Why some colors appear more memorable than others: A model combining categories and particulars in color working memory

Gi-Yeul Bae*,^, Maria Olkkonen#, Sarah R. Allred%, & Jonathan I. Flombaum^

*Center for Mind and Brain, University of California, Davis
#Department of Psychology, University of Pennsylvania
%Department of Psychology, Rutgers—The State University of New Jersey
^Department of Psychological and Brain Sciences, Johns Hopkins University

Address correspondence to any author:
GYB: gybae@ucdavis.edu
MO: mariaol@sas.upenn.edu
SRA: srallred@scarletmail.rutgers.edu
JIF: flombaum@jhu.edu

Mailing address for correspondence:
Jonathan Flombaum
JHU / 3400 N. Charles Street
Ames Hall / PBS
Baltimore, MD 21218

Running Head: A model incorporating categories into color working memory
Key words: visual working memory, delayed estimation, color perception, categorization
Word count: 1156 incl. Abstract; 16 Figures.
Draft: March 2015; In Press; *JEP:G*

Abstract (225)

Categorization with basic color terms is an intuitive and universal aspect of color perception. Yet research on visual working memory capacity has largely assumed that only continuous estimates within color space are relevant to memory. As a result, the influence of color categories on working memory remains unknown. We propose a dual content model of color representation in which color matches to objects that are either present (perception) or absent (memory) integrate category representations along with estimates of specific values on a continuous scale (“particulars”). We develop and test the model through four experiments. In a first experiment pair, participants reproduce a color target, both with and without a delay, using a recently influential estimation paradigm. In a second experiment pair, we use standard methods in color perception to identify boundary and focal colors in the stimulus set. The main results are that responses drawn from working memory are significantly biased away from category boundaries and toward category centers. Importantly, the same pattern of results is present without a memory delay. The proposed dual content model parsimoniously explains these results, and it should replace prevailing single content models in studies of visual working memory. More broadly, the model and the results demonstrate how the main consequence of visual working memory maintenance is the amplification of category related biases and stimulus-specific variability that originate in perception.

Introduction

Visually guided behavior requires both perception and working memory. For example, choosing the ripest avocado at the store requires a comparison between avocados experienced in the past and those observable now. Distinguishing between objects that differ on color —or any other basic visual feature— may seem effortless. But like many other tasks in perception and cognition, it is enormously challenging in practice. Because of inherent uncertainty in perception, inescapably noisy neural processing, and the complexity of viewing conditions, even comparing two side-by-side avocados is computationally difficult. Adding memory demands compounds the difficulty.

Despite considerable interest in the role of visual working memory in behaviors such as detecting changes and reproducing remembered features, little contact has been made between research on the perception of basic visual features and research that uses those features to investigate the nature of visual working memory. Here we focus on color, which has received the majority of attention in studies targeting visual working memory. We test three hypotheses: (1) that working memory maintenance exhibits color-specific biases, (2) that biases originate in perception, and (3) that observers functionally use two kinds of color information when matching colors between objects. These are an estimate of hue on a continuous scale —what has been called a “particular” in other contexts (e.g. Huttenlocher et al., 2000)— and a probabilistic category assignment. The results are central for theories of visual working memory, where inferences about memory processing rest on assumptions that are contravened by our hypotheses. More generally, our results demonstrate that visual perception and working memory share a common vocabulary for describing the material properties of surfaces in the world.

Delayed estimation

Recent and influential work in the domain of visual working memory has examined the mechanisms that support detection of object change, detection of object similarity (match), and more generally, the mechanisms involved in the reproduction of features seen in the recent past. Research on visual working memory has typically framed such tasks in the language of estimation; a participant must estimate the feature of an object seen in the past, given noisy inputs, and then compare it with an estimate of what is seen currently. Appropriately, a paradigm called 'delayed estimation' has been devised and proven productive for investigating working memory mechanisms associated with matching (**Figure 1a;** Wilken & Ma, 2004; Zhang & Luck, 2008).

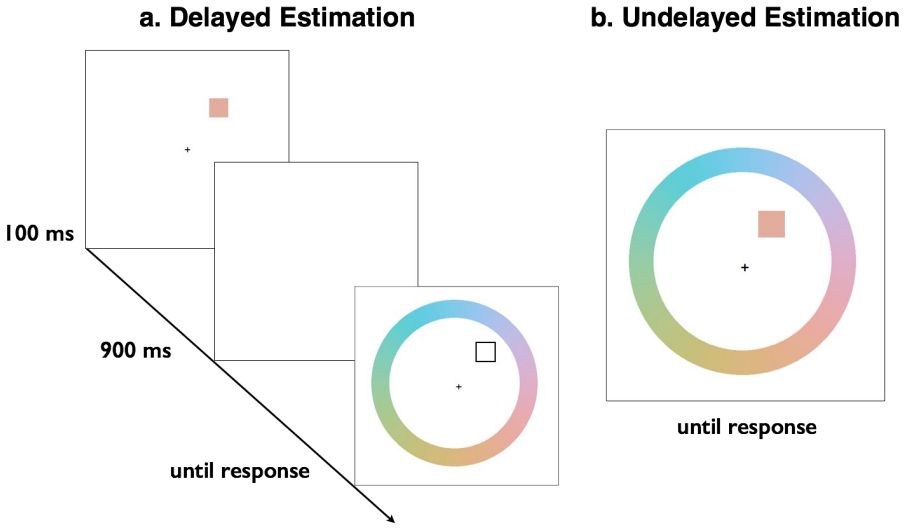

Figure 1. Procedure for color estimation with a delay (a) and without a delay (b).

The majority of studies using this task focus on color working memory —as we will here— and so we describe the basic methodology in that context. In a typical experiment, participants remember the individual hues in a set of circles or squares. After a short delay period, participants report the hue value of one of the study objects on a con-

tinuous response scale, a hue circle (usually with 180 exemplars) comprising all the hues utilized in the study. Response variability —measured as angular deviation between selected and true hues— differs between trials and by condition, motivating inferences concerning the structure of visual working memory (Anderson & Awh, 2012; Bays, Catalao, & Husain, 2009; Bays, Wu, & Husain, 2011; Emrich & Ferber, 2011; 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, 2011; 2009; 2008).

Ultimately, interpreting the results of this and any related paradigm depends on one's expectations about performance *without* memory maintenance (without an enforced memory delay), situations that are constrained more by perception than by the attendant challenges arising from an absent stimulus and working memory maintenance. Fortunately, the same paradigm can be manipulated minimally to investigate this performance. Simply removing the delay period allows one to measure variability of responses when there are no externally enforced memory demands, what we will call 'undelayed estimation' (**Figure 1b**; see also Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014). Practically, undelayed estimation supplies an opportunity to build empirical expectations about performance for use when interpreting effects of memory. And theoretically, it supplies a good methodological opportunity to directly relate perception and working memory in the same task (Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Bae et al., 2014; Gold et al., 2010; Souza, Rerko, & Lin, 2014).

However, we have recently demonstrated that several unwarranted assumptions are built into expectations about undelayed and delayed responses in the literature on vis-

ual working memory (Bae et al., 2014). In our previous study, we investigated responses on a color-specific basis, while also employing what appears to be standard color rendering practice in published reports using delayed estimation. We discovered considerable stimulus-dependent differences in response variability. This is a problem because standard practice with delayed estimation has been to collapse responses across colors, characterizing response variability under the implicit assumption that all colors would elicit more or less similar response distributions.

Further scrutiny of these color-specific response properties led to several additional discoveries. First, color-specific differences correlated across independent observers, demonstrating that they were not random. Second, color-specific differences appeared in undelayed experiments and were correlated with delayed color-specific differences, demonstrating that they originate in perception. Third, color-specific differences were large: in some instances, differences between colors were larger than differences caused by memory load, the primary phenomenon that theories of visual working memory seek to explain. Fourth, color-specific response properties were reliably related to category structure within the set of color samples, suggesting that color categories likely play a role in visual working memory. Finally, we discovered that omitting the calibration and rendering techniques prescribed in research on color perception has likely caused many studies to include rendered colors that differ in meaningful ways from intended ones. Notably, in our study, which specified equiluminant *intended* colors, *rendered* colors differed considerably in luminance.

These results motivate the present study. They suggest that color working memory may not behave uniformly, even with equiluminant stimuli, and that it may rely on encod-

ing of stimulus categories —along with continuous values— to support comparative stimulus judgments.

Indeed, there are good reasons to expect such effects (Allred & Flombaum, 2014). With respect to stimulus-specific response properties, it is known that even equiluminant but different hues will elicit meaningfully different response distributions in a matching context (Witzel & Gegenfurtner, 2013). These effects can originate in perception, as opposed to arising only through an interaction with working memory maintenance (Nemes, Parry, & McKeefry, 2010; Olkkonen & Allred, 2014; Olkkonen, McCarthy, & Allred, 2014). More generally, careful work on color discrimination in psychology and color science has shown that no color space is ever likely to be perceptually uniform (for discussion, see Brainard, 2003; Wyszecki & Stiles, 1982).

In addition, color perception has a salient categorical aspect, at least intuitively. English speakers generally feel comfortable using only 11 terms, often even fewer, to describe a space including a million discriminable shades (Pointer & Attridge, 1997; Linhares, Pinto, & Nascimento, 2008). The development of color terms seems to follow a seemingly universal hierarchical structure suggesting that people using different languages share broadly similar intuitions about color categories (Berlin & Kay, 1969). Additionally, both continuous and categorical representations of colors are present in mammalian brains, although the latter representation (Bird, Berens, Horner, & Franklin, 2014; Brouwer & Heeger, 2013; Koida & Komatsu, 2007) is perhaps less established than the former (e.g. Johnson, Hawken, & Shapley, 2001, 2004; Conway & Tsao, 2006; Horwitz & Hass, 2012).

We therefore sought to use the estimation paradigm to test three related proposals about the contents of color working memory and their relationship to perceptual inputs. We propose that reproducing a *perceived* hue relies on both continuous and categorical representations of hue, that reproducing a *remembered* hue relies on these same two representations, and that the joint reliance on these contents produces stimulus-specific biases. This challenges prevailing assumptions in color working memory research, which include only a continuous hue estimate and no stimulus-specific biases.

Dual contents: continuous estimates ("particulars") and probabilistic categories

To explain how joint continuous and categorical representations can produce reproduction biases, the well-known relationship between spatial working memory and local landmarks serves as an elegant example. Consider an empty piece of paper with a dot on it. If asked to reproduce the dot on another, entirely empty piece of paper, your responses will likely form a cloud —probably a two-dimensional Gaussian— characterized by the uncertainty in your position estimates and noise in your motor machinery. Now consider a case in which the dot is placed in the same place on the paper, but within a larger circle and near its perimeter. Assuming the circle is also on the reproduction paper, your responses over many trials will form a different cloud. None of your responses will cross the perimeter of the circle. The presence of a salient landmark will bias responses. These and related experiments conducted by Huttenlocher and colleagues (2000; Crawford, Huttenlocher, & Hedges, 2006; Duffy, Huttenlocher, Hedges, & Crawford, 2010) demonstrate that spatial working memory relies on both continuous position estimates — in their terms, "particulars"— and categorical descriptions relative to either inductively developed categories, such as distributions of stimuli used during an experiment, or land-

marks, such as “within the circle and near the perimeter.” Combining these contents produces biased reproductions.

We propose that color working memory (and perception) work in much the same way. In the case of the delayed estimation task, each stimulus in a memory sample is represented both by a noisy estimate of a particular hue value on a continuous scale and also by a category label from the set comprising the basic color terms (e.g. blue, green, orange etc.). In our model, the category label is itself assigned probabilistically, so that hues near a category boundary will be assigned to different categories on different occasions. The combination of these two contents will result in biases that differ by stimulus. Colors near the center of categories are unlikely to produce biased estimates, because continuous and categorical estimates align. But colors near boundaries will exhibit large biases in the focal direction of their categories. In the same way that an observer will not place a dot outside a circle when she remembers it as being inside the circle, she should not respond with hues she would label as green to reproduce one she remembers as blue.

To test our proposal, we employ two approaches. In behavioral experiments, we first characterize stimulus-specific response distributions elicited by each of the colors (i.e. 180 colors) in a complete hue circle using delayed and undelayed estimation. In order to establish the relationship between continuous and categorical contents of colors, we independently identify probabilistic category boundaries and focal exemplars using category assignment and focal identification procedures typical in research on color appearance (**Figure 2**; see also Bae et al, 2014; Witzel & Gegenfurtner, 2013). The results of these experiments are reported first. We then describe a computational model designed

to predict empirically obtained response distributions by combining continuous estimates and probabilistic category assignments.

