaptation and constancy, and its relationship to higher-level reasoning and language (for reviews of color vision see, e.g., Gegenfurtner & Kiper, 2003; Solomon & Lennie, 2007). This creates a unique opportunity for combining expertise across areas to relate visual working memory to visual perception and cognition more broadly.

Keywords: visual working memory, delayed estimation, color

Acknowledgments

We thank George Alvarez, Tim Brady, Daryl Fougnie, and Weiwei Zhang for sharing details of their ?18 experimental procedures with us.

Commercial relationships: none.
Corresponding author: Jonathan Isaac Flombaum.
Email: flombaum@jhu.edu.
Address: Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA.

References

Allen, E. C., Beilock, S. L., & Shevell, S. K. (2012). Individual differences in simultaneous color constancy are related to working memory. *Journal of the Optical Society of America A, 29*, A52–A59.

Anderson, D. E., & Awh, E. (2012). The plateau in mnemonic resolution across large set sizes indicates discrete resource limits in visual working memory. *Attention, Perception, and Psychophysics, 74*, 891–910.

Bachy, R., Dias, J., Alleysson, D., & Bonnardel, V. (2012). Hue discrimination, unique hues and naming. *Journal of the Optical Society of America A, 29*, A60–A68.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10):7, 1–11, http://www.journalofvision.org/content/9/10/7, doi:10.1167/9.10.7.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia, 49*, 1622–1631.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. City, CA: University of California Press. ?19

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Brainard, D. H. (1998). Color constancy in the nearly natural image II. Achromatic loci. *Journal of the Optical Society of America A, 17*, 307–325.

Brainard, D. H. (2003). Color appearance and color difference specification. In S. K. Shevell (Ed.), *The science of color* (pp. 191–216). Oxford, UK: Elsevier Science.

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). *Display characterization. Encyclopedia of imaging science and technology*. City, State: John Willey and Sons. ?20

Burnham, R. W., & Clark, J. R. (1954). A color memory test. *Journal of the Optical Society of America, 44*, 658–659.

Burnham, R. W., & Clark, J. R. (1955). A test of hue memory. *Journal of Applied Psychology, 39*, 164–172.

Collins, M. (1931). Some observations on immediate colour memory. *British Journal of Psychology, 22*, 344–351.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114.

Danilova, M. V., & Mollon, J. D. (2012). Foveal color perception: Minimal thresholds at a boundary between perceptual categories. *Vision Research, 62*, 162–172.

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision, 12*(4):12, 1–16, http://www.journalofvision.org/content/12/4/12, doi:10.1167/12.4.12.

Fairchild, M. D. (1998). *Color appearance models*. Reading, MA: Addison-Wesley.

Fougnie, D., & Alvarez, G. A. (2011). Object features

fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12):3, 1–12, http://www.journalofvision.org/content/11/12/3, doi:10.1167/11.12.3.

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12):27, 1–11, http://www.journalofvision.org/content/10/12/27, doi:10.1167/10.12.27.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications, 3*, *1229*,doi:10.1038/ncomms2237.

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Neuroscience, 26*, 181–206.

Giesel, M., & Gegenfurtner, K. R. (2010). Color appearance of real objects varying in material, hue, and shape. *Journal of Vision*, *10*(9):10, 1–21, http://www.journalofvision.org/content/10/9/10, doi:10.1167/10.9.10.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14*, 926–934.

Gold, J. M., Hahn, B., Zhang, W., Robinson, B. M., Kappenman, E. S., Beck, V. M., et al. (2010). Reduced capacity but shared precision and maintenance of working memory representations in schizophrenia. *Archives of General Psychiatry, 67*, 570–577.

Higgins, J. J. (2004). *An introduction to modern nonparametric statistics*. Pacific Grove, CA: Brooks/Cole.

Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A, 13*, 1981–1991.

Keshvari, S., van den Berg, R., & Ma, W. (2013). No evidence for an item limit in change detection. *PLoS Computational Biology, 9*(2), e1002927, doi:10.1371/journal.pcbi.1002927.

Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A, 25*, 1215–1226.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Nemes, V. A., Parry, N. R., Whitaker, D., & McKeefry, D. J. (2012). The retention and disruption of color information in human short-term visual memory. *Journal of Vision*, *12*(1):26, 1–14, http://www.journalofvision.org/content/12/1/26, doi:10.1167/12.1.26.

Nilsson, T. H., & Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 141–150.

Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PloS One, 9*, e8648, 1–11.

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: 1. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 261–282.

Solomon, S. G., & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience, 8*, 276–286.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*, 1–29.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10):9, 1–8, http://www.journalofvision.org/content/13/10/9, doi:10.1167/13.10.9.

van den Berg, R., Shin, H., Chou, W., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, USA, 109*, 8780–8785.

Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000a). Variations in normal color vision. I. Cone-opponent axes. *Journal of the Optical Society of America A, 17*, 1535–1544.

Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000b). Variations in normal color vision. II. Unique hues. *Journal of the Optical Society of America A, 17*, 1545–1555.

Wilken, P., & Ma, W. J. (2004). A detection theory of change detection. *Journal of Vision, 4*(12):11, 1120–1135, http://www.journalofvision.org/content/4/12/11, doi:10.1167/4.12.11.

Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision*, *13*(7):1, 1–33, http://www.journalofvision.org/content/13/7/1, doi:10.1167/13.7.1.

Wolfe, J. M., Klempen, N. L., & Shulman, E. P. (1999). Which end is up? Two representations of orientation in visual search. *Vision Research, 39*, 2075–2086.

Wyszecki, G., & Stiles, W. S. (1982). *Color science:*

Concepts and methods, quantitative data and formulae (2nd ed.). New York: John Wiley and Sons.

Xiao, B., Hurst, B., MacIntyre, L., & Brainard, D. H. (2012). The color constancy of three-dimensional objects. *Journal of Vision*, *12*(4):6, 1–15, http://www.journalofvision.org/content/12/4/6, doi:10.1167/12.4.6.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20*, 423–428.