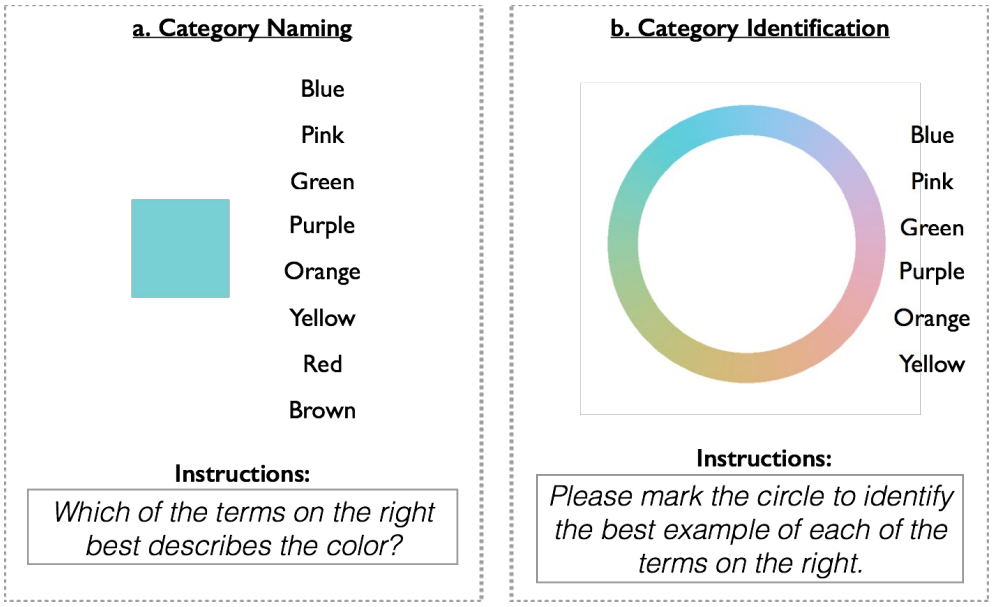

Figure 2. Procedure for Category Naming (a) and Category Identification (b).

Experiments: Categories and Stimulus-Specific Response Properties

The experiments included in this study encompass several goals. The first is to characterize any systematic stimulus-specific properties of matching responses to colors on a hue circle (with constant luminance), using both delayed and undelayed estimation. By 'systematic,' we mean stimulus-specific properties that do not arise randomly, which we diagnose through correlations across independent observers. Toward this end, we identified a circle of 180 equally spaced colors (CIELAB) with a constant luminance that we employ in delayed and undelayed estimation experiments. Both experiments include only a single sample item in each trial, either presented simultaneously with a response wheel (undelayed estimation) or followed by a delay and then a response wheel (delayed estimation; **Figure 1**).

Our second goal is to identify category boundaries and focal colors within the hue circle. We do this using a pair of experiments similar to those common in research on color perception (c.f. Witzel & Gegenfurtner, 2013). One group of participants completed a category naming experiment, in which they indicate which color term best describes each of the 180 hues. Another group of participants completed a category identification experiment, in which they select the best example of each of the basic color terms from the complete hue circle. The third goal is to characterize any reliable relationships between stimulus-specific response properties in the estimation experiments and the category landmarks derived from the category experiments.

Methods

Participants. All participants were Johns Hopkins University undergraduates who received course-related credit in exchange for participation: Delayed estimation, n=3; undelayed estimation, n=8; category naming, n=10; category identification, n = 5. All participants had normal or corrected-to-normal visual acuity and reported normal color vision. Each completed only one of the four experiments. Protocol was approved by the Johns Hopkins University Homewood IRB.

Apparatus. The experiment took place in a dark, sound-attenuated room. There was no light source except for a CRT monitor at a viewing distance of 60 cm, such that the display subtended approximately 39.56° by 25.35° of visual angle.

Stimuli: We chose 180 equally spaced stimuli that only varied in hue in CIELAB space (L*=70, a*=0, b*=0, radius of 38; **Figure 3**). This ring is similar to, but not identical with prevalently used rings in the literature on delayed estimation. We found that more commonly used settings were outside the monitor gamut. RGB values correspond-

ing to the CIELAB coordinates were generated by performing a standard monitor calibration (Brainard, Pelli, & Robson, 2002). In color conversions from device-independent to device-dependent spaces, we used the measured monitor white point of CIE xyY [0.3184, 0.3119, 48.64]. Conversions between color spaces were performed with colorimetric routines implemented in the Psychophysics Toolbox (Brainard, 1997) and radiometer measurements (PR655, PhotoResearch Inc, Chattsworth, CA). Stimuli were always presented on a uniform background that was the center point of the chosen CIELAB hue ring (L*a*b* = [70,0,0]) in order to ensure equal saturation and chromatic contrast with respect to background across hues.

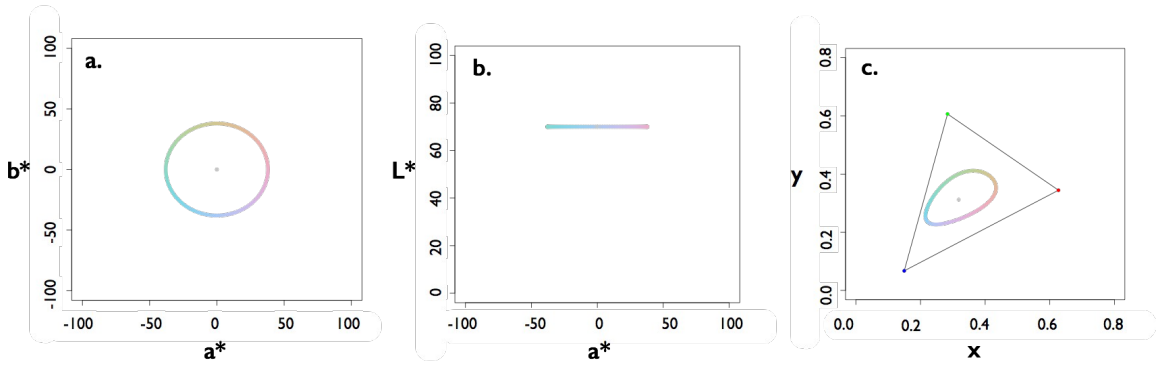

Figure 3. Hue circle used in experiments. a) Hue circle a* and b* coordinates in CIELAB space. b) L* values of all hues, and c) x and y values of hue circle, shown within monitor gamut (triangle; CIE xyY space).

Procedures and analyses: In the **undelayed estimation** experiment, participants made color matches to study stimuli as follows. Each trial began with a white fixation cross (0.5° x 0.5°) displayed in the center of the monitor. After 500 ms, the study stimulus (a 2° x 2° colored square) appeared at one of eight possible positions (4.5° from fixation) together with the matching wheel (8.2° radius and 2° thick) that surrounded the space in which study stimuli could appear (**Figure 1**). The matching wheel consisted of all 180 stimuli, organized as a hue circle. On each trial, the matching wheel was randomly ro-

tated to prevent position-color associations. The task was to click the color on the matching wheel that was perceived as most similar to the study color. Both the study stimulus and the matching wheel remained on the screen until response, at which time a black line superimposed on the matching color indicated the clicked position.

The undelayed experiment included eight participants. This is because we sought approximately 60 measurements across the experiment in response to each individual hue, a typical number of measurements obtained in delayed estimation experiments within a condition (see e.g. Bays et al., 2009). Because obtaining 60 measurements per color, per participant in this case would have produced an excessively long experiment, we divided the 180 study colors into two sets of 90 colors. Arbitrarily setting one of the colors as number one and then moving around the circle until color 180, the two sets were made by grouping odd and even colors together, so that colors within each set formed a color wheel of 90 exemplars with an equal spacing of four degrees (instead of two) between hues. Half of the participants were presented only odd exemplars as study stimuli, and the other half were presented only even ones. All participants, however, encountered the *entire* color wheel for response selection. Each participant completed four blocks of 360 trials, totaling 1440 trials. Within a block, each color appeared four times in a random order, producing 16 measurements per color per participant, and 64 observations per color overall.

The delayed estimation task was identical to undelayed estimation, with the following exceptions. Most importantly, the study color remained on the screen for 100 ms, and then disappeared from view for 900 ms. Only after the delay did the matching wheel

appear (**Figure 1**). Participants were asked to remember the presented color as precisely as possible.

This experiment included three participants, again, in order to obtain approximately 60 measurements per color across the experiment. In this case each participant completed ten blocks of 360 trials, totaling 3600 trials. In each block, each of the 180 colors was presented twice, in a random order, resulting in 20 observations per color and participant, and 60 observations per color overall. The ten blocks were distributed over three consecutive days (with four blocks on the last day). This experiment was actually run before the undelayed experiment. We found it difficult to find participants that would reliably return to the lab over three consecutive days, which led us to the design of the undelayed experiment with more participants in shorter sessions, but producing approximately the same number of observations per color.

We used a mixture model comprised of a von Mises and a uniform distribution to analyze the results of each estimation experiment (Zhang & Luck, 2008). The model includes three free parameters: the proportion of target-based responses (β , $0 \leq \beta \leq 1$), bias (μ, $-\pi \leq \mu \leq +\pi$), and the concentration parameter of the von Mises distribution (κ, $0 \leq \kappa \leq 700$), which is the inverse variance and is often called 'precision.' Larger κ values reflect less dispersed distributions. In the remainder of the paper we refer to precision of color matches. The complete model is as follows:

$$p(\widetilde{X} \vee S_i) = \beta \phi(S_i + \mu_i, \kappa_i) + (1-\beta)\frac{1}{2\pi}(1)$$

$\widetilde{X}$ denotes the angular position of an estimated hue to a particular target stimulus, S_i , so that $p(\widetilde{X}|S_i)$ is the probability of a response sampled by an observer given the target color. Note that we use the subscript *i* to denote individual stimulus values,

emphasizing the fact that we fit the model to each individual color stimulus with its own parameters. The first term in the model denotes the von Mises density (ϕ , circular normal distribution) described by the two free parameters –-μ and κ–- multiplied by a mixture coefficient, β . By fitting μ along with κ we are able to determine whether individual colors elicit differentially biased distributions, that is, whether they elicit response distributions not centered on the correct sample color.

The second term of the mixture model denotes the uniform density attributed to guessing; thus, (1- β) is typically interpreted as the guessing rate, reflecting trials with encoding or maintenance failures.

All model fitting was performed by maximum likelihood inference. Parameters were initialized to multiple starting values in an attempt to avoid local maxima. Importantly, we fit the model to each study color individually.

The **category naming** experiment (**Figure 2a**) was designed to identify boundaries on the hue circle. On each trial, a square (2° x 2°) filled with one of the 180 study colors was presented at the center of the screen. On the right side of the square, the chromatic color terms comprising Berlin and Kay's eight basic color categories were presented vertically (Berlin & Kay, 1969: 'Red', 'Brown', 'Orange', 'Yellow', 'Green', 'Blue', 'Purple', and 'Pink'). Participants selected the color term that most closely described the study color. The study square and color terms remained on the screen until a response. Each participant completed six trials for each of the 180 study colors, presented in random order, for a total of 1080 trials per participant. We included ten participants. Our previous study using this method included eight observers (Bae et al., 2014). We in-

cluded ten here using a slightly shorter design per participant, intending to obtain the same number of observations nonetheless.

The **category identification** experiment (**Figure 2b)** was designed to identify focal exemplars for each of the basic color terms. Participants selected the study color that best exemplified each color category as follows. On each trial, the matching wheel appeared in the center of the screen together with the basic color terms to the right of the wheel. Participants clicked on the matching wheel to indicate the best example of each color term. A black line appeared after the mouse click at each location to prevent multiple responses for the same color term. The matching wheel randomly rotated on each trial to prevent any association between color and position.

The terms 'Red' and 'Brown' were excluded because very few study colors were identified with these terms in the color naming experiment (See **Figure 4a**). This is likely due to the saturation level and luminance selected for the hue circle. Thus, participants made 6 responses —one for each color term— per trial, and they each completed 30 trials, resulting in 30 responses for each color term per participant.

The purpose of the category experiments pair was to derive distributions describing category membership for the six basic color terms. By collapsing responses across participants (within each experiment) we obtained two empirical frequencies for each color describing the probability that it was assigned a particular name, as the best name for that color in the category naming experiment, or as the best example of a given name in the category identification experiment.

Through the category naming experiment we operationalized color boundaries as colors that were equally likely to be named with adjacent category terms. To interpret the

results of the identification experiment, we fit six von Mises distributions, one to the responses elicited by each of the six color terms. The means of these distributions were considered estimates of focal exemplars.

Results

Categorization experiments. **Figure 4a** plots the results of the category naming experiment. Most colors were assigned a single term repeatedly. But some were likely to elicit more than a single response, and a handful received two adjacent terms with equal probability. These can be thought of as category boundaries. (Note that 'red' and 'brown' were rarely attributed to any of the samples). This pattern of response is similar to that in previous category experiments (e.g. Boynton & Olson, 1990; Sturges & Whitfield, 1997).

Figure 4b plots the frequency with which each color was selected as the best example of any of the six colors terms, along with best-fit von Mises densities. If all samples within a category were perceived as equally good exemplars of the categories, these frequencies would have been relatively uniform, much like the distributions in the naming experiment. But distributions in the identification experiment were clearly peaked, reflecting agreement among observers about best exemplars. We treat the peaks of these distributions, operationalized as the mean of a von Mises distribution, as focal colors in the analyses reported below.

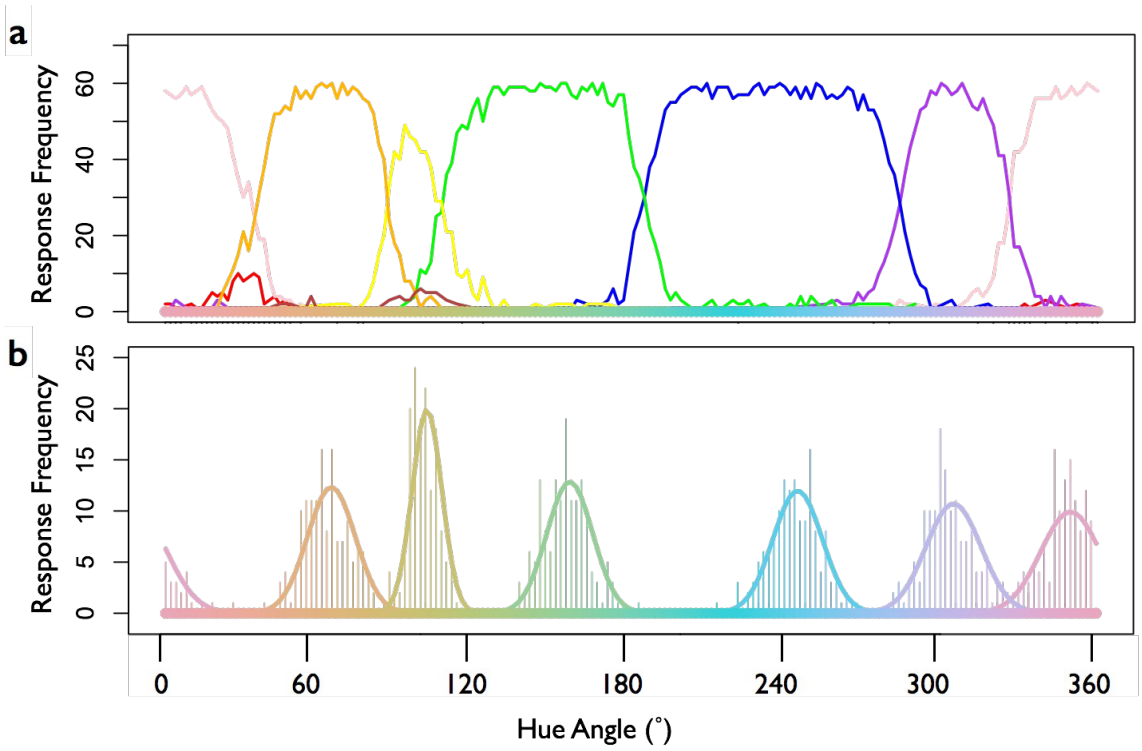

Figure 4. Results of the *category naming* (a) and *category identification* (b) experiments. In (a), the response frequency with which each color term was used is shown for each of the 180 hues, and in (b) the response frequency with which each hue was labeled as the best exemplar for each color term. von Mises distributions fit to the response frequencies are shown in (b) as well.

The qualitative take away from this pair of experiments is that many colors were best described by a single color term, but not all of those colors were equally good examples of their respective terms. And some colors were neither good examples nor well characterized by a single term.

Precision and bias in the estimation experiments. As expected, both estimation experiments produced uniformly low guessing rates ($1 - \beta$; no-delay average: 0.8%; delay average: 2.1%). We could therefore use model-free measures of dispersion and bias to characterize stimulus-specific response characteristics. Indeed, all the results reported are similar when viewed in this way. But we employ the model-based parameters to accommodate the broader project of supplying a model of working memory contents that can be used to analyze responses in situations with higher expected guessing rates.

Figure 5 shows response distributions with and without delay for two target examples. It is meant to illustrate three broadly applicable points. First, distributions to different hues were not equally dispersed. In these cases, responses to the blue example were more dispersed than to the yellow one (high kappa values correspond to low dispersion). Second, responses were biased; the average of a response distribution (represented by the dotted lines in the figure) was usually not the veridical study hue (triangles in the figure). The degree of bias, which was computed as the distance between the mean response and the study color, was also stimulus-specific, with some study hues showing more bias than other study hues, and as in the two examples shown, biases were not in the same angular direction for all hues.

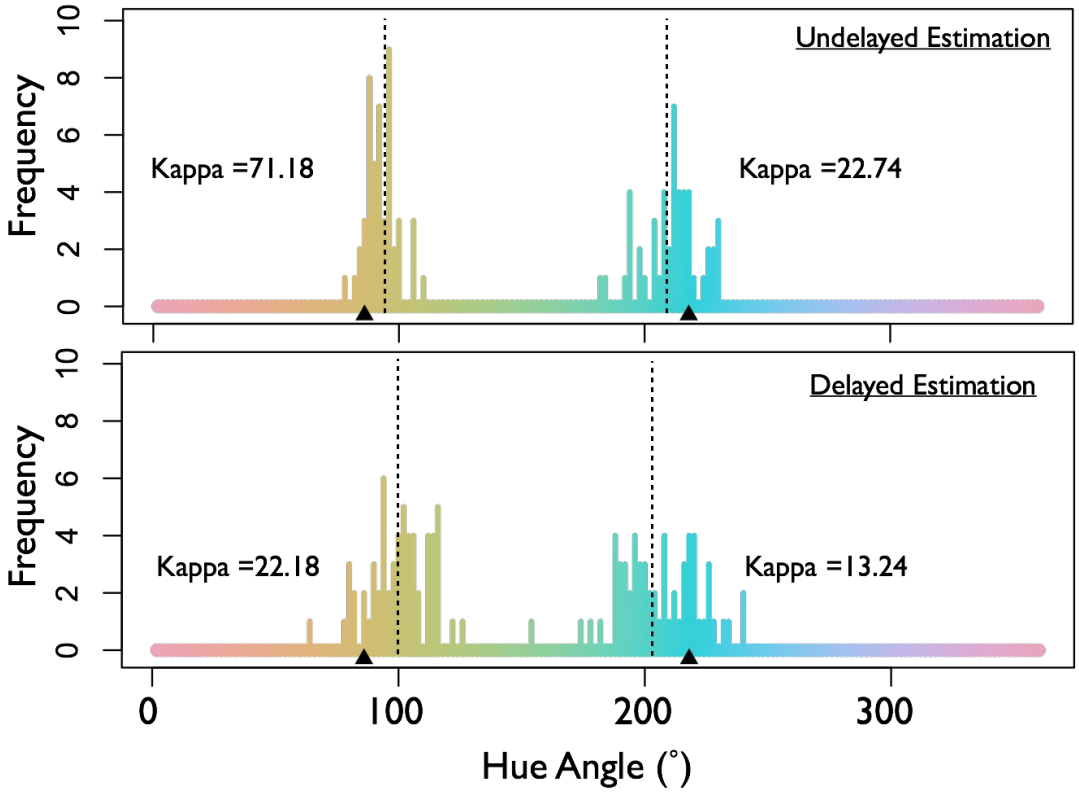

Figure 5. Response distributions for two study hues, in undelayed (top) and delayed (bottom) estimation. Triangles on the graphs designate the true hue values, and dotted lines identify the distribution means.

These patterns were evident in the dataset as a whole. **Figure 6** makes the point theory-free: we plot the frequency with which each color was selected as a response across the whole of each experiment. If hues generally elicited similar and unbiased response distributions, these overall distributions should be close to uniform (each color was the target equally often). The distributions clearly are not uniform. **Figures 7** and **8** plot precision and bias estimates for each color with and without delay. Again, there was considerable color-by-color variability. Crucially, color-by-color κ estimates were significantly correlated in two out of three pairwise observer-relationships, and the third correlation was marginally significant ($t(178) = 4.51$, $r = 0.32$, $p < .01$; $t(178) = 2.81$, $r = 0.21$, $p < 0.01$; $t(178) = 1.81$, $r = 0.13$, $p = 0.07$). μ estimates were also significantly correlated across all pairwise comparisons ($t(178) = 12.88$, $r = 0.70$, $p < 0.001$; $t(178) = 6.80$, $r = 0.45$, $p < 0.001$; $t(178) = 11.18$, $r = 0.64$, $p < 0.001$).

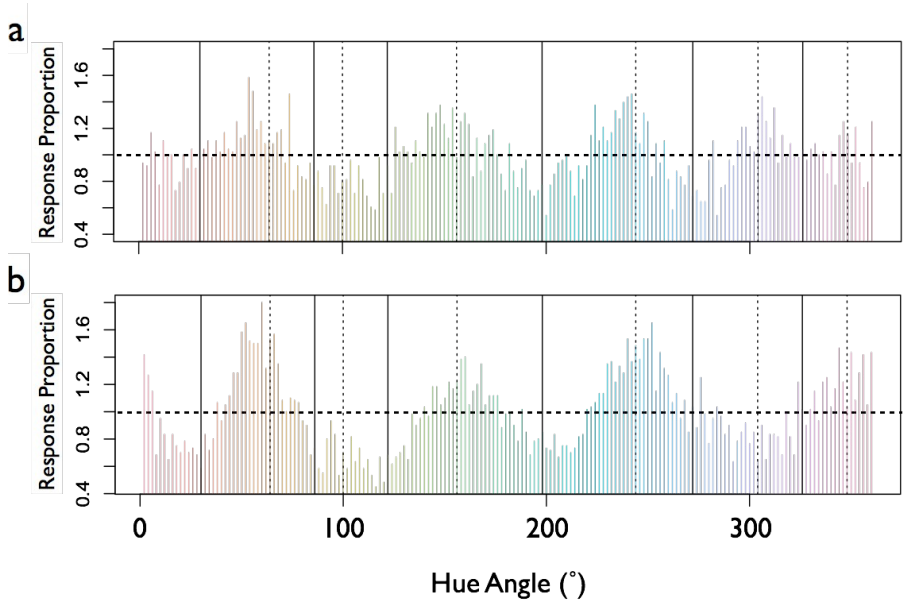

Figure 6. Normalized response frequencies for each individual hue across all observers in the undelayed (a) and delayed (b) estimation experiments. Each hue appeared as a target with equal frequency (60 or 64 times depending on the experiment). A response proportion of one (the dashed horizontal lines) thus indicates that a matching hue was selected with the same frequency it appeared as a study hue. Vertical dotted lines indicate focal colors, and vertical solid lines indicate border colors (see methods).

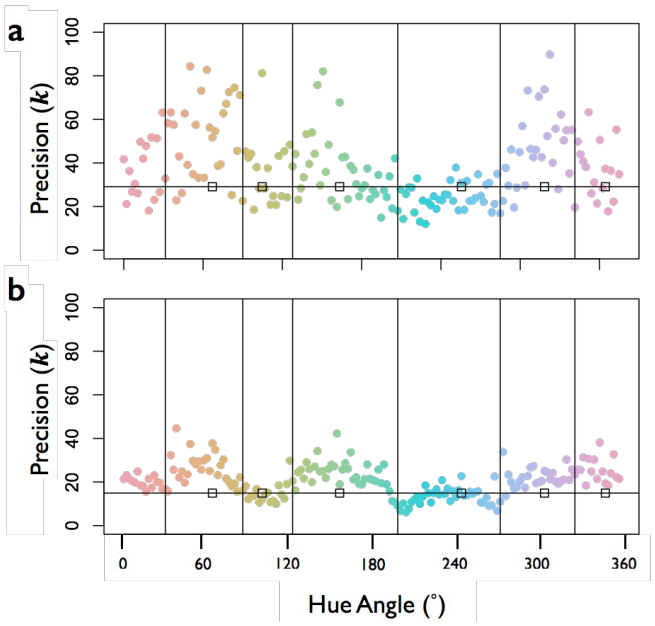

Figure 7. Hue-specific precision estimates (κ) in undelayed (a) and delayed (b) estimation. Vertical dotted lines indicate focal colors, and vertical solid lines indicate border colors (see methods). The solid horizontal line in each figure is the κ value obtained when the mixture model was fit to responses collapsed across hues.