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science, 22*, 1431–1441.

Appendix A

	Intended CIELAB coordinates			Rendered CIELAB coordinates			Measured xyY coordinates		
Color	L*	a*	b*	L*	a*	b*	x	y	Y
1	70	80.000	38.000	62.693	71.424	27.188	0.518	0.311	23.917
2	70	79.963	40.094	62.605	71.274	29.222	0.528	0.314	23.229
3	70	79.854	42.185	62.541	71.066	31.297	0.535	0.319	23.192
4	70	79.671	44.272	62.503	70.799	33.410	0.542	0.324	23.052
5	70	79.416	46.350	62.491	70.474	35.557	0.548	0.328	22.955
6	70	79.088	48.419	62.503	70.092	37.737	0.554	0.332	22.909
7	70	78.689	50.475	62.539	69.653	39.944	0.559	0.337	22.938
8	70	78.218	52.515	62.601	69.158	42.176	0.566	0.340	22.818
9	70	77.676	54.538	62.686	68.610	44.428	0.570	0.345	22.970
10	70	77.063	56.541	62.794	68.010	46.696	0.574	0.349	22.948
11	70	76.382	58.521	62.925	67.360	48.977	0.578	0.354	23.137
12	70	75.631	60.476	63.078	66.661	51.264	0.581	0.358	23.255
13	70	74.813	62.404	63.253	65.916	53.554	0.584	0.361	23.296
14	70	73.928	64.302	63.447	65.129	55.841	0.585	0.364	23.449
15	70	72.977	66.168	63.661	64.300	58.121	0.587	0.367	23.501
16	70	71.962	68.000	63.893	63.433	60.387	0.588	0.370	23.753
17	70	70.883	69.795	64.170	62.612	61.015	0.588	0.373	23.994
18	70	69.742	71.552	64.469	61.775	61.251	0.587	0.375	24.253
19	70	68.541	73.267	64.780	60.904	61.497	0.587	0.377	24.406
20	70	67.281	74.940	65.103	60.001	61.752	0.585	0.379	24.697
21	70	65.963	76.567	65.437	59.071	62.016	0.583	0.381	24.957
22	70	64.589	78.148	65.781	58.116	62.288	0.580	0.382	25.316
23	70	63.160	79.680	66.134	57.139	62.567	0.579	0.384	25.638
24	70	61.680	81.160	66.496	56.143	62.852	0.576	0.386	25.987
25	70	60.148	82.589	66.864	55.130	63.143	0.573	0.388	26.291
26	70	58.567	83.963	67.239	54.104	63.438	0.571	0.389	26.632
27	70	56.940	85.281	67.619	53.067	63.738	0.569	0.391	27.009
28	70	55.267	86.541	68.004	52.022	64.041	0.566	0.394	27.419
29	70	53.552	87.742	68.392	50.971	64.347	0.563	0.395	27.815
30	70	51.795	88.883	68.783	49.916	64.655	0.561	0.397	28.113
31	70	50.000	89.962	69.177	48.859	64.965	0.559	0.399	28.573
32	70	48.168	90.977	69.571	47.803	65.275	0.556	0.401	29.016
33	70	46.302	91.928	69.966	46.750	65.586	0.553	0.403	29.329
34	70	44.404	92.813	70.326	45.629	65.853	0.551	0.405	29.829
35	70	42.476	93.631	70.335	43.761	65.679	0.548	0.407	30.294
36	70	40.521	94.382	70.344	41.865	65.504	0.546	0.409	30.800
37	70	38.541	95.063	70.351	39.943	65.328	0.544	0.410	30.968