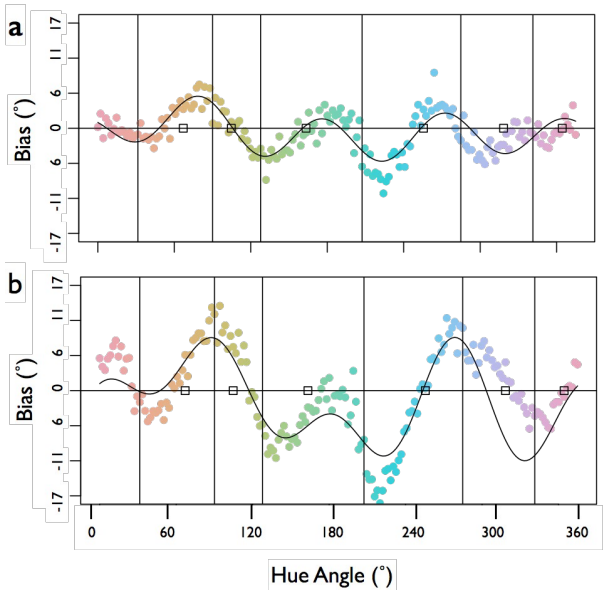

Figure 8. Hue-specific bias estimates, the difference in degrees between each hue value and the estimated mean (μ**)** of the response distribution in trials in which the hue was the target, for undelayed (a) and delayed (b) estimation. Positive values indicate leftward bias and negative values indicate rightward. Vertical dotted lines indicate focal colors, and vertical solid lines indicate border colors (see methods). The black smoothing curves are superimposed to emphasize the pattern of bias estimates.

More importantly —for the purpose of characterizing the contents and mechanisms of color working memory— delayed κ estimates were significantly smaller (more variable) than undelayed estimates (Delayed κ = 20.56; Undelayed κ = 38.39; $t(179) = 12.26$, $p < 0.001$), and estimates of μ were larger (more biased; mean μ: 5.61 vs. 2.84, $t(179) = 11.204$, $p < 0.001$). Additionally, patterns of stimulus-specific response correlated significantly between delayed and undelayed estimation experiments for both κ ($t(178) = 5.50$, $r = 0.38$, $p < 0.001$) and μ ($t(178) = 10.82$, $r = 0.63$, $p < 0.001$), as shown in **Figure 9**. Recall that these experiments included distinct groups of participants.

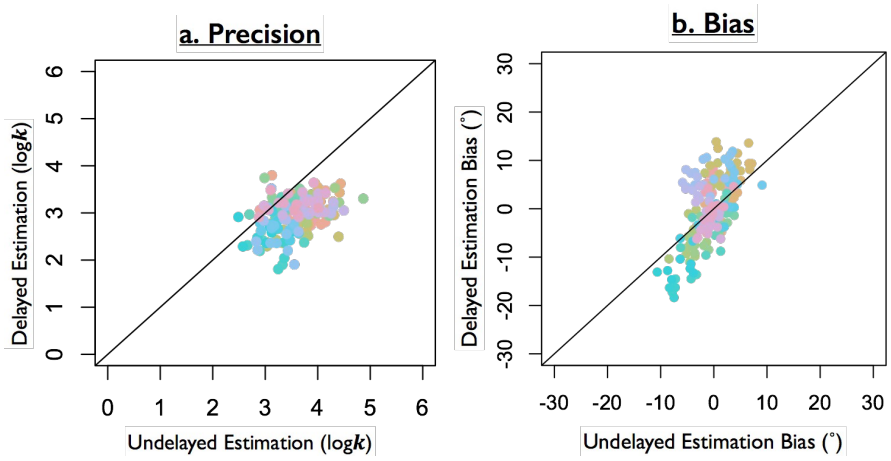

Figure 9. Correlations of hue-specific precision (a) and bias (b) estimates between undelayed and delayed estimation.

Having established reliable patterns of color-specific responses (with and without a delay) we considered whether these properties contained interpretable variability. A number of systematic effects become qualitatively apparent in the patterns of hue-specific performance. First, some regions of the hue circle acted as attractors. Hues on either side of these regions showed oppositely directed biases *towards* the attractor region. Second, some regions of the color space seemed to repel responses. Responses to their surrounding colors were biased away from these regions.

We sought to determine whether these patterns relate to category structure within the set of colors. First, we computed the angular distance between each study color and the nearest focal color on the wheel. There were 23 unique distances, which we used as bins (each with a 2 degree width). We then correlated distance with the average κ and absolute bias of each bin. If the properties of response distributions are independent from the category structure of the color space, then there should be no effects on bias and precision of distance from the focal colors on the wheel. However, all four correlations were significant (**Figure 10**, Undelayed: κ $t(21) = -2.79$, $p < 0.05$, $r = -0.52$; Bias $t(21) = 2.97$, $p < 0.01$, $r = 0.54$; Delayed: κ $t(21) = -7.04$, $p < 0.001$, $r = -0.84$; Bias $t(21) = 6.85$, $p < 0.001$, $r = 0.83$).

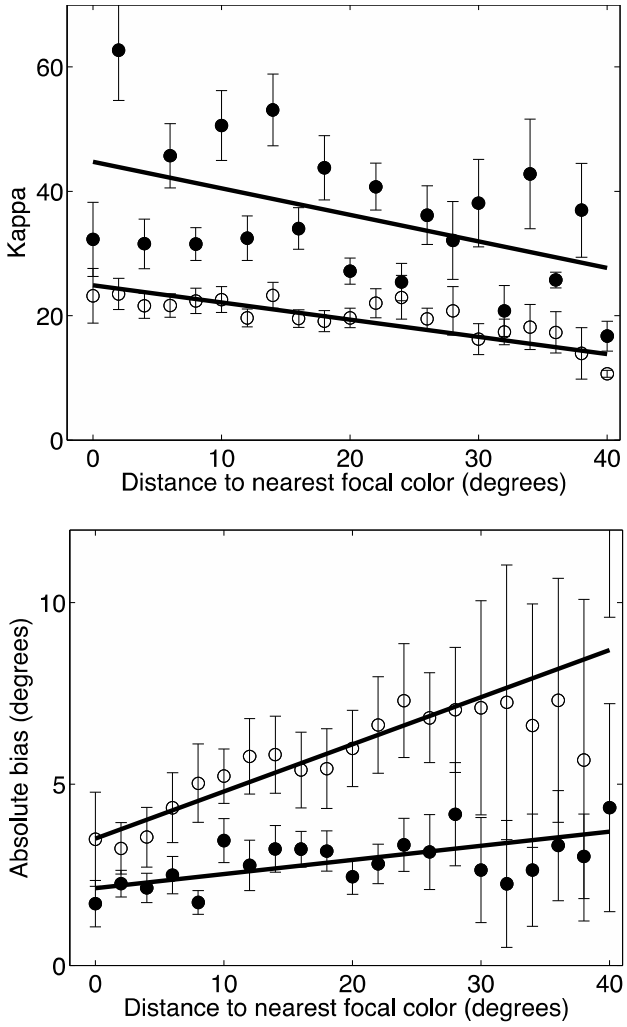

Figure 10. Relationship between kappa (top) and bias (bottom) and distances to focal colors in delayed (empty symbols) and undelayed (filled) estimation. Error bars are s.e.m of parameters in each distance bin.

Discussion

With a hue circle comprising stimuli of equal luminance, equal saturation, and equal chromatic contrast with the background, we discovered systematic stimulus-specific response variability in estimation tasks. Surprisingly, these patterns were evident in estimation without a delay. The precision and degree of bias for a given hue were predicted by its relative position within a color category, that is, its distance from focal colors and category boundaries. The results are consistent with our hypothesis that estimation responses rely on dual contents, including a noisy, continuous estimate of a particular

hue value, and a category assignment. The model presented in the next section is meant to further support this point; we reserve most discussion of dual contents for the time being.

Even without a model, however, it is readily apparent that patterns of stimulus-specific responses that depend on imposed working memory maintenance have the same basic properties as those less reliant on maintenance (in the experiment without a delay). Stimuli that exhibit biases without a delay exhibit even greater biases with a delay. This is consistent with part of our hypothesis, that working memory maintenance amplifies biases in estimation responses that originate prior to maintenance.

This is relevant for considering the role of verbal rehearsal, which is sometimes thought to be involved in memory experiments for colored stimuli. The correlation between biases in the delayed and undelayed conditions indicates that if category rehearsal plays a role in the delayed condition, it also plays a role in the undelayed condition. Yet it is odd to think of explicit rehearsal playing a role without a delay. Thus stimulus-specific bias and precision cannot be attributed to a verbal rehearsal process that is solely present when a maintenance period is imposed.

Since including a delay appears to amplify biases present without a delay, an important question centers on the origin of the biases without a delay. Two classes of cause suggest themselves. First, it is possible that the categorical bias observed in undelayed estimation is caused by working memory. Estimation without a delay may involve memory to some extent. For example, if an observer saccades between targets and match positions, working memory is presumably involved in stimulus maintenance during the saccade (Hollingworth, Matsukura, & Luck, 2013; Schneegans, et al., 2014).

On such a view, the difference between the two conditions would presumably be explained by greater maintenance demands with a delay (compared to without). Note that with such a theory, biases caused by memory would still need to be related to color categories in order to produce the specific patterns of effects we have observed.

The second possibility is that stimulus-specific biases are caused by processing that originates in perception. What we mean by this is that the visual system may spontaneously assign category labels to signals, and as we have proposed, that these labels interact with encoded hue content to produce bias during response. On this view, bias with a delay is greater than without because increased uncertainty in metric signals lead to a greater impact of category encoding. If memory for hue value is noisier than perception of hue value —as it is in all theories that we are aware of— then category encoding should produce greater bias when it interacts with noisier metric signals. Our formal model makes this clear, and we discuss it further in the General Discussion.

These two possibilities are not mutually exclusive. What is crucial from our perspective is that both require that color categories be assigned to signals at some stage in order to impact responses. Below, we advance a formal version of the second possibility —where category encoding (and thus bias) emerges in a categorical perceptual channel. But a theory in which category encoding occurs in working memory would also be importantly different from prevailing theories, which assume that a hue is described only as a point in a continuous space.

Two conclusions are therefore warranted based on the empirical findings reported. First, working memory contents include category labels, though it remains unclear if they are assigned during perception or later. Second, the effects of a *minimal increase* in work-

ing memory demands—those differentiating a stimulus that can be re-inspected, even kept in view during response, and a stimulus that is absent when a response needs to be made—is an amplification of category-related stimulus-specific biases.

Categories and Particulars: A Dual Content Model of Color Working Memory

The objective of the dual content model that we propose below is twofold. Theoretically, the model is an implementation to test the hypothesis that color estimation combines a continuous value with a probabilistic category assignment. Towards this end, we propose a probabilistic model that combines these two sources of information, and we compare it to a model that only utilizes a continuous value (the prevailing approach in the working memory literature), and a model that only utilizes a probabilistic category assignment.

Practically, the objective of our modeling effort is to supply a revised model for use in studies of working memory, one that efficiently predicts stimulus-specific response variability and provides transparent parameters for building theories of working memory limits. To demonstrate the presence of stimulus-specific bias and precision in the experimental section above, we fit a three-parameter mixture model to each of the 180 individual hues on our color circle. But this is an inefficient approach. It ultimately includes many free parameters, requires long experiments, and it is not obvious how it can be used to build theories of working memory limits, or to engage in rigorous comparisons of those theories. Fortunately, the significant relationships we observed between stimulus-specific responses and category-landmarks suggest a systematic cause —or at least, a reliable predictor— of inter-stimulus response differences. We thus sought to leverage the

results of the category experiments to build a more compact model, one that could replace the prevailing mixture model and eventually accommodate further alterations in the service of better understanding visual working memory under a variety of experimental conditions.

CATMET: A dual content model

In broad strokes, the model receives a study hue permuted by noise, termed a noisy sample, and then it estimates the study hue most likely to have caused the noisy sample. Crucially, noisy samples are encoded in two ways: The model infers a distribution of stimulus hues likely to have caused the noisy sample, what we will term a metric distribution, corresponding to an encoding of “particulars” in the terms of Huttenlocher and colleagues (2000). And the model assigns a category descriptor to the noisy sample, on the basis of which it produces a distribution of hues likely to generate that category descriptor, what we will term a category distribution. The initial assignment of category is also noisy, with probabilities derived empirically from the category experiment pair. Thus, identical study stimuli can be assigned to different color categories on different encounters. Both metric and category distributions are in continuous color space. The main difference between this model and prevailing models is in the implementation of a categorical encoding. In prevailing models, the stimulus hue is permuted by noise, and the noisy sample that results is encoded as metric value (a “particular”); CATMET also encodes it as a member of a coarse category. This is the dual content component of the model.

The model then produces an estimate of the stimulus hue by sampling from a joint distribution, achieved by multiplying the metric and category distributions. These steps

are laid out schematically in **Figure 11.** To summarize, the model involves three stages. In the first, it encodes a sample through a high-resolution metric channel, and also through a coarse, category channel. In the second step it generates distributions of continuous values likely to have produced the content encoded through each channel. And it finally combines those distributions to arrive at a single distribution of probable study stimuli. Model details follow.

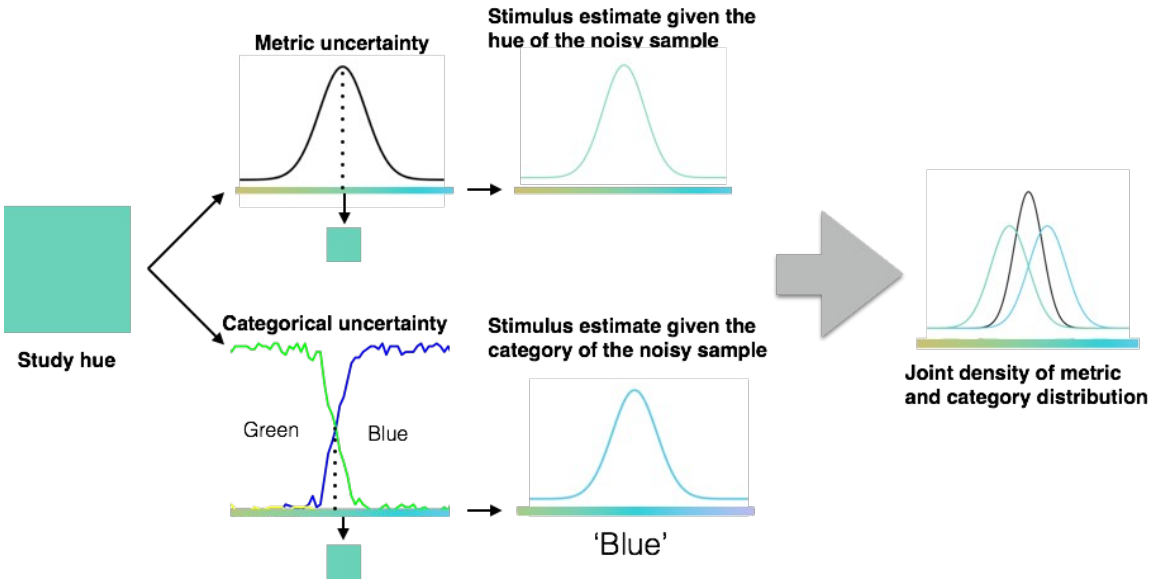

Figure 11. Schematic depiction of the dual content (CATMET) model. The study stimulus leads to a noisy sample received by the observer, and encoded through two channels. In the top panel the particular hue of the sample is encoded, leading to a distribution of study hues likely to have produced that sample, the 'metric distribution'. In the bottom panels, the sample is encoded through a coarse categorical channel: a category is assigned to the sample, and then a distribution of hues likely to produce that category is generated, a category distribution. Finally, the distributions are combined to produce a joint distribution of likely study hues (shown in black).

Step 1: The noisy sample: As in most perceptual models, we assume that the incoming sensory signal is noisy. Thus on each run (simulation) of the CATMET model, the study hue will be encoded based on a noisy sample. Here we describe how we generate a noisy sample on each run given a study hue.

The probability of the model receiving a particular noisy sample, denoted $\hat{S}$, given a study hue, S_i, is determined by a von Mises distribution with two parameters, μ_i and κ_i:

$$S \vee ¿\, S_i \quad \hat{¿} \quad ¿ \quad p\,¿ \tag{2}$$

We use the subscript *i* to denote stimulus specific μ and κ values, the values that apply to the i^{th} exemplar on the color wheel. But our goal with this model is to characterize stimulus-specific differences *without* stimulus-specific parameters. Indeed, we assume that the metric distribution is unbiased, and instead, that observed biases result from the interaction with a category distribution. Accordingly, we assume *unbiased* sensory signals, endowing each study hue with μ equal to zero, and we use a single κ value for all stimuli. With stimulus-independent von Mises parameters, Equation 2 can be rewritten as follows:

$$S \vee ¿\, S_i \quad \hat{¿} \quad ¿ \quad p\,¿ \tag{3}$$

Rather than fit κ within the model, we choose an easily obtainable estimate. In modeling the results of the experiment without a delay, we obtain κ by fitting the prevailing mixture model (Zhang & Luck, 2008; Equation 1) to the responses from that experiment across all colors simultaneously, a value of 29.10 (which is the value of the horizontal line in **Figure 7**). Our goal here is to quickly obtain a reasonable, color-neutral estimate in order to see the behavior that arises from the model generally. We discuss this further after presenting the model results. The same is done when we model the experiment with a delay; we fit the mixture model in Equation 1 across all responses and

colors in that experiment, obtaining a single κ value of 14.89 to utilize in testing the model.

Step 2. Assigning noisy samples to categories. Unlike extant models, the CATMET model assigns a category label to any noisy sample received. Simply put, it labels the sample with one of a set of basic color terms. For most samples this should be a straightforward and uncontroversial process; as shown in Figure 4a, most individual hues were reliably named with only one basic color term in the Category Naming experiment. But some colors received two adjacent labels, such as 'Green' and 'Blue' with high probabilities. On each simulation, we therefore assign labels to samples probabilistically, as follows.

First, we derive distributions of boundary colors by identifying the colors in the category naming experiment that were closest to receiving two adjacent color names with equal frequency. We then set these values as border colors. To implement the assumption of noisy borders, we use von Mises distributions, centered on each border color, and with the color-independent κ values from Step 1.

We now have six border colors, which we denote as B_j . On each model simulation, a discrete border between each category *j* and *j*+1 is selected randomly by drawing a color from the probabilistic distributions defined by the parameters μ_j and κ_B as described above. (The 'B' subscript here is just meant to denote the fact that we use the same precision value for all boarders).

$$B_j = \phi(\mu_j, \kappa_B) \quad (4)$$

With the sampled border colors, B_j , the category of a target color is determined by the relative position of a target color and each border color (Maddox & Ashby, 1993).

Suppose a target color is S_i , and there are six alternative categories.

if the angular position of S_i (5)

$> B_1$ and $\leq B_2$; then $S_i \in$ category 1
$> B_2$ and $\leq B_3$; then $S_i \in$ category 2
$> B_3$ and $\leq B_4$; then $S_i \in$ category 3
$> B_4$ and $\leq B_5$; then $S_i \in$ category 4
$> B_5$ and $\leq B_6$; then $S_i \in$ category 5
$> B_6$ or $\leq B_1$; then $S_i \in$ category 6

Straightforwardly, if the angular position of a target color S_i is between B_j and B_{j+1}, the model determines that the target color is a member of category *j*. By using noisy samples and noisy borders, a single stimulus (especially one near a border) will be assigned to different categories on different simulations. Thus on each model simulation the noisy sample $\hat{S}$ that is used (in Step 4, below) to generate the metric distribution of likely stimulus hues, is also assigned a category which we denote $\hat{C}$.

Step 3. Probability of study hues given a category: With a category label assigned, the model now engages in a process to ensure that a response generated will be a good example of the category assigned. The coarse encoding of category leads the model to prefer responses that are better examples of a particular category. To do this, the model calculates the probability that each study-hue would have produced the *category* description encoded in Step 2, Formally, the probability of drawing a hue from a distribution of hues that are likely to belong to category $\hat{C}$:

$$p\left(\widetilde{X}_C \middle| \hat{C} ¿ = \phi\left(\widetilde{X}_C \vee \mu_c, \kappa_c\right)\right. \quad (6)$$

We denote this distribution $\widetilde{X}_C, \in order$ to distinguish it from the distribution reflecting the probability of study hues obtained on the basis of a sample's encoded metric value in extant models (and also in Step 4 upcoming, and denoted $\widetilde{X}_S$). μ_c and κ_c are parameters describing a distribution of hue values in category C. We estimate their

values using the data from the category experiment pair.

To do so, we combine the data from both category experiments into a frequency distribution, as follows. The raw data on each trial of those experiments —a total of 10800 color naming and 150 category identification trials— are a color value and color term that were associated by a participant. On the basis of each experiment, we thus compute the probability of each of 180 colors being associated with a given color term. Since for each color we now have two association probabilities (one from each of the category experiments), we average the probabilities, producing a unified probability of association between each of the six basic color terms and each of the 180 color values. In other words, for each individual color term —the six possible color categories— we now have a distribution of normalized association strengths with each of the 180 hues. To each of these six distributions we fit a von Mises, thus obtaining estimates for μ_c and κ_c for each category distribution. With these parameters, we can now use Equation 6 to compute $p(\widetilde{X}_C|\hat{C}¿$ for each color category and each of the 180 hues.

Step 4. Probability of study hues given a noisy sample: In addition to encoding the noisy sample (Equation 3) through a coarse categorical channel, the model encodes it through a higher-resolution channel. That is, it records the exact sample hue from among the set of 180 possible hues. And it then generates a distribution of study hues likely to have generated the encoded sample hue. This is accomplished using Bayes theorem:

$$p(\widetilde{X}_S \vee \hat{S}) \alpha\, p(\hat{S} \vee \widetilde{X}_S)\, p(\widetilde{X}_S) \tag{7}$$

Here, $p(\widetilde{X}_S)$ is a uniform density —all colors are equally likely to occur— such that $p(\widetilde{X}_S \vee \hat{S})$ is simply identical to $p(\hat{S} \vee \widetilde{X}_S)$, a value obtainable by using Equation 3 (with $\widetilde{X}_S$ replacing S_i). Step 4 thus implements what is the typical metric model

applied widely in previous work (e.g. Zhang & Luck, 2008).

Step 5: Estimating the study hue. To arrive at a final estimate of the study hue, we combine the metric information about the noisy sample in Step 4 with the category information about the noisy sample in Step 3. This joint probability distribution is created by combining the two distributions in Step 3 and Step 4 (Equations 6 and 7). We denote the final joint distribution $\widetilde{X}_{JD}$.

$$p\left(\widetilde{X}_{JD}\middle|\hat{S},\hat{C}\right)=\frac{p\left(\widetilde{X}_{C}\middle|\hat{S}\right)p\left(\widetilde{X}_{S}\vee\hat{C}\right)}{\sum p\left(\widetilde{X}_{C}\middle|\hat{S}\right)p\left(\widetilde{X}_{S}\vee\hat{C}\right)}(8)$$

A single hue estimate for the response in a given simulation is obtained by sampling from the distribution $p\left(\widetilde{X}_{JD}\middle|\hat{S},\hat{C}\right)$.

Analysis

We used the CATMET model to generate simulated responses to the delayed and undelayed estimation experiments that participants completed. As noted above, in places where the model employed a color-neutral κ value, it was derived from the data in the appropriate experiment (i.e. delay or undelayed estimation). This was the only parameter derived from the estimation experiments themselves. The parameters employed in the assignment and use of category information were fit to responses in the categorization experiments, which involved unique groups of participants, and which did not involve estimation responses.

The model generated 100 simulated responses to each of the 180 hues, in a simulated version of the undelayed as well as the delayed estimation experiments. Once simulated responses had been generated, we repeated the analyses that had been applied to the empirical results of the estimation experiments; we fit a mixture model to each individual

color, thus allowing us to characterize the stimulus-specific response properties (dispersion and bias) that arose in practice (from a model with no initial representational biases). We then compared these parameters to those that we had obtained from the responses of human participants.

Results

The CATMET model produced biased responses that are similar to the biases measured in the responses of human observers (**Figure 12**). The mean-response (μ) fits we obtained from the model were highly correlated with those of human observers (no-delay; $r = 0.55$, $p < 0.001$; delay: $r = 0.65$, $p < 0.001$). Estimates of response precision, on a color-by-color basis (**Figure 13**) fit to model responses also correlated significantly with the estimates fit to responses from experimental participants (no-delay: $r = 0.16$, $p < 0.05$; delay: $r = 0.39$, $p < 0.001$). While significant, these correlations were weaker than those for bias. In participants, between-observer correlations were also weaker for matching precision than for bias. Thus, the precision of color matches appears less systematic than the bias of color matches.