	Intended CIELAB coordinates			Rendered CIELAB coordinates			Measured xyY coordinates		
Color	L*	a*	b*	L*	a*	b*	x	y	Y
38	70	36.538	95.676	70.359	37.998	65.152	0.541	0.412	31.559
39	70	34.515	96.218	70.366	36.033	64.976	0.537	0.416	31.378
40	70	32.475	96.689	70.372	34.050	64.801	0.533	0.418	31.454
41	70	30.419	97.088	70.378	32.051	64.625	0.530	0.422	31.129
42	70	28.350	97.416	70.383	30.039	64.451	0.524	0.425	31.022
43	70	26.272	97.671	70.388	28.016	64.278	0.519	0.430	31.089
44	70	24.185	97.854	70.392	25.985	64.106	0.514	0.434	31.015
45	70	22.094	97.963	70.396	23.948	63.936	0.510	0.436	30.760
46	70	20.000	98.000	70.399	21.907	63.767	0.504	0.441	30.784
47	70	17.906	97.963	70.402	19.865	63.600	0.500	0.444	30.655
48	70	15.815	97.854	70.404	17.825	63.435	0.494	0.448	30.668
49	70	13.728	97.671	70.406	15.788	63.272	0.490	0.451	30.439
50	70	11.650	97.416	70.407	13.757	63.112	0.484	0.456	30.526
51	70	9.581	97.088	70.408	11.734	62.954	0.479	0.460	30.480
52	70	7.525	96.689	70.408	9.721	62.798	0.475	0.463	30.230
53	70	5.485	96.218	70.408	7.721	62.645	0.469	0.468	30.318
54	70	3.462	95.676	70.407	5.736	62.495	0.463	0.472	29.992
55	70	1.459	95.063	70.405	3.768	62.347	0.458	0.476	29.817
56	70	−0.521	94.382	70.403	1.818	62.202	0.454	0.479	29.971
57	70	−2.476	93.631	70.400	−0.111	62.060	0.449	0.483	29.728
58	70	−4.404	92.813	70.397	−2.018	61.920	0.441	0.489	30.346
59	70	−6.302	91.928	70.393	−3.900	61.783	0.436	0.494	30.302
60	70	−8.168	90.977	70.387	−5.756	61.649	0.431	0.497	30.121
61	70	−10.000	89.962	70.381	−7.585	61.518	0.426	0.500	30.021
62	70	−11.795	88.883	70.375	−9.384	61.389	0.420	0.506	29.929
63	70	−13.552	87.742	70.367	−11.152	61.263	0.415	0.509	29.917
64	70	−15.267	86.541	70.358	−12.888	61.139	0.411	0.512	29.588
65	70	−16.940	85.281	70.348	−14.590	61.018	0.405	0.517	29.483
66	70	−18.567	83.963	70.337	−16.258	60.899	0.400	0.520	29.442
67	70	−20.148	82.589	70.324	−17.890	60.783	0.395	0.524	29.516
68	70	−21.680	81.160	70.311	−19.485	60.669	0.390	0.528	29.297
69	70	−23.160	79.680	70.296	−21.042	60.557	0.386	0.531	29.123
70	70	−24.589	78.148	70.280	−22.560	60.447	0.382	0.533	29.161
71	70	−25.963	76.567	70.262	−24.039	60.339	0.378	0.537	29.154
72	70	−27.281	74.940	70.242	−25.477	60.232	0.374	0.540	28.812
73	70	−28.541	73.267	70.221	−26.874	60.127	0.369	0.544	29.010
74	70	−29.742	71.552	70.198	−28.230	60.024	0.365	0.547	29.076
75	70	−30.883	69.795	70.174	−29.543	59.923	0.362	0.549	28.925
76	70	−31.962	68.000	70.147	−30.814	59.822	0.358	0.552	28.911
77	70	−32.977	66.168	70.119	−32.042	59.723	0.354	0.555	28.712
78	70	−33.928	64.302	70.088	−33.227	59.625	0.351	0.557	28.974
79	70	−34.813	62.404	70.056	−34.367	59.527	0.347	0.558	28.821
80	70	−35.631	60.476	70.021	−35.464	59.431	0.344	0.558	28.722
81	70	−36.382	58.521	70.000	−36.382	58.521	0.339	0.557	29.004
82	70	−37.063	56.541	70.000	−37.063	56.541	0.334	0.555	28.750
83	70	−37.676	54.538	70.000	−37.676	54.538	0.330	0.551	28.663
84	70	−38.218	52.515	70.000	−38.218	52.515	0.326	0.547	28.965
85	70	−38.689	50.475	70.000	−38.689	50.475	0.321	0.542	29.055
86	70	−39.088	48.419	70.000	−39.089	48.419	0.317	0.537	28.913
87	70	−39.416	46.350	70.000	−39.416	46.350	0.312	0.530	28.751
88	70	−39.671	44.272	70.000	−39.671	44.272	0.307	0.521	28.857
89	70	−39.854	42.185	70.000	−39.854	42.185	0.303	0.514	28.798
90	70	−39.963	40.094	70.000	−39.963	40.094	0.298	0.508	29.134