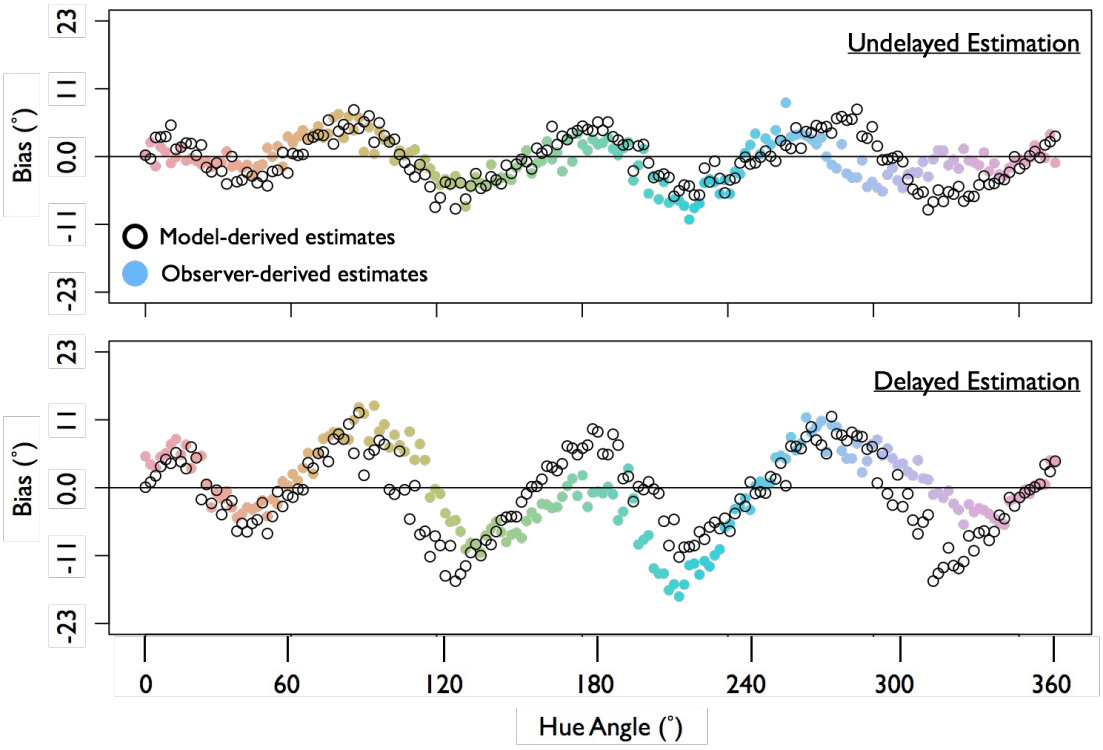

Figure 12. CATMET model- (black open circles) and observer-derived (filled circles) bias estimates for undelayed and delayed estimation, four-category model shown.

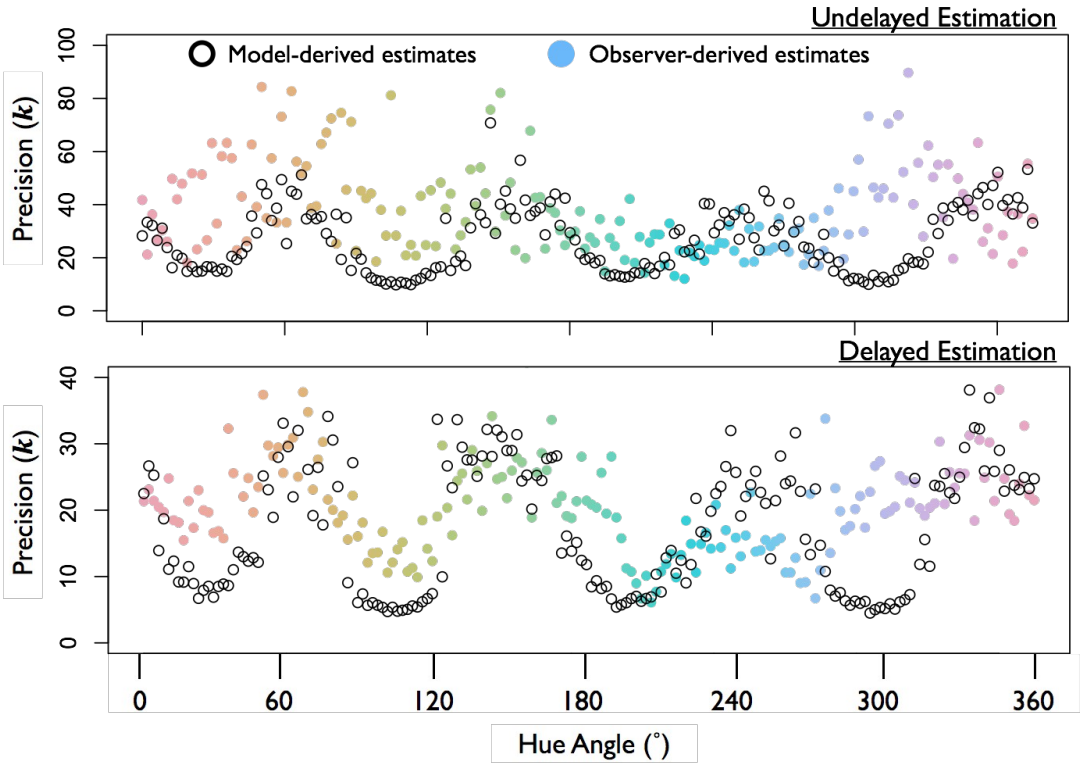

Figure 13. CATMET model- (black open circles) and observer-derived (filled circles) precision estimates for undelayed and delayed estimation, four-category model shown.

These correlations were obtained from a version of the CATMET model utilizing only four (instead of six) categories, ‘orange’, ‘green’, ‘blue’, and ‘pink’. The four-category model performed better than the six-category model, and inspection of observer responses suggests that these categories are more obviously present in the set of colors, with yellow and purple less well represented. But the six-category model faired worse only by a small margin, as can be seen in **Figure 14**, which plots summed squared error for each model’s hue-specific predictions compared to estimates obtained from observer responses.

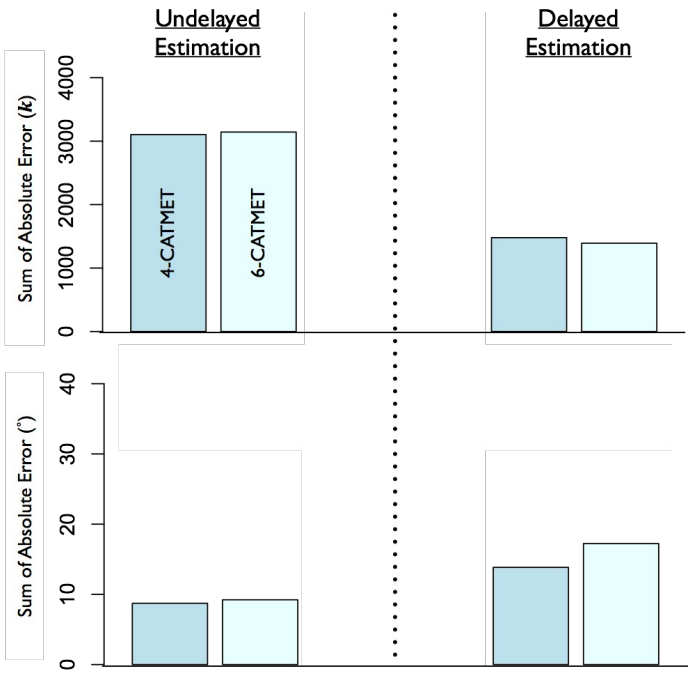

Figure 14. Comparison of precision (top panels) and bias (bottom panels) estimates in undelayed (left) and delayed (right) estimation via sum of absolute error (absolute value) for the four-category (4-CATMET) and the six-category (6-CATMET) CATMET models.

Comparison with other models

We also compared the CATMET model to two additional models, one that uses only category encoding (CATONLY), and one that is more similar to the prevailing approach, using only continuous values, without categories (the METRIC model). Implementation of these models is straightforward. The METRIC model omits all steps apart from 1 and 4 in the CATMET model. It receives a noisy sample, encodes the hue of that sample, which then becomes the basis for an inferred distribution of likely stimulus values (from which responses are sampled). The CATONLY model, in contrast, omits steps 4 and 5. It encodes a noisy sample only in terms of its category. It then generates a distribution of stimulus hues likely to belong to the encoded category, and it samples responses from only that distribution.

We simulated each of these models 100 times for each of the 180 hues, then fitting hue-specific μ and κ estimates to the generated responses (Equation 1), as we initially did for the responses of human participants. These estimates were then correlated with those obtained from the human participants, with r values for each correlation shown in **Figure 15**. The CATMET model produced stronger correlations than the CATONLY model, while the correlations with the METRIC model were uniformly close to zero.

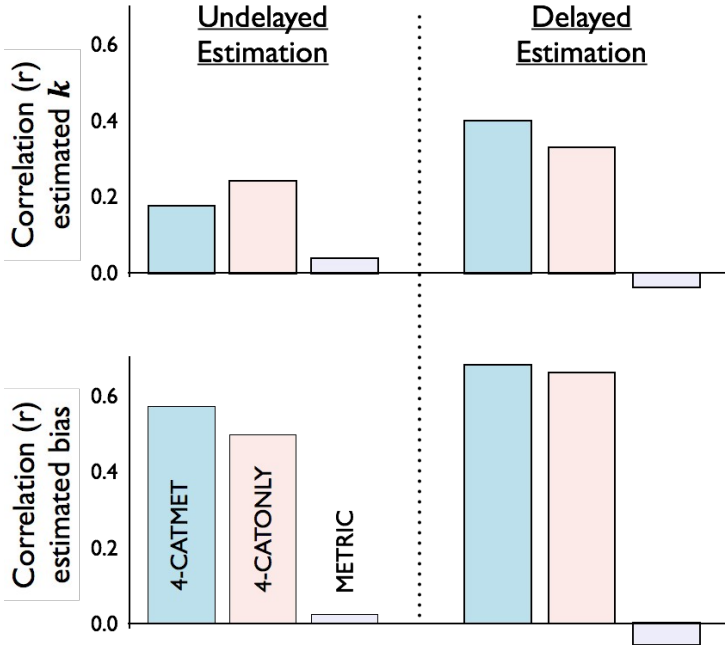

Figure 15. Comparison of correlation values obtained for four-category CATMET model, the four category CATONLY model and the METRIC model, with responses of human observers. Correlations are based on hue-specific model- and observer-derived parameter estimates. Top panel shows precision correlations, and bottom panel shows bias correlations.

Discussion

To summarize, the CATMET model produced stimulus estimates and responses that correlated relatively strongly and significantly with biases observed in human responses. Crucially, it achieved this outcome with underlying representations that were unbiased. Bias emerged by combining category-dependent and value-dependent estimates obtained through simultaneous encoding channels. Devising the model in this way, we

sought to capture what seems to us a commonsense way of characterizing individual colors, as particular cases within categories, as opposed to particular cases within a general and entirely continuous color space.

General Discussion

We sought to test a three-part hypothesis: (1) that working memory maintenance exhibits color-specific biases; (2) that these biases originate in perception, and (3) that observers functionally use two kinds of color information when matching colors between objects, an estimate of hue on a continuous scale —what has been called a “particular” in other contexts (e.g. Huttenlocher et al., 2000)— and a probabilistic category assignment.

First, to test for color-specific estimation biases subsequent to working memory maintenance, we conducted a standard delayed estimation experiment, collecting 60 responses to each of 180 study hues. We found color-specific biases: average estimates frequently deviated from the study hue. Importantly, these biases correlated significantly across independent observers. Second, these color-specific delayed biases were significantly correlated with color-specific biases measured in an undelayed version of the task. This suggests perceptual origins for these effects, or minimally, origins that are not dependent on imposed memory maintenance and an absent target stimulus. Additionally, we found reliable patterns of differences in response precision across hues, suggesting differences in the fidelity with which observers estimate hue values among exemplars with equal contrast and luminance. To our knowledge, this is the only study to investigate delayed estimation with confirmed equal luminance and background contrast among rendered stimuli.

To investigate the role of color categories in these effects, we utilized a pair of experiments in which (different) groups of observers either selected a best name for each of 180 hues, or selected a best example for each of six color names from the basic color terms (Berlin & Kay, 1969). Consistent with previous results using similar tasks (e.g. Witzel and Gegenfurtner, 2013; Boynton & Olson, 1990, Sturges & Whitfield, 1997) we observed systematic responses, with most hues receiving a single color term reliably, some —which we interpreted as category boundaries— receiving two names with nearly equal proportion, and with a few hues repeatedly tagged as best examples —which we interpreted as focal colors. The degree of bias and response precision were both significantly predicted by a hue's distance from the nearest category focal color.

Finally, we presented a dual content model that can account for the observed hue-specific estimation properties and interactions with category landmarks. The model is critically different from prevailing models in that it encodes noisy chromatic signals through two channels, a high-resolution channel that records the signal hue in continuous terms, and a coarse channel that records only a signal's category. It then uses each of these contents to assess the probability that any given stimulus would have induced the encoded contents, and it combines these assigned probabilities to produce a jointly determined estimate of the stimulus. In this model, the first channel is bias-free. Bias emerges, through the interaction with the category assignment: hues that are already good category exemplars will show less bias than hues near boundaries, since the category distribution generated in response to an encoded category describes the strength of each hue's association with a given category. These results have important practical and theoretical implications for the study of color working memory and perception, in

particular, and visual working memory, in general, which we discuss in detail below.

Previous evidence for hue-specific bias and precision

Previous work has yielded contradictory results about the relationship between color categories and color memory. For example, Uchikawa and Shinoda (1986) reported that colors near category borders are remembered more precisely than focal colors are (see also Bornstein & Korda, 1984; Boynton et al., 1989; Roberson & Davidoff, 2000; Pilling et al., 2003). In contrast, Bartleson (1960) reported that focal colors are remembered better than boundary colors, and others reported that they are remembered more precisely (Heider, 1972). Still other studies have failed to find systematic relationships between categories and fidelity of color memory; Witzel & Gergenfurtner (2013) found that category boundaries are not broadly predictive of stimulus-specific differences in discrimination thresholds and others have reported a lack of systematic bias as a function of hue (Siple & Springer, 1983; Allred & Olkkonen, in press; Jin & Shevell, 1996).

One potential explanation for mixed results involves differences in methodology. Alternative forced choice (AFC) methods for example may lead observers to rely on category and particular encodings differently than they do in estimation tasks. However, several observations suggest that our findings may generalize to other tasks. First, we previously reported hue-specific response precision using an estimation task with a different response method (Bae et al., 2014): an aperture through which participants rotated a color wheel to reveal one hue at a time (see also van Den Berg et al., 2012). Hue-specific responses in this experiment correlated significantly with responses in the standard estimation experiment. The effects in a standard estimation task therefore

generalize to an adjustment procedure. And second, the relative size of the biases we have reported here are consistent with those reported elsewhere in tasks using AFC methods (Olkkonen & Allred, 2014; Nemes et al., 2010). We found values up to 10°, but with significant and systematic effects as small as 2° near focal colors, which is the smallest measurable effect in these experiments.

Two other methodological issues, both involving sampling, may produce differences between studies. First, if study stimuli sample only a small region of color space, or coarsely sample large regions of color space, they are ill-quipped to uncover patterns of responses across a hue circle (Allred & Olkkonen, in press; Hedrich et al., 2009, Ling & Hurlbert, 2008). Second, if study stimuli are sampled too coarsely, this could also produce the impression of relatively discrete and precise —as opposed to probabilistic— category boundaries. To see why, consider the pattern of results in Figures 6 and 10. We have demonstrated that bias near boundaries is toward focal colors. Imagine that colors on either side of the blue/green border are sampled—a between-border discrimination. If the border colors sampled are very far from the border, the focal bias will pull the just-green toward green and the just-blue toward blue, and the between-category discrimination will appear very good. If, on the other hand, the colors sampled are very close to the border region, study colors will be easily confused. Thus many small-spaced samples across a relatively large space may be necessary to identify the kinds of effects found in our study.

Finally, it is important to note that Zhang and Luck (2008) in their original report did investigate the possibility of category effects, and found none. Specifically, Zhang & Luck (2008) were concerned that participants may encode stimuli *only* in terms of color

categories, then selecting a nearby focal color value, but respecting category boundaries when making responses. To investigate this possibility they conducted an appropriate analysis, generating a heat map for responses given each target value with a memory load of one. A category-only representation, they predicted, would produce a staircase pattern in such a heat map; but they found a continuous distribution, with average responses near target values. The problem is that this analysis assumes clear, 'noiseless' boundaries and focal colors. The noisy nature of category boundaries, in practice, means that responses near boundaries will appear 'fuzzy,' not staircase-like, even if observers respect boundaries. (Indeed, we were able to replicate their analysis with our data). Likewise, the noisy focal colors will lead to continuous distributions of category responses rather than discrete ones.

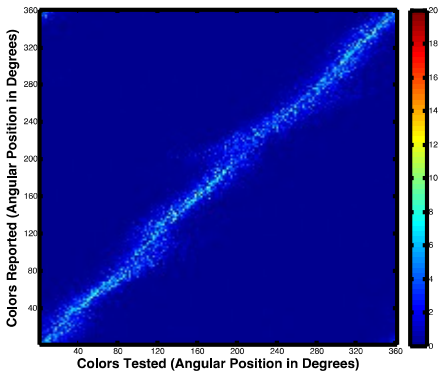

Figure 16. Heat map showing color reports as a function of a target's true color in delayed estimation, replicating an analysis conducted by Zhang & Luck (2008; see their supplementary material).

With the data from our delayed estimation experiment —which clearly include category effects— we were able to produce a heat map of responses very similar to the one produced by Zhang & Luck (2008) and meant to suggest an absence of category effects (Figure 16). In contrast, Figure 6 presents an alternative route to detecting non-uniformity in responses, one that many groups can easily apply to their data sets (assuming each hue has been presented as target a sufficient number of times). There, we plotted normalized response frequency for each hue. There are clear peaks and valleys; retrospectively, it is clear that the biggest effects are at the category prototypes, not the boundaries. If hues generally elicited similar and unbiased response distributions, these overall distributions should be close to uniform (each color was the target equally often). The distributions clearly are not uniform. Figures 7 and 8 plot precision and bias estimates for each color with and without delay.

Overall, a contribution of this work to ongoing research on precision and bias as a function of category structure is in demonstrating that the estimation paradigm —devised for, and until now, used only to study working memory— can serve as an efficient paradigm for studying color perception. Forced choice and related psychophysical approaches require too many trials to design experiments with 180 hues and sufficient numbers of comparative observations. Future work should continue to investigate border and focal color performance, perhaps using estimation as a means to select smaller subsets of important comparisons for use with forced choice and related methods.

Color terms and categories; verbal versus visual memory

Throughout this report we have used ‘color categories’ and ‘color terms’ interchangeably. But some have drawn a distinction between linguistic labels that do not

necessarily map onto underlying representations, and category markers in visual processing of color. This distinction may also relate to a common distinction between verbal and visual working memory (Baddeley & Hitch, 1974), with some arguing that color terms can be stored in verbal working memory, while visual working memory traffics only in continuous coordinates (Luck & Vogel, 2013).

Importantly, the practical implications of our work are independent of whether the categorical channel is verbal or non-verbal. We have demonstrated empirically that participant responses vary by hue in ways that relate to color terms, and that these responses can be modeled by combining probabilistic categorization with continuous hue estimates. Regardless of the underlying cause, this fact is important for understanding behaviors guided by visual working memory.

Theoretically, though, we would suggest that our results are consistent with the hypothesis that categorization occurs as part of visual processing, before any additional verbal labeling takes place. Category effects emerged in undelayed estimation, when resorting to verbal encoding is unnecessary since the study hue remained perpetually in view during response selection. Similarly, in our previous study (Bae et al., 2014) category effects were present with very short exposure and delay periods (100 ms each) and with large memory loads, where verbal encoding and rehearsal would be difficult and unlikely.

Note that categorical processing need not involve verbal rehearsal in principle. In the case of object orientation, for example, degrees of tilt are coded within the context of associated category labels related to object-internal axes and external frames of references. Roughly, this can be thought of as coding an object as ‘the top of the object is

tilted to the left, by 30 degrees' in contrast with 'tilted 330 degrees.' Categorical, non-verbal encoding of orientation appears critical for explaining neuropsychological dysfunction as well as performance asymmetries with healthy participants (adults and children; Gregory et al., 2011; Gregory & McCloskey, 2010; McCloskey, 2009; Valtonen et al., 2008; McCloskey et al., 2006). Similar conclusions have been reached in the context of orientation and visual search, where it has been suggested that objects are *preattentively* categorized as "steep," "shallow," "tilted-left," and "tilted-right," with attentive processing then augmenting these representations with continuous angular values (Wolfe et al., 1992; See also Foster & Ward, 1991; Treisman & Gormican, 1988).

In the case of color, whether non-verbal categorization takes place has long been an important question (along with broader questions about the impacts on perception of verbal categorization). Evidence that is consistent with nonverbal categorization taking place within perception includes neural evidence of early categorical encoding in the brain (Stoughton & Conway, 2008; Bird et al., 2014), categorical color constancy in perception of real-world scenes (Olkkonen et al., 2010), and categorical effects on visual search for colored targets (Daoutis et al., 2006). The reported results contribute to this body of evidence by demonstrating that categorization influences matching performance even with an in-view stimulus. Strengthening this evidence, as well, is the reported reliability of inter-observer category judgments and the ability to predict categorical influences on matching performance in one group of participants based on category landmarks identified by other individuals.

Non-uniform visual memory

The empirical results presented here falsify key assumptions built into current

models of working memory. Specifically, we have demonstrated that the fidelity of working memory —both in terms of bias and precision— is not uniform across hues with equal luminance and equal chromatic contrast with the background. These results suggest that conclusions previously drawn about working memory utilizing delayed estimation should be reexamined, having incorporated inaccurate assumptions into data analysis and interpretation.

As one example, consider the debate about whether or not observers ever 'drop' items from memory —perhaps because of a fixed capacity limit (see e.g. Ma et al., 2014; Luck & Vogel, 2013). Because the question is about whether some responses amount to random guesses, average angular error cannot be used to compare theories; it would conflate target-directed and 'guess' responses. This calculus led Zhang & Luck (2008) to their influential mixture model (Equation 1), designed to estimate average guessing rate and average response precision by best accounting for the individual angular errors that participants produce on each trial. The fitting seeks parameters that manage the tradeoff between lowering precision and, effectively, counting fewer responses as guesses (see also Suchow et al. 2014). But in the way the fitting has been done, it incorporates the assumption that all target-directed responses should look more or less the same, or equivalently, that no target-directed responses should look more guess-like than ones directed to any other target. Our results invalidate this assumption: some color targets do tend to elicit responses that are more distributed than others and with means distant from the target, responses that would look like guesses under a high-precision, unbiased assumption applied to all colors equally.

The same concerns apply to many modifications, extensions, and proposed

alternatives to the Zhang & Luck model. For example, Bays and colleagues (2008) proposed that in addition to target-directed and guess responses, observers sometimes make *non*target-directed responses, arising from feature and object misbindings (Treisman & Gelade, 1980). To estimate the frequency of these occurrences, they added a misbinding term to the Zhang & Luck (2008) model. In this case, the misbinding term included the same precision parameter as the target-directed term in the model. In other words, it implemented a uniformity assumption in two places. On this basis, Bays and colleagues argued that previous models produced the *appearance* of high guessing rates —interpreted to demonstrate fixed capacity limits— because they misattributed misbinding as guessing.

It may turn out that deriving an estimated misbinding rate with a base that is more similar to our model (or some other set of non-uniform expectations) will produce very similar estimates as those obtained previously. But at this stage the question remains open, empirical, and non-trivial. Assigning a probability to a given response under the assumption that it reflects a misbinding depends on the probability one would assign were it actually a response to the same hue in the case that the hue were the actual target. Just like estimating guessing rates, accurately estimating misbinding rates depends on one's expectations about what target-directed responses will look like for each hue, and we have demonstrated that those expectations should not be uniform.

A final example concerns recent models that propose stochastic causes of inter-trial and inter-item precision (e.g. van den Berg et al., 2012; see also Fougnie et al., 2012). Essentially, these models propose that representational precision does not have the same value at all moments in time, and should itself be thought of as drawn from a

(gamma) distribution. To account for seemingly unlikely responses as nonetheless target-directed, the models ultimately suggest that with some frequency, precision is very low, making large angular-error responses more likely than they might otherwise appear (i.e. given a single precision value applied to all trials). The radical significance of this hypothesis is in the suggestion that there may be no fixed capacity limits in working memory whatsoever, evident in the complete absence of guessing responses in model fits.

But the methodological and analytical problem here should be clear: if each trial has a different target color, and different colors tend to produce different response distributions —some that are relatively biased and imprecise— then color-driven trial-by-trial variability needs to be accounted for before further stochastic variability can be evaluated. The relevant models were fit under an assumption of color uniformity.

None of the studies just mentioned are unique with respect to a uniformity assumption. In fact, all delayed estimation experiments we are aware of, including those investigating other visual features, appear to assume uniformity. And there are reasons to expect that non-uniformity extends to other stimulus domains. Orientation is probably the second most common feature in studies of visual working memory with delayed estimation. All the relevant studies in this domain also seem to assume representational uniformity. But there are extant results that should give pause. There are known orientation-dependent asymmetries in visual search (Wolfe et al., 1992; Foster & Ward, 1991; Treisman & Gormican, 1988), there are theories of orientation representation that rely on categorical variables (McCloskey, 2009), and the accuracy of orientation estimation is known to depend on orientation, apparently driven by prior expectations over orientation frequencies (Girshick, Landy, & Simoncelli, 2011).

Thus we suspect that uniformity assumptions are violated in practice in all or nearly all estimation experiments where they have been applied, certainly in all cases pertaining to color. Recognizing this may turn out to be a positive development. Debates concerning the underlying structure of visual working memory appear intransigent. Perhaps the impasse is to some degree caused by unexpected perceptual non-uniformity interacting with individual stimulus and data sets.