	Intended CIELAB coordinates			Rendered CIELAB coordinates			Measured xyY coordinates		
Color	L*	a*	b*	L*	a*	b*	x	y	Y
91	70	−40.000	38.000	70.000	−40.000	38.000	0.293	0.497	29.186
92	70	−39.963	35.906	70.000	−39.963	35.906	0.289	0.488	29.205
93	70	−39.854	33.815	70.000	−39.854	33.815	0.285	0.479	29.304
94	70	−39.671	31.728	70.000	−39.671	31.728	0.281	0.469	29.384
95	70	−39.416	29.650	70.000	−39.416	29.650	0.277	0.460	29.394
96	70	−39.088	27.581	70.000	−39.089	27.581	0.273	0.448	29.137
97	70	−38.689	25.525	70.000	−38.689	25.525	0.270	0.438	29.330
98	70	−38.218	23.485	70.000	−38.218	23.485	0.266	0.428	29.331
99	70	−37.676	21.462	70.000	−37.676	21.462	0.262	0.418	29.482
100	70	−37.063	19.459	70.000	−37.063	19.459	0.259	0.408	29.649
101	70	−36.382	17.479	70.000	−36.382	17.479	0.257	0.401	29.698
102	70	−35.631	15.524	70.000	−35.631	15.524	0.254	0.389	29.588
103	70	−34.813	13.596	70.000	−34.813	13.596	0.251	0.379	29.639
104	70	−33.928	11.698	70.000	−33.928	11.698	0.249	0.372	29.937
105	70	−32.977	9.832	70.000	−32.977	9.832	0.247	0.361	29.638
106	70	−31.962	8.000	70.000	−31.962	8.000	0.245	0.353	29.795
107	70	−30.883	6.205	70.000	−30.883	6.205	0.243	0.346	30.062
108	70	−29.742	4.448	70.000	−29.742	4.448	0.241	0.336	29.855
109	70	−28.541	2.733	70.000	−28.541	2.733	0.240	0.330	30.120
110	70	−27.281	1.060	70.000	−27.281	1.060	0.239	0.322	30.023
111	70	−25.963	−0.567	70.000	−25.963	−0.567	0.237	0.314	30.341
112	70	−24.589	−2.148	70.000	−24.589	−2.148	0.236	0.306	30.461
113	70	−23.160	−3.680	70.000	−23.160	−3.680	0.235	0.298	30.388
114	70	−21.680	−5.160	70.000	−21.680	−5.160	0.234	0.293	30.643
115	70	−20.148	−6.589	70.000	−20.148	−6.589	0.235	0.288	30.611
116	70	−18.567	−7.963	70.000	−18.567	−7.963	0.234	0.281	30.589
117	70	−16.940	−9.281	70.000	−16.940	−9.281	0.234	0.278	30.932
118	70	−15.267	−10.541	70.000	−15.267	−10.541	0.234	0.271	30.985
119	70	−13.552	−11.742	70.000	−13.552	−11.742	0.235	0.267	30.890
120	70	−11.795	−12.883	70.000	−11.795	−12.883	0.236	0.262	30.955
121	70	−10.000	−13.962	70.000	−10.000	−13.962	0.236	0.258	31.013
122	70	−8.168	−14.977	70.000	−8.168	−14.977	0.237	0.254	31.128
123	70	−6.302	−15.928	70.000	−6.302	−15.928	0.239	0.250	31.112
124	70	−4.404	−16.813	70.000	−4.404	−16.813	0.240	0.248	31.150
125	70	−2.476	−17.631	70.000	−2.476	−17.631	0.242	0.244	31.318
126	70	−0.521	−18.382	70.000	−0.521	−18.382	0.243	0.242	31.338
127	70	1.459	−19.063	70.000	1.459	−19.063	0.245	0.239	31.575
128	70	3.462	−19.676	70.000	3.462	−19.676	0.247	0.235	31.295
129	70	5.485	−20.218	70.000	5.485	−20.218	0.249	0.233	31.540
130	70	7.525	−20.689	70.000	7.525	−20.689	0.252	0.230	31.533
131	70	9.581	−21.088	70.000	9.581	−21.089	0.254	0.229	31.686
132	70	11.650	−21.416	70.000	11.650	−21.416	0.258	0.228	31.798
133	70	13.728	−21.671	70.000	13.728	−21.671	0.261	0.226	31.699
134	70	15.815	−21.854	70.000	15.815	−21.854	0.264	0.226	31.814
135	70	17.906	−21.963	70.000	17.906	−21.963	0.268	0.225	32.178
136	70	20.000	−22.000	70.000	20.000	−22.000	0.270	0.223	32.179
137	70	22.094	−21.963	70.000	22.094	−21.963	0.274	0.223	32.531
138	70	24.185	−21.854	70.000	24.185	−21.854	0.278	0.223	32.236
139	70	26.272	−21.671	70.000	26.272	−21.671	0.282	0.222	32.348
140	70	28.350	−21.416	70.000	28.350	−21.416	0.287	0.222	32.621
141	70	30.419	−21.088	70.000	30.419	−21.089	0.292	0.223	32.669
142	70	32.475	−20.689	70.000	32.475	−20.689	0.295	0.221	32.596
143	70	34.515	−20.218	70.000	34.515	−20.218	0.301	0.223	32.921

	Intended CIELAB coordinates			Rendered CIELAB coordinates			Measured xyY coordinates		
Color	L*	a*	b*	L*	a*	b*	x	y	Y
144	70	36.538	−19.676	70.000	36.538	−19.676	0.305	0.223	32.854
145	70	38.541	−19.063	70.000	38.541	−19.063	0.311	0.224	32.944
146	70	40.521	−18.382	70.000	40.521	−18.382	0.315	0.224	33.003
147	70	42.476	−17.631	70.000	42.476	−17.631	0.321	0.225	33.193
148	70	44.404	−16.813	70.000	44.404	−16.813	0.327	0.227	33.344
149	70	46.302	−15.928	70.000	46.302	−15.928	0.333	0.228	33.543
150	70	48.168	−14.977	70.000	48.168	−14.977	0.339	0.230	33.561
151	70	50.000	−13.962	70.000	50.000	−13.962	0.345	0.232	33.509
152	70	51.795	−12.883	70.000	51.795	−12.883	0.352	0.234	33.769
153	70	53.552	−11.742	70.000	53.552	−11.742	0.355	0.234	33.458
154	70	55.267	−10.541	70.000	55.267	−10.541	0.360	0.236	32.979
155	70	56.940	−9.281	70.000	56.940	−9.281	0.363	0.236	32.415
156	70	58.567	−7.963	70.000	58.567	−7.963	0.366	0.237	31.910
157	70	60.148	−6.589	70.000	60.148	−6.589	0.370	0.237	31.386
158	70	61.680	−5.160	70.000	61.680	−5.160	0.374	0.239	30.824
159	70	63.160	−3.680	69.651	62.606	−4.234	0.378	0.240	30.336
160	70	64.589	−2.148	69.204	63.353	−3.409	0.384	0.241	29.642
161	70	65.963	−0.567	68.763	64.081	−2.525	0.389	0.243	29.187
162	70	67.281	1.060	68.327	64.789	−1.583	0.392	0.244	28.945
163	70	68.541	2.733	67.898	65.475	−0.583	0.399	0.248	28.406
164	70	69.742	4.448	67.477	66.136	0.476	0.405	0.250	27.815
165	70	70.883	6.205	67.065	66.770	1.594	0.411	0.252	27.362
166	70	71.962	8.000	66.664	67.376	2.771	0.417	0.254	26.965
167	70	72.977	9.832	66.275	67.949	4.006	0.423	0.257	26.599
168	70	73.928	11.698	65.900	68.490	5.299	0.431	0.260	26.149
169	70	74.813	13.596	65.538	68.994	6.651	0.437	0.264	25.817
170	70	75.631	15.524	65.192	69.459	8.061	0.443	0.266	25.472
171	70	76.382	17.479	64.862	69.883	9.528	0.451	0.271	25.190
172	70	77.063	19.459	64.550	70.265	11.052	0.458	0.273	24.838
173	70	77.676	21.462	64.257	70.600	12.633	0.465	0.277	24.622
174	70	78.218	23.485	63.984	70.888	14.270	0.473	0.281	24.300
175	70	78.689	25.525	63.731	71.127	15.961	0.481	0.285	24.069
176	70	79.088	27.581	63.500	71.314	17.706	0.487	0.289	23.889
177	70	79.416	29.650	63.291	71.448	19.504	0.496	0.293	23.756
178	70	79.671	31.728	63.106	71.527	21.353	0.502	0.297	23.534
179	70	79.854	33.815	62.944	71.550	23.251	0.510	0.302	23.316
180	70	79.963	35.906	62.806	71.516	25.196	0.518	0.307	23.217
Background	100	0	0	100	0	0	0.286	0.294	37.530

Table A1. CIELAB values and corresponding CIE xyY values for stimulus color values used in experiments. The xyY values were obtained by measuring light emission spectrum using a spectroradiometer (PR-655 SpectraScan, Photo Research). Unit of Y is candelas per square meter.