Finally, we note that there are many ways to formally characterize non-uniformities in a relevant feature space. The CATMET model does so on the basis of category identification experiments of a manageable size, and it is a natural extension of the original Zhang & Luck mixture model. In particular, CATMET uses a single precision value, but produces non-uniform estimates through combination of hue information with category information. In this way, it may supply a quick, initial method for establishing parameter estimates for guessing rates, precision, and misbinding rates as a function of memory load. We hope that further research will identify alterations that can more completely model stimulus-specific response properties and also illuminate the nature of visual working memory limits.

Categories as priors

The main theoretical contribution of this work is to support the hypothesis that estimation abilities for color rely on both continuous and categorical representations, even when a stimulus is in view. What appears, in aggregated responses, as differences in the memorability of different colors is the consequence of a tendency to categorize colors such that some are better examples of a given category than others, and with some as

reasonable examples of more than one category. Colors are more accurately and precisely remembered when they are good examples of their respective categories.

The CATMET model was inspired by related models described by Huttenlocher and colleagues (2000) in the case of spatial memory, and thus it relied on a category-encoding channel. This seems intuitive to us in the case of color, where typical discourse will refer to particulars within a category, as opposed to just particulars. To pump intuition: it seems that a paint buyer is more likely to hold up a sample and say, “We want this blue”, than to say, “We want this color.”

But there are other ways one might arrive to similar outcomes. One important possibility is that perceptual context effects elicit the bias: embedding a hue in the color wheel may alter its perception compared to the study hue. Perhaps the color wheel itself draws responses to particular points —category centers. Could the results be a response bias caused by perception of the *color wheel*, rather than any actual encoding of the study hue’s category? Although this kind of perceptual context effect may play a role in estimation without delay, we note that the effects were even larger in the memory experiment. Thus the bias is not purely a perceptual context effect.

A more thorny issue concerns whether the samples were actually *encoded* as categories, as our model and theorizing suggest. Perhaps a purely metric encoding interacts with a perceptual context effect at the wheel to produce response bias. In this case, the noisier metric encoding during estimation with delay would increase the relative weight of the perceptual context effect. For example, an observer may encode the sample as #136, but upon inspecting the wheel, note that #139 is a better example of the kind of category that #136 belongs to. There are a number of reasons to think this is not the best

explanation for the effects. Specifically, in our previous work (Bae et al., 2014), we found the same pattern of stimulus specific effects using a different response method, an aperture for viewing a single color at a time with the wheel rotating below. Similar biases have also been found in research with alternative-forced-choice methods (Olkkonen & Allred, 2014; Nemes et al., 2010). Thus it seems inaccurate to describe the effects as merely response bias, driven by the response method.

But there is one other alternative that may suggest a useful distinction between *working* memory and *long-term* memory in the mechanisms that support color matching. This alternative relies on long-term memory to encode a prior over hues that can reflect category structure. That is, from a more typical Bayesian perspective, a non-uniform prior over hues —with higher probabilities at focal colors— might produce the effects without an explicitly categorical encoding of each instance. We cannot exclude this possibility based on our current analyses, and we welcome future investigation of related models that are more traditionally Bayesian. Indeed, the consequences of a categorical encoding channel in the CATMET model are not very different from those that would be expected from a general prior over colors. The latter would bias participants away from any unlikely hues. In the case of CATMET, the impact of category encoding is ultimately to bias participants away from unlikely hues within a known category.

Operationalizing the impact of categories through a Bayesian prior has the advantage of connecting delayed and undelayed hue estimation to the much larger program of research involved in resolving *memory* as well as *perceptual* uncertainty. Bayesian priors are expected to apply to the perceptual appearance of stimuli, even in view. In perceptual contexts, Bayesian models have successfully explained stimulus-specific patterns of bias

in many domains, including size (Ashourian & Loewenstein, 2011), time (Jazayeri & Shadlen, 2010), motion speed (Stocker & Simoncelli, 2006), and orientation (Girshick et al., 2011). Given noisy signals that depend on interactions with viewing conditions, priors facilitate perception by directing observers away from generally unlikely conclusions, and towards generally likely ones. Such priors —whether implemented as priors or as category encoding— should have stronger effects when signals are noisier. Under the presumption that signals associated with absent objects are noisier than signals associated with viewable ones, it makes sense that an imposed memory delay appears to have the impact of increasing category-related biases compared to undelayed conditions. From this perspective, perception and working memory are perhaps less distinct than typically portrayed. Both face the challenge of estimating properties of the physical world from noisy sensory signals.

Conclusion

Interest in working memory has largely focused on the nature of underlying limits that restrict the amount and quality of content that the system can store. Relatively neglected, however, has been the nature of the content itself —the variables whose values the system stores in order to describe a stimulus. We have shown that in the case of color working memory, assumed contents inaccurately omit categorical variables, and as a result, produce unwarranted assumptions about content uniformity in the system's outputs. This demonstrates how limits on content cannot be studied effectively without also characterizing content empirically. Moreover, a research program that considers the contents of working memory systems inherently situates the system within a broader suite of behavior-guiding mechanisms. The contents of working memory are usually acquired from

perceptual inputs, and the nature of working memory outputs depends not only how much it stores, but also on what it stores.

References

Allred, S. R. & Flombaum, J. I. (2014). Relating color working memory and color perception. *Trends in Cognitive Sciences*, *18*, 562-565.

Allred, S. R. and Olkkonen, M. (In press). The effect of memory and context changes on color matches to real objects. *Attention, Perception, & Psychophysics*

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, & Psychophysics*, *74*, 891-910.

Ashby, F. G. & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(3), 372-400.

Ashourian, P. & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One*, *6*(5), e19551.

Baddeley, A. D. & Hitch, G. J. (1974). *Working memory*. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 8, pp. 47–89). New York: Academic Press.

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. F. (2014). Stimulus specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 1–23.

Bartleson, C. J. (1960). Memory colors of familiar objects. *Journal of the Optical Society of America*, *50*, 73–7.

Bays, P. M., Catalo, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 1-11.

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. University of California Press.

Bird, C. M., Berens, S. C., Horner, A. J., & Franklin, A. (2014). Categorical encoding of color in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4590–5.

Bornstein, M. H. & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological research*, *46*(3), 207-222.

Boynton, R. M., Fargo, L., Olson, C. X., Smallman, H. S. (1989). Category effects in color memory. *Color Research and Application*, *14*, 229-234.

Boynton, R. & Olson, C. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research*, *30*(9), 1311–1317.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity : Beyond individual items and toward structured representations. *Journal of Vision*, *11*, 1–34.

Brady, T.F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G.A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*, 981-990.

Brainard, D. H. (2003). Color appearance and color difference specification. In S. Steven K. (Ed.), *The Science of Color* (2nd ed., Vol. 116, pp. 191–216). Washington, D.C.: Optical Society of America.

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

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. *Encyclopedia of Imaging Science and Technology.* John Willey and Sons, Inc.

Brouwer, G. J. & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *The Journal of Neuroscience*, *33*(39), 15454-15465.

Conway, B. R. & Tsao, D. Y. (2006). Color architecture in alert macaque cortex revealed by FMRI. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(11), 1604–13.

Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2006). Within-category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin & Review*, *13*(2), 245–50.

Duffy, S., Huttenlocher, J., Hedges, L. V, & Crawford, L. E. (2010). Category effects on stimulus estimation: shifting and skewed frequency distributions. *Psychonomic Bulletin & Review*, *17*(2), 224–30.

Emrich, S. M.,& Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, *12(4):12*, 1-16.

Foster, D. H. & Ward, P. A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *243*(1306), 75-81.

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*.

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.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature communications*, *3*, 1229.

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., & Luck, S. J. (2010). Reduced capacity but shared precision and maintenance of working memory representations in schizophrenia. *Archives of General Psychiatry*, *67*, 570-577.

Gregory, E., Landau, B., & McCloskey, M. (2011). Representation of Object Orientation in Children: Evidence from Mirror-Image Confusions. *Visual Cognition*, *19*, 1035-1062.

Gregory, E., & McCloskey, M. (2010). Mirror-image confusions: Implications for representation and processing of object orientation. *Cognition*, *116*, 110-129.

Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2009). Color constancy improves for real 3D objects. *Journal of Vision*, *9*, 1–16. doi:10.1167/9.4.16

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1), 10–20.

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological science*, 0956797612459767.

Horwitz, G. D., & Hass, C. A. (2012). Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing. *Nature Neuroscience*, *15*(6).

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology. General*, *129*(2), 220–41.

Jazayeri, M. & Shadlen, M. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026.

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

Johnson, E. N., Hawken, M. J., & Shapley, R. (2001). The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nature Neuroscience*, *4*(4), 409–416.

Johnson, E. N., Hawken, M. J., & Shapley, R. (2004). Cone inputs in macaque primary visual cortex. *Journal of Neurophysiology*, *91*(6), 2501–2514.

Koida, K., & Komatsu, H. (2007). Effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. *Nature Neuroscience*, *10*(1), 108–16.

Linhares, J. M., Pinto, P. D., & Nascimento, S. M. (2008). The number of discernible colors in natural scenes. *Journal of the Optical Society of America A*, *25*(12), 2918–2924.

Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, *25*(6), 1215-1226.

Luck, S.J. & Vogel, E.K.(2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences, *Trends in Cognitive Sciences*, *17*, 391-400.

Ma, W.J., Husain, M., & Bays, P.M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*, 347-356.

McCloskey, M. (2009). *Visual reflections: A perceptual deficit and its implications*. New York: Oxford. Oxford Scholarship Online.

McCloskey, M., Valtonen, J., & Sherman, J. (2006). Representing orientation: A coordinate-system hypothesis, and evidence from developmental deficits. *Cognitive Neuropsychology, 23*, 680-713.

Nemes, V. A., Parry, N. R. A., & McKeefry, D. J. (2010). A behavioural investigation of human visual short term memory for colour. *Ophthalmic & Physiological Optics*, *30*(5), 594–601.

Olkkonen, M., & Allred, S. R. (2014). Short-Term Memory Affects Color Perception in Context. *PLoS ONE*, *9*(1), e86488

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*(11), 1–15.

Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition*, *31*(4), 538-551.

Pointer, M. R. & Attridge, G. G. (1997). The Number of Discernible Colours. *Color Research & Application*, *23*(1), 17–19.

Roberson, D. & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*,*28*(6), 977-986.

Schneegans, S., Spencer, J. P., Schöner, G., Hwang, S., & Hollingowrth, A. (2014). Dynamic interactions between visual working memory and saccade target selection. *Journal of Vision*, *14*, 1-23.

Siple, P. & Springer, R. M. (1983). Memory and preference for the colors of objects. *Perception & Psychophysics*, *34*(4), 363–70.

Stocker, A. A. & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585.

Stoughton, C. M. & Conway, B. R. (2008). Neural basis for unique hues. *Current Biology : CB*, *18*(16), R698–9.

Souza, A., Rerko, L., & Lin, H-Y. (2014). Focused attention improves working memory: implications for flexible-resource and discrete-capacity models. *Attention, Perception & Psychophysics*, *76*(7), 2080-102.

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception & Psychophysics*, *76*(7), 2071-9.

Sturges, J. & Whitfield, T. W. A. (1997). Salient features of Munsell colour space as a function of monolexemic naming and response latencies. *Vision Research*, *37*(3), 307–13.

Tresiman, A. & Gelade, G. (1980). A feature integration theory of attention. Cognitive Psychology, 12, 97-136.

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological review*, *95*(1), 15.

Uchikawa, K. & Shinoda, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research & Application*, *21*(6), 430-439.

Valtonen, J., Dilks, D. D., & McCloskey, M. (2008). Cognitive representation of orientation: A case study. *Cortex*, *44*, 1171-1181.

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*, *109*, 8780-8785.

Wilken, P. & Ma, W. J. (2004). A detection theory of change detection. *Journal of Vision*, *4*, 1120-1135.

Witzel, C., & Gegenfurtner, K. (2013). Categorical sensitivity to color differences. *Journal of Vision*, *13*, 1–33.

Wolfe, J.M., Friedman-Hill, S.R., Stewart, M I., & O'Connell, K. M. (1992) The Role of Categorization in Visual Search for Orientation. Journal of Experimental Psychology, 18(1), 34-49

Wyszecki, G., & Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae.* New York: John Wiley and sons.

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.

Author Note

This research was supported in part by grant NSF CAREER BCS-0954749 to SA, and by a Research Expansion Award to GYB administered by the Johns Hopkins University Department of Psychological and Brain Sciences and funded by the Walter L. Clark Fellowship Fund. The authors also thank Ed Vogel, Brent Strickland, and Daryl Fougnie for thoughtful comments and suggestions